1-1-2019

# Youth Screening Protocol For General Psychopathology

Steven Bishop

# YOUTH SCREENING PROTOCOL FOR GENERAL PSYCHOPATHOLOGY

A Thesis

presented in partial fulfillment of requirements

for the degree of Master of Arts

in Clinical Psychology

The University of Mississippi

Steven M. Bishop

December, 2019

ABSTRACT

There is a high prevalence rate of psychopathology among US youth, with at least one in four youths meeting clinical criteria for a mental health condition. Those youths who experience psychopathology tend to demonstrate greater functional impairments and have more adverse outcomes compared to youth who do not have these frequently occurring conditions. It is unfortunate that many of these conditions, along with their deleterious effects, are poorly identified in pediatric settings despite the availability of screening instruments. Most screening instruments, however, assess for domain-specific areas of psychopathology only, and can require substantial time to administer and interpret within the typical timeframe of a pediatric visit. The present study sought to design a brief, psychometrically sound, general youth screening instrument using data obtained from the NCS-A. Approximations for sensitivity, specificity, positive predictive value, negative predictive value, and predictive accuracy did not support the development of a broad youth screening instrument to assess psychopathology more generally. The current lack of construct validity for a more general youth screening instrument are discussed, along with future areas of research.

TABLE OF CONTENTS

# I. INTRODUCTION

## Prevalence.

Psychiatric disorders among youth in the United States (US) are highly prevalent, with at least one in every four youths meeting diagnostic criteria for at least one condition over their lifetimes (Merikangas, He, Burstein, Swanson, et al., 2010). Similarly, the high prevalence of these disorders can be found in earlier well-known regional studies of children and adolescents (Costello, Mustillo, Erkanli, Keeler, and Angold, 2003; Canino, Shrout, Rubio-Stipec, Bird, et al., 2004; Roberts, Roberts, and Xing, 2006), as well as international studies (Polanczyk, Salum, Sugaya, Caye, and Rohde, 2015). For perspective, mental health disorders among US children and adolescents are more prevalent than most pediatric medical conditions, including obesity, Diabetes, and asthma (Perou, Bitsko, Blumberg, Pastor, Ghandour, Gfroerer, et al., 2013; Suryavanshi and Yang, 2016).

## Impact and Comorbidity.

Youth who experience these frequently occurring disorders tend to exhibit greater impairments in functioning and have more adverse outcomes compared to youth who do not, particularly in areas of school and work performance (McGee, Prior, Williams, Smart, and Sanson, 2002; Barbaresi, Katusic, Colligan, Weaver, and Jacobsen, 2007; Merikangas, He, et al., 2010, Merikangas, He, Burstein, Swendsen, et al., 2011; Green, McLaughlin, Alegria, Costello, Gruber, Hoagwood, et al., 2013; Costello and Maughan, 2015), involvement with law enforcement (Merikangas, He, Burstein, Swendsen, et al., 2011; Coker, Smith, Westphal,

Zonana, and McKee, 2014; Abram, Zwecker, Welty, Hershfield, Dulcan, and Teplin, 2014; Kim-Cohen, Caspi, Moffitt, Harrington, Milne, and Poulton, 2003), physical health (Costello and Maughan, 2015), social withdrawal (Avenevoli, Swendsen, He, Burstein, and Merikangas et al., 2015), and suicidal behavior (Nock, Borges, Bromet, Cha, Kessler, and Lee, 2008; Avenevoli, Swendsen, et al., 2015; Nock, Green, Hwang, McLaughlin, Sampson, et al., 2013). A substantial percentage of afflicted youth also struggle with more than one disorder simultaneously, making it even more difficult for them to function (Merikangas, He, et al., 2010; Angold, Costello, and Erkanli, 1999; Angold, Erkanli, Farmer, Fairbank, Burns, et al., 2002; Wilens, Biederman, Brown, Tanguay, Monuteaux, et al., 2002). Additionally, these adverse effects can have lifelong implications, in that these disorders tend to be sustained from childhood and adolescence through adulthood (Costello, Egger, & Angold, 2005; Kessler, Chiu, Demler, and Walters, 2005; Kessler, Avenevoli, Costello, Georgiades, Green, et al., 2012; Kim-Cohen, Caspi, et al., 2003).

**Identification of Youth Psychopathology.**

Given the prevalence and deleterious effects of psychiatric disorders for youth, it is unfortunate that these conditions are not identified or treated more routinely by practitioners in primary care settings (Gardner, Kelleher, Pajer, and Campo, 2003). This is surprising given that the general medical setting has long been considered an appropriate location to identify and treat or direct the treatment of youth with mental health needs, provided that most children and adolescents engage in at least annual visits with a pediatrician (Mayne, Ross, Song, McCarn, Steffes, Liu, et al., 2016). Nevertheless, research indicates that pediatricians infrequently recognize or treat the mental health needs of youth despite national initiatives to expand their capabilities and integrate these services (Mayne, Ross, et al., 2016). For instance, Gardner, Kelleher, Pajer and Campo (2004) recruited a sample of 377 primary care practitioners to

investigate whether or not practitioners used standardize psychiatric diagnostic criteria during office visits to identify youth with mental health problems. Specifically, the study examined practitioners use of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) when encountering psychiatric symptoms/conditions. Across 3,674 unique visits, the study indicated that practitioners used DSM criteria in 858 (23%) of visits. This finding suggests that practitioners seldom use standardized criteria to identify and address the mental health needs of youth. Interestingly, the study also calculated an adjusted odds ratio to indicate the likelihood that practitioners used DSM criteria. The analysis indicated that primary care practitioners were very unlikely to use DSM criteria at wellness visits (OR = 0.79), but much more likely in visits previously identified as being for the purposes of mental health concerns (OR = 1.99; Gardner, Kelleher, et al., 2004). Regardless of the approach to diagnosis, fewer than 20% of youth with an identified psychiatric disorder actually receive mental health services from their primary care physician (Merikangas, He, et al., 2011). To address these issues, research has argued for some time that existing psychiatric diagnostic screening tools should be habitually taught to medical providers and implemented in primary care settings (e.g., Cassidy and Jellinek, 1998; Grayson and Carlson, 1991).

Practitioners can utilize numerous readily available child and adolescent screening instruments to inform their clinical decisions, many of which have been extensively researched for their feasibility and psychometrics as mental health screeners (Wissow et al., 2013; Lavigne, Meyers, & Feldman, 2016). There are pediatric screening tools that assess for a wide-range of psychopathology (e.g., Behavioral Health Screen and Pediatric Symptom Checklist), while other instruments are more disorder-specific, focusing on a single domain (e.g., State-Trait Anxiety Inventory and Children's Depression Inventory). That said, the following review limits its focus

to psychiatric screeners designed to identify mental health disorders and psychosocial dysfunction among youth in primary care settings. The review is not intended to be an exhaustive list of every available screening instrument used currently in this setting. Instead, it features the more well-known tools created for use in pediatric primary care, researched in that setting, and potentially used to inform actual practitioner decision-making.

**Screening Instruments.**

The Primary Care Mental Health Screener (PCMHS; Hartung & Lefler, 2010) measures a range of mental health problems in children ages 3 to 12 years old across eight subscales (i.e., anxiety, depression, inattention, hyperactivity, oppositional, conduct, developmental, and learning), which more or less correspond to *DSM-IV* symptom criteria for disorders commonly experienced in childhood. The first version was 69 items long; however, the authors shortened the instrument to 35 items using principal components analyses (PCA). In the initial development study, Hartung & Lefler (2010) examined a sample of 328 parents/caregivers (mostly Caucasian (74%) mothers (86%) from a small town private pediatric clinic). The majority of children referenced in this psychometric study were males (60%) between the ages of 3 and 9 years old (78%). The authors conducted PCA independently for the eight subscales of each instrument to determine if the subscale items formed coherent factors. Additionally, they calculated Cronbach's alpha to measure the internal consistency of each instrument's subscale in the total sample and also separately by gender and age group. The results of the PCMHS long version indicated adequate to excellent performance across most calculations ($\alpha$ = .76 to .96), with the only exception being the depression subscale when applied to preschoolers ($\alpha$ = .67). Similarly, the PCMHS short form indicated adequate to excellent performance across most calculations ($\alpha$ = .72 to .95). For preschoolers, however, the shortened version had inadequate

internal consistency for conduct ($\alpha$ = .63), depression ($\alpha$ = . 55), and pervasive developmental ($\alpha$ = . 64) subscales.

The Patient Health Questionnaire for Adolescents (PHQ-A; Johnson et al., 2002) is a 67-item self-report questionnaire designed for use in primary care settings. The PHQ-A assesses for mental health disorders prevalent in youth ages 13 to 18 (i.e., generalized anxiety and panic disorders; major depression disorder and dysthymic disorder (now referred to as persistent depression); eating disorders; and substance use disorders). It is an adaptation of the adult Patient Health Questionnaire (PHQ; Spitzer, Kroenke, & Williams, 1999), which was developed as an efficient, self-administered method of quickly discerning psychiatric symptoms. In their initial validity study, Johnson et al. (2002) examined a sample of 403 adolescents recruited from primary care clinics and school nurse offices in California, New Jersey, New York, and Ohio. The participants were asked to complete the PHQ-A and the Medical Outcomes Study Short-Form General Health Survey (SF-20) at their primary care visit. Those participants who were available later by phone also completed a semi-structured interview administered by Ph.D.-level clinical psychologists blind to the preliminary results of PHQ-A. The adolescents enrolled in the study were mostly Caucasian (77.2%) females (63.3%) between the ages of 13 and 18 years. The researchers calculated the sensitivity, specificity, diagnostic accuracy and strength of agreement of the PHQ-A with a semi-structured interview. The results indicated excellent specificity (.92) and fair sensitivity (.75) across psychiatric disorders. The overall diagnostic accuracy of the PHQ-A was good (.88), with a high negative predictive power (≥ 93%) but a lower positive predictive power (71%). In terms of overall diagnostic agreement with a clinical interview as the standard, the PHQ-A demonstrated fair to substantial agreement ($\kappa$ = 0.40 to 0.65).

The Pediatric Symptom Checklist (PSC; Jellinek, Murphy, & Burns, 1986) is a 35-item parent questionnaire that screens for a broad range of psychosocial dysfunction in children ages 6 to 12 years old. Parents are instructed to rate each symptom item as "never," "sometimes," or "often," corresponding to scores of 0, 1, 2 points, respectively, which are summed for a total score. The initial PSC was developed as a revised version of the Washington Symptom Checklist (WSC; Wimberger & Gregory, 1968). Subsequently, the original form of the PSC was shortened to a 17-item version, and subscales were added to enhance its sensitivity and clinical utility for practitioners (PSC-17; Gardner, Murphy, Childs, Kelleher, et al., 1999). Jellinek, Murphy, & Burns (1986) conducted two preliminary studies designed to examine the validity and reliability of the PSC. The first examined a sample of 206 parents of children aged 6 to 12 years from three independent pediatric clinics. Participants were asked to complete the PSC and Child Behavior Checklist (CBCL) for comparative purposes. The enrolled sample consisted almost entirely of Caucasian (99%) patients, who had mostly male children (54%) and were from various socioeconomic backgrounds (i.e., 18% were identified as upper SES; 44% were middle; and 38% were lower). The second study examined a sample of 31parents of school-aged children recruited from two different outpatient mental health specialty areas: a hospital based clinic and a community mental health center. Though study two followed similar administrative procedures as study one, there were administrative irregularities with data collection, which contributed to the study's smaller sample size. The participants were all Caucasians, who had mostly male children (77%) and were from mostly from middle (29%) and lower (54%) socioeconomic backgrounds.

In study one, the researchers investigated the validity of the PSC for accurately screening cases by comparing its rate of agreement with the CBCL behavior problems and social

competence subscales using Cohen's Kappa. The researchers utilized a cut score of 28 for the PSC because it provides the optimal level of agreement with the CBCL for predicting general psychiatric risk in the pediatric and mental health samples. Oddly, specific analyses related to this cutoff score were not provided but referenced as being published elsewhere. Regardless, the results indicated that the PSC cutting score of 28 concurred with the CBCL in classifying 80% of the 206 participants as not-at-risk for psychosocial dysfunction and 8% as children at risk. The test-retest reliability for the PSC was also shown to be fair to good for classifying cases ($\kappa = .69$), and the Cronbach alpha for internal reliability was good ($\alpha = 0.86$).

In study two, researchers sought to validate the PSC in a mental health sample by mean score comparison with the pediatric sample and through similar analyses performed in study one. The PSC mean score in the mental health sample was 34.7, whereas the PSC mean score in the pediatric sample was 18.0. The results of the *t* test analysis indicated that there was a significant mean score difference between the two samples ($t = 10.3$, 235 df, $p < 0.001$). The PSC correctly identified 27 (87%) of the 31 children in the mental health sample compared to 165 (80%) of 206 in the pediatric sample. The results also indicated an overall excellent rate of agreement between the PSC and the CBCL in correctly classifying cases ($\kappa = 0.76$). The test-retest reliability was not performed as it had been in study one. Nevertheless, the internal reliability results of the PSC for the mental health sample was also good ($\alpha = 0.85$) similar to that of the pediatric sample in study one ($\alpha = 0.86$).

A number of follow-up studies have indicated that the PSC has fair to good strength of agreement with the CBCL (Jellinek & Murphy, 1988; Simonian & Tarnowski, 2001) and with a structured interview (Anderson et al., 1998; Simonian & Tarnowski, 2001; Walker, LaGrone & Atkinson, 1989). Likewise, its validity and reliability as a screening instrument of psychiatric

risk has been widely supported in medical care settings (Jellinek et al., 1986; Jellinek, Murphy, Robinson, et al., 1988; Murphy, Reede, Jellinek & Bishop, 1992; Murphy, Arnett, Bishop, Jellinek & Reede, 1992; Jellinek, Little, Murphy & Pagano, 1995; Anderson et al., 1998; Walker, LaGrone & Atkinson, 1989).

The original cutoff score of 28 for the PSC has a sensitivity ranging from 0.88 to 0.95 (good to excellent accuracy rating), a specificity from 0.68 to 1.00 (poor to excellent accuracy rating), and can correctly classify between 12% and 22% of children in primary care settings (Jellinek & Murphy, 1988). It should be noted, however, that this original cutoff score was validated in a homogeneous group (i.e., almost entirely Caucasian males) from higher socioeconomic backgrounds (Jellinek et al., 1986). Although the validity and reliability of the PSC has also been examined in more ethnically and socioeconomically diverse populations (Murphy, Reede, et al., 1992; Murphy, Arnett, et al., 1992; Jellinek, Little, et al., 1995), its original cutoff score of 28 was reduced in order to optimally classify cases. Simonian and Tarnowshi (2001), for example, examined a sample of 187 mothers (102 Caucasian, 85 African American) of children (103 females, 84 males) who were all from low socioeconomic backgrounds. They found that the empirically validated cutoff score of 28 when applied to their disadvantaged sample had a 35% false negative rate with a sensitivity of 0.65 (poor accuracy rating). Nevertheless, when they reduced the PSC cutoff score from 28 to 24, its sensitivity improved from 0.65 to 0.89, while its specificity decreased only slightly from 0.94 to 0.87. In another sample of predominantly Hispanic children (N = 199 of 210) who were also from disadvantaged backgrounds, Jutte, Burgos, Mendoza, Ford, and Huffman (2003) showed that it was necessary to reduce the original PSC cutting score to 12 in order to best classify children in this population (sensitivity, 0.74 (a fair accuracy rating); specificity, 0.94 (an excellent accuracy

rating)).

The Eyberg Child Behavior Inventory (ECBI; Eyberg and Ross, 1978) is a 36-item parent informant questionnaire that assesses behavior problems (e.g., refuses to obey until threatened with punishment; argues with parents about rules; hits their parents; steals; lies; physically fights with siblings; argumentative; temper tantrums) in children ages 2 to 16 years old on two dimensions: (1) the prevalence of the behavior, and (2) whether or not the behavior is problematic. The 36 items were the result of a two-year data collection that asked parents of behaviorally disordered children to record their problematic behaviors. In its current form, the ECBI asks parents to rate the frequency of their child's behavior on a 7-point scale from *never* (1) to *always* (7), which is scored by summing each item to yield a total Intensity Score between 36 and 252 points. Additionally, parents indicate whether or not each behavior is problematic, which sums to yield a total Problem Score that ranges from 0 to 36.

In the initial validity study, Eyberg & Ross (1978) collected four different samples of children between the ages of 2 and 7 years old. The target control group consisted of the first 22 families of children (11 females, 11 males) who responded to advertisements posted throughout the community (i.e., local newspapers, pediatric clinics, nursery schools, and parent-child services). The behavior problem group consisted of 43 children (34 males, 9 females) who were referred for psychological evaluation of behavioral difficulties. The clinic control group consisted of 20 children (13 males, 7 girls) referred specifically for intellectual and developmental evaluation, but who had no evidence of prior or current behavioral problems. The researchers also incorporated a treatment intervention group, which consisted of 10 children from the behavior problem group who provided pre- and post-treatment scores on the ECBI. The study utilized the ECBI as its sole dependent measure for data collection, which was analyzed

using both an independent sample *t* test and paired sample *t* test. The results indicated a significant mean score difference in the ECBI Problem Score for the behavior problem group (18.6) compared to the target control group mean score of 4 ($t$ (63) = 8.98, $p < .001$) and the clinic control group mean score of 5 ($t$ (61) = 7.34, $p < .001$). Similarly, there was a significant mean difference in the ECBI Intensity Score for the behavior problem group (158.26) compared to the target control group mean of 100 ($t$ (63) = 7.37, $p < .001$) and the clinic control group mean of 105.10 ($t$ (61) = $p < .001$). Further analysis of ECBI indicated that provisional cut points of 11 for the Problem Score and 127 for the Intensity Score accurately classified 9 of 10 children from the treatment intervention group.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a 25-item informant-rated questionnaire that screens for positive and negative attributes in youth between the ages of 4 and 16 years old. The 25 items are evenly divided between five scales (i.e., hyperactivity, emotional symptoms, conduct problems, peer problems, and prosocial), which was based on prior factor analyses and frequency distributions from a modified version of the Rutter parent questionnaire (Goodman, 1994). Responses are scored on a 3-point scale ("not true," "somewhat true," or "certainly true") and all items within a specific domain are summed to yield scale scores (ranging from 0 to 10). With the exception of the prosocial scale, the other four scale scores are summed for a total difficulties score, which ranges from 0 to 40. Subsequently, there have been several refinements to the original SDQ, which include a self-report version for youth aged between 11 and 16 (Goodman, Meltzer, & Bailey, 1998) and an impact supplement that improved detection of psychopathology and determinants of service use (Goodman, 1999).

In the preliminary development study, Goodman (1997) recruited the parents and teachers of 403 youth between the ages of 4 and 16 years old from two different practice settings

in London, England (a dental clinic and a psychiatric clinic). Respondents were asked to complete the informant-rated version of the SDQ (i.e., same questionnaire for both groups) and the Rutter questionnaires (i.e., Rutter A parent version and Rutter B teacher version). The sample's demographic information was limited to the mean ages of the dental (10.8 years) and psychiatric (9.8 years) samples and the overall distribution of gender (males, 59%; females, 41%). The study utilized Receiver Operating Characteristic (ROC) curves for statistical comparison (with analyses for both scales based on their total scores) and Pearson product-moment correlations (with cross-measure correlations for each type of rater and cross-informant correlations for each type of measure). The ROC curve results were .87 for the parent-informant SDQ, .85 for the teacher-informant SDQ, .87 for the Rutter A parent questionnaire, and .84 for the Rutter B teacher questionnaire. The cross-measure correlations between the SDQ and Rutter scores for the parent report ranged from .78 to .88 and for the teacher report between .87 and .92. The cross-informant correlations between parent and teacher scores ranged from .37 to .65 for the SDQ and from .47 to .57 for the Rutter. Goodman (1997) concluded that these initial findings support the concurrent validity of the SDQ and also its predictive validity.

The study also included provisional cutoff scores in its Appendix, although none of the data or statistical analysis to support these values was provided. Goodman (1997), however, did note that these cutoff scores were preliminary on the basis of samples from this study and other epidemiological surveys that utilized the SDQ (which were not referenced). That said, there are subsequent studies that did provide more definite and empirically-informed cutoff data (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000; Goodman, Renfrew, & Mullick, 2000; Bourdon et al., 2005). Additionally, several studies indicate that the SDQ has excellent agreement with the CBCL (Goodman & Scott, 1998; Vogels et al., 2009), good comparability to

the PSC (Vogels et al., 2009), and even outperforms the CBCL on case identification for inattention and hyperactivity when both instruments were compared to a semi-structured interview (Goodman & Scott, 1999). Furthermore, its validity and reliability as a screening tool of psychopathology has been extensively supported in several community samples (Goodman, Meltzer, & Bailey, 1998; Goodman, Ford et al., 2000; Jee et al., 2011) and across two nationally representative samples of youth in Britain (Goodman, 2001) and the United States (Bourdon et al., 2005; Merikangas, Avenevoli, Costello, Koretz, & Kessler, 2009). As noted by Lavigne, Meyers, and Feldman (2016), the SDQ is also likely the most validated self-report in comparison to a structured interview among extant instruments, which provides a wealth of data concerning its predictive validity of ultimate diagnosis.

The DISC Predictive Scales (DPS; Lucas et al., 2001) collectively comprise a computerized self-report screener developed as a series of concise predictive scales for psychopathology in youth between the ages of 9 and 18 years old. They encompass 12 diagnostic categories (i.e., simple phobia; social phobia; agoraphobia; overanxious (since removed from diagnostic taxonomy); obsessive-compulsive; separation anxiety; eating; major depressive; attention-deficit/hyperactivity; oppositional defiant; conduct; and alcohol/substance use disorders). The original development study entailed using stem items from the Diagnostic Interview Schedule for Children (DISC) to predict impairment criterion scores in a sample of 1,286 youth between 9 and 17 years old. Each stem item was examined via logistic regression and its relative prediction of diagnostic severity was compared to the other items within the same subscale. Those items having the highest regression weights (i.e., predictive utility) and positive responses served as "gate" items in the fully assembled DPS. Following its assembly, the DPS was validated against the previous study's findings in a residential care sample of 884 youth

between 10 and 18 years old. The results indicated that the DPS obtained sensitivity ranges from .70 to 1.00, specificity from .76 to .95, and AUC from .82 to .96. Exceptions to this can be seen in DPS scales for agoraphobia with a sensitivity of .40, and OCD with a specificity of .49. Once validated, the DPS was analyzed in a separate residential care sample of 80 youth between 9 and 17 years old. The analysis indicated that the DPS achieved similar results, with sensitivity ranges from .71 to 1.00, specificity from .61 to .96, and AUC from .72 to .98. The end result of this process was reduction of the 206-item DISC to 76 items in the DPS. Additionally, predicting DISC diagnoses, the DPS demonstrated good to excellent results for sensitivity, specificity, and AUC.

Although the DPS was not designed specifically as a pediatric screener, Husky, Miller, McGuire, Flynn, and Olfson (2011) conducted a medical records review of 483 youth aged 13 to 17 who were given the option to complete the measure at the time of their appointment. The goal was to investigate the practicality of mental health screening integration at one pediatric clinic. The procedures were such that the pediatric clinic was simply provided with the DPS and offered no further support from researchers. The study indicated that this was a feasible method of implementation, in that 44.7% of the medical records reviewed contained a completed DPS. Of those youths who completed the DPS, 13.9% were positively identified as having at least one or more mental health disorder (Husky et al., 2011). Finally, the study also mentioned that this clinic may have been somewhat atypical in that they had already been offering the Columbia University TeenScreen Program since 2005, which could limit the generalizability of results to other clinics without similar attention to mental health issues.

This review featured readily available psychiatric screeners designed or modified specifically for use in pediatric primary care settings. Each screener selected for review has been

researched for its effectiveness to accurately diagnose psychopathology among youth in this setting and reduce the time involved in administration and interpretation, along with its utility to inform the clinical decisions of practitioners. All of the measurements studied included statistical data and were validated with some other gold standard measurement or structured interview across different samples of youth (clinical and nonclinical). Most of the screeners reviewed assessed for a broad range of mental health disorders except for the ECBI, which assessed for behavioral disorders only. Almost all of the screening measurements reviewed ranged between 25 and 76 items (the exception being the PSC-17), potentially making their administration burdensome in pediatric settings. It is possible to reduce the number of items necessary for broad psychopathological screening, however, without sacrificing validity or compromising instrument sensitivity or specificity, through the application of decision-tree analysis. To the author's knowledge, this technique has not yet been applied to the development of a general screener for youth psychopathology, which informs the need for the current study.

**Goals of the Present Study.**

The present study investigated the application of decision-tree statistics to a sizeable and publicly available data set of US adolescents. The researcher aimed to develop a youth screening protocol for general psychopathology from the Composite International Diagnostic Interview (CIDI). It was hypothesized that applying decision-tree statistics to the full-length version of CIDI should converge to establish a brief, psychometrically sound, broad screening tool for youth aged 13 to 18 years. Furthermore, in a manner consistent with the findings from Stewart, Tuerk, Metzger, Davidson, and Young, (2015), it was hypothesized that the resultant decision-tree algorithm used to classify the presence or absence of any psychopathological diagnosis could reduce the number of items considerably (i.e., to between 2 and 10).

# II. METHODS

The Scientific Infusion That Helps (SITH) Lab obtained access to the *Collaborative Psychiatric Epidemiology Surveys* (CPES) database through the Inter-University Consortium for Political and Social Research (ICPSR). The objectives of CPES are to provide publicly available data regarding the prevalence of psychiatric disorders, severity of resultant impairment, and typical treatment patterns. The CPES database combines three nationally representative surveys, which are the National Comorbidity Survey Replication (NCS-R), the National Survey of American Life (NSAL), and the National Latino and Asian American Study (NLASS). Of these initiatives, the NCS-R is a nationally representative sample of adult psychiatric disorders taken from 2001 to 2004 (Kessler & Merikangas, 2004; Kessler et al., 2009). The NCS-R survey extended its lower age range to assess the base rates and correlates of psychiatric disorders in a sample of adolescents, commonly known as The National Comorbidity Survey – Adolescent Supplement (NCS-A; Kessler et al., 2004; Kessler & Merikangas, 2004). The NCS-A involved gathering information from nationally representative youth and at least one of their parents. Child interviews were conducted using a gold-standard psychodiagnostic assessment (the World Health Organization (WHO) Composite International Diagnostic Interview Version 3.0, or CIDI; Andrews & Peters, 1998; Kessler & Ustun, 2004; Merikangas et al., 2009) and parent information was collected via self-report instrument (Kessler et al., 2009; Merikangas et al., 2009; Merikangas et al., 2010). The structure of the child/adolescent data is optimal for the current study given item-level information from a large sample.

**Participants**

The sample included 10,123 U.S. adolescents aged 13 – 18 years, demographics for which appear below. Weighted sampling procedures were employed to validate that the sample was representative of U.S. youth (Kessler et al., 2009). The NCS–A demographic variables included sex, age, ethnicity, parental education and marital status, urbanicity and degree of poverty (Merikangas et al., 2010; see Table 1). See Table 1 Characteristics of the NCS–A (*N* = 10,123) adapted from Merikangas et al., 2010, p. 982.

**Measure**

**The World Health Organization (WHO) Composite International Diagnostic Interview Version (CIDI)**. The CIDI is a fully structured diagnostic interview administered by highly trained lay interviewers to determine *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychological Association, 1994) diagnoses (Merikangas et al., 2010). Even though the version of the CIDI conducted in the NCS–A sample is similar to the one administered in the NCS-R, a work group of researchers and clinicians modified the CIDI to fit the language and experiences of youth, and included an extensive amount of supplemental material in the revised instrument (Kessler & Ustun, 2004; Merikangas et al., 2009; Merikangas et al., 2010). The CIDI assesses lifetime disorders including anxiety (i.e., panic, agoraphobia, social phobia, specific phobia, posttraumatic stress, generalized anxiety, separation anxiety), mood (i.e., major depressive or dysthymia, bipolar I and II), behavior (i.e., attention-deficit/hyperactivity, oppositional defiant, conduct disorder), eating (i.e., anorexia nervosa, bulimia nervosa, binge eating), and substance use (i.e., alcohol abuse/dependence, drug abuse/dependence; Merikangas et al., 2009; Merikangas et al., 2010). As a diagnostic interview, the CIDI is the most widely used instrument in epidemiological studies involving psychiatric conditions (Kessler et al., 1994), which provides lifetime prevalence and present-day diagnoses

of mental disorders based on the conventional definitions of *DSM-IV* (Kessler et al., 1994; Merikangas et al., 2009). Multiple versions of the CIDI have demonstrated robust psychometric properties (Wittchen, 1994; Andrews & Peters, 1998) and concordance with *DSM-IV* psychopathology with a clinical reappraisal sample (Kessler et al., 2009; Merikangas et al., 2010).

**Procedure**

Decision-tree analysis was conducted using the Classification Tree feature in SPSS. This involved assigning all variables to either predictor or outcome roles. Outcome variables were coded as either categorical or continuous, but predictors must be categorical. The process was conducted using Classification and Regression Trees (CART) to evaluate each predictor variable in terms of how well it differentiated subgroups at a particular cut score (King & Resick, 2014). For the purposes of this analysis, diagnostic status (i.e., yes or no) served as the outcome variables and questions about individual symptoms served as the predictor variables. Given that the focus of the study was to develop a general screener, outcome variables were summed across all diagnostic categories to create a single, aggregate variable. In other words, all predictor variables were analyzed in terms of their ability to accurately predict whether or not a given individual has any diagnosis of any kind. Following the recommendations of Onwuegbuzie and Collins (2010), this study set the minimum number of root and terminal nodes at 50 and 25, respectively. Assuming that the model converged to accurately predict outcome, one or more adaptive question sets could be derived. These should form the basis for an algorithm to create the final instrument, which could ideally minimize the number of questions necessary to predict probable diagnosis. Additionally, the sensitivity and specificity of the final measure were calculated.

## III. RESULTS

For the purposes of this analysis, 18 *DSM-IV* disorders diagnosed at 12 months (i.e., ADHD, agoraphobia with/without panic, agoraphobia with panic, alcohol abuse, anorexia, bulimia, conduct, drug abuse, dysthymia, generalized anxiety, major depressive episode, oppositional-defiant, panic attacks, panic disorder, PTSD, separation anxiety, social phobia, and separation anxiety) were aggregated to establish a single outcome variable for analysis (i.e., diagnostic status). As seen in Table 2, in determining the presence of any *DSM-IV* diagnosis at 12 months, 40.9% (*N = 2912*) of the training sample and 41.4% (*N = 1261*) of the test sample met criteria according to the fully administered version of the CIDI. The CART decision-tree algorithm classified the training sample with an overall accuracy rate of 71.9%, which is in the acceptable or fair range, whereas the test sample was classified with an overall accuracy rate of 68.9%, which is in the poor range. Regarding the training sample, the CART algorithm categorized the presence of any diagnosis with a sensitivity rate of 72.0%, and the absence of any diagnosis with a specificity rate of 71.7%, both of which fall into the acceptable range. Furthermore, the CART algorithm generated a positive predictive value of 63.9% (poor range) and a negative predictive value of 78.7% (acceptable range) for the training sample. Considering the test sample, the CART algorithm categorized the presence of any diagnosis with a sensitivity rate of 72.1%, falling into the acceptable range, and the absence of any diagnosis with a specificity rate of 66.7%, falling into the poor range. Similarly, for the test sample, the CART algorithm generated a positive predictive value of 60.5% (poor range) and a negative predictive value of 77.2% (acceptable range) for the test sample.

As previously mentioned (see Methods), predictor variables consisted of individual symptom questions contained within the full-length version of the CIDI. More specifically, particular symptom-related questions were selected and pooled from 17 different CIDI modules (i.e., initial screener questions; ADHD; agoraphobia, conduct, depression, eating, generalized anxiety, intermittent explosive, irritable depression, mania, opposition-defiant, panic, PTSD, separation anxiety, social phobia, specific phobia, and substance) that resulted in a total of 713 questions. Next, the pooled questions were analyzed simultaneously by the CART algorithm to find those items with the best binary split for classifying the outcome variable of interest or, in this case, the diagnostic status of any diagnosis (i.e., presence or absence). Table 3 contains the resultant 20 questions that had the best capability of classification for the current sample.

Figure 1 in the Appendix represents the actual tree structure for predicting the diagnostic status from the 20 questions contained in Table 3. As Figure 1 shows, the best overall predictor of diagnostic status (i.e., node 0) was item 1 (irritable episode). Subsequently, item 2 (during fear-sick to stomach) was the next best predictor for respondents who indicated that an irritable episode was ongoing (i.e., node 1). For individuals who specified a "No" response to item 2, the next best predictor was item 4 (avoidance interfered with work/social life/relations), whereas item 5 (used weight control strategy) was the best predictor for those who indicated a "Yes" response to item 2. Dichotomous classifications were achieved for those individuals who responded with a "little" to item 4 in the succeeding tree branches. In contrast, only a single dichotomous decision was reached for those who endorsed "no/some/a lot/extremely" to item 4 through subsequent branching with item 15 (social situation: felt distant; not really there). At item 5, however, no further splitting was necessary with an endorsement of "greater than a year ago" (i.e., classify present), while a confirmation of "within the past year" resulted in further

splitting with item 10 (recently had alcohol problems). Following this sequence, additional splits occurred with item 17 (laugh/see funny side) and item 18 (treated for eating problems in past year) but neither item achieved a dichotomous decision.

A similar tree pattern can be observed for individuals who indicated that an irritable episode had ended (i.e., node 2), where item 3 (recently had problem because of drug use) was nominated as the best predictor variable. An indication of "greater than a year ago" to item 3 resulted in item 6 (in most severe fear-feel surrounding unreal) being the next best predictor. With item 6, a "No" response indicated the absence of any diagnosis, whereas a "Yes" response continued branching with successive items in order to achieved a dichotomous classification. Likewise, for those who specified "within the past year" to item 3, item 7 (in most severe fear-fear losing control) was chosen as the best predictor variable. As with item 6, a "No" response to item 7 stopped further branching that also indicated the absence of any diagnosis, while an affirmative answer continued the branching process to reach a dichotomous decision. For the current sample, this tree-branching process resulted in tree structures that were between 2-5 items for optimal classification, depending on responses (i.e., either absence or presence of any diagnosis).

# IV. DISCUSSION

The present study sought to apply decision-tree statistics to construct a short, psychometrically sound, general tool to broadly screen for psychopathology in youth aged 13 to 18 years. This involved access of a sizeable, familiar, item-level database that was divided into two samples and independently analyzed. The results indicated that the best classification possible was within the poor-to-fair range across all measures (i.e., sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy). The hypothesized outcome, therefore, of a convergent algorithm with sound psychometric properties was not supported in the present study.

This lack of support is inconsistent with prior research studies that applied a similar decision-tree approach for instrumentation development. Stewart et al. (2015), for example, developed a sequence of diagnostic algorithms from the CAPS that predicted the presence or absence of PTSD at a reliably high rate of accuracy. In another study, Sattler, Bentley, and Young (2019) derived a series of algorithms from the CIDI that exhibited generally high rates of accuracy across diagnostic domains. In an earlier, similar study, Sattler, Whiteside, Bentley, and Young (2018) successfully abbreviated the OCD subscale of the Spence Children's Anxiety Scale to as little as a single item without comprising its psychometric properties. Although it was a slightly different study, Ebesutani, Bernstein, Chorpita, and Weisz (2012) constructed assessment algorithm protocols from multiple self-report scales that informed the treatment selection and clinical decisions of practitioners. In other words, unlike the present study, prior

studies demonstrated the feasibility of harnessing decision-tree statistics to derive highly efficient and psychometrically sound algorithms from various instruments.

There are, however, notable differences that exist between those previous efforts and the current one, which might have contributed to the observed inconsistencies. One dissimilarity can be seen with the way prior studies employed measurement items relative to diagnosis. For instance, Stewart et al. (2015) used the CAPS instrument to focally determine a diagnosis of PTSD only; Sattler et al. (2019) used the CIDI domain-specific module items to predict a diagnosis in only that particular domain; and Sattler et al. (2018) used items within the OCD subscale to categorize OCD alone. Thus, these studies used measurement items unitarily within a single diagnostic domain, contributing to greater correlation between questions and measurement of directly relevant symptoms. That is to say, individual algorithms were derived from measurements to categorize a specific diagnosis only, whereas the present study applied measurement items broadly relative to any diagnosis. This was in keeping with the primary goal of examining the potential for an algorithm useful in predicting the presence/absence of *any* diagnosis rather than individual algorithms that corresponded to a specific diagnosis. Thus, when compared to the findings of prior studies the discrepancies in the current study could have resulted from methodological differences.

On the other hand, the observed discrepancies can also be explained as a function of construct validity. Following Haynes, Richard, and Kubany (1995), construct validity indicates the extent to which obtained scores from an instrument measure the intended construct when the instrument is implemented. Construct validity encompasses all categories of validity related to instrument construction (i.e., criterion validity, content validity, concurrent validity, convergent and divergent validity, and predictive validity). These other areas of validity, however, cannot be

determined until the content validity of an instrument is established. Thus, content validity indicates that the contents or items within an instrument are relevant to and representative of the stated construct for a specific reason (connected to a theoretical rationale). Relevance is understood to mean that the items contained in an instrument are suitable and purposeful relative to the desired construct, while representativeness means that the items contain all components of the desired construct when measured (Haynes et al., 1995). The CAPS instrument clearly illustrates this point in that its 17 items measure all the dimensions of a PTSD diagnosis. A similar example can be observed in the domain-specific modules of the CIDI, where the items within each module correspond to all of the relevant and representative factors of a particular diagnosis. Considering this framework for instrument construction, it is potentially easier to detect why discrepancies between previous research and current findings resulted. The present study expected sufficient levels of content validity (i.e., relevance and representativeness) in each domain-specific module item to translate to accurate measurement of the aggregated diagnostic construct, which was clearly not the case. This conceptualization inadvertently caused a mismatch among the constructs for two reasons, the first being that domain-specific module items were constructed to function with a particular diagnosis assessed in a specific domain. The second is that individual diagnoses when aggregated created a dichotomous construct (i.e., diagnosis present or absent) where some unique components of an individual diagnosis were no longer relevant to and representative of the aggregate construct. This is clearly illustrated in the poor-to-fair measurements in sensitivity, specificity, and predictive accuracy in the current study (i.e., construct validity). Although the content validity of module items remained with respect to the domains in which they were developed, they inadequately measured the aggregated construct. Therefore, it is quite possible that many of the module items overlapped with one

another, were unrelated to, excluded from, or over-represented the overall construct of some form of diagnosis being present or absent. The agoraphobia module items, for example, could have easily overlapped with panic module items in relation to the aggregated construct. Additionally, items from the social anxiety, panic, and substance-use modules might have been over-represented because of the prevalence of these symptoms among adolescents.

In addition to the lack of support for a convergent algorithm, the present study had several limitations. First, although the CART method conducts a comprehensive analysis of possible splits on every predictor variable, it can be constrained because of its bias towards predictors with more values or that allow for more splits (King & Resick, 2014; Loh, He, & Man, 2015). The present study included many of these predictors that the CART analysis identified and incorporated into the creation of homogenous subgroups in its model, which could have limited its overall viability. There are also several limitations related to the CIDI instrument itself. For example, the assessment of 12-month history of disorders is based on retrospective reports of adolescents and parents (Merikangas et al., 2010), which could be less accurate than reports of contemporaneous symptoms. Additionally, the instrument took a considerable amount of administration time that resulted in multiple interviews for many respondents (Merikangas, Avenevoli, Costello et al., 2009), which could have further limited accuracy in some cases. Despite these possible limitations, this version of the CIDI remains a valid measurement of psychopathology for adolescents and the sample size employed would typically reduce or eliminate threats to statistical conclusion validity. Despite this validity in terms of its intended purposes, the lack of convergent construct validity for predicting a more amorphous categorization of possible disturbance (i.e., any form of diagnosis present or absent) was likely propelled by the domain-specificity of the item generation process. Accordingly, future studies

could apply factor analysis to identify relative constructs or latent variables across diagnostic constructs then construct screener items to match, thus reducing the risk to overall construct validity and potentially enhancing generation of a viable model. Another possibility is for future studies to construct screening items with strong construct validity related to a dichotomous diagnostic variable. For instance, it possible that a face-valid question like, “Have you been diagnosed with a mental health disorder within the past year?” could be content-valid related to the presence or absence of any diagnosis. Construction of a completely new instrument designed for these more general purposes, therefore, could be more successful in achieving the goals of the current study. Along this same line, future studies could adapt a network analysis model approach to develop screening items (i.e., determining symptoms common to multiple disorders that are indicative of generalized risk). Once screening items are constructed for a specified diagnostic construct, future studies could also apply another decision-tree approach such as the Quick, Unbiased Estimation Statistic (QUEST), which reduces some of the biases associated with CART, or optimal decision-tree analysis (i.e., contemporary techniques from computer science that are much more complex and resource-intensive in terms of computer hardware needs). Regardless, programmatic study and adaptation of the mindset that provided the foundation for the current study will be necessary to advance understanding in this domain.

LIST OF REFERENCES

Abram, K.M., Zwecker, N.A., Welty, L.J., Hershfield, J.A., Dulcan, M.K., and Teplin, L.A. (2014). Comorbidity and Continuity of Psychiatric Disorders in Youth After Detention: A Prospective Longitudinal Study. *Journal of American Medical Association of Psychiatry, 72, 1,* 84-93.

American Psychiatric Association, & American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (DSM). *Washington, DC: American Psychiatric Association,* 143-147.

Anderson, D.L., Spratt, E.G., Macias, M.M., Jellinek, M.S., Murphy, J.M., Pagano, M., Griesemer, D.A., Holden, K., and Barbosa, E. (1998). Use of the Pediatric Symptom Checklist in the Pediatric Neurology Population. *Journal of Pediatric Neurology, 20,* 116-120.

Andrews, G., and Peters, L. (1998). The psychometric properties of the composite international diagnostic interview. *Social Psychiatry and Psychiatric Epidemiology, 33*(2), 80-88.

Angold, A., Costello, J.E., and Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40,* 57-87.

Angold, A., Erkanli, A.E., Farmer, E.Z., Fairbank, J.A., Burns, B.J., Keeler, G., and Costello, J.E. (2002). Psychiatric Disorder, Impairment, and Service use in Rural African American and White Youth. *Archive of General Psychiatry, 59*, 893-901.

Avenevoli, S., Swendsen, J., He, J.P., Burstein, M., and Merikangas, K.R. (2015). Major Depression in the National Comorbidity Survey-Adolescent Supplement: Prevalence,

Correlates, and Treatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 55, 1,* 37-44.

Bourdon, K.H., Goodman, R., Rae, D.S., Simpson, and G. Koretz, D.S. (2005). The Strengths and Difficulties Questionnaire: U.S. Normative Data and Psychometric Properties. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 557–564.

Canino, G., Shrout, P. E., Rubio-Stipec, M., Bird, H. R., Bravo, M., Ramirez, R., Chavez, L., Alegria, M., Bauermeister, J., Hohmann, A., Ribera, J., Carcia, P., and Martinez-Taboas, A. (2004). The *DSM-IV* Rates of Child and Adolescent Disorders in Puerto Rico: Prevalence, Correlates, Service Use, and the Effects of Impairment. *Archives of General Psychiatry*, *61*, 85–93.

Cassidy, L.J., and Jellinek, M.S., (1998). Approaches to Recognition and Management of Childhood Psychiatric Disorders in Pediatric Primary Care. *Pediatric Clinics, 45,* 1037-1052.

Coker, K.L., Smith, P.H., Westphal, A., Zonana, H.V., and McKee, S.A. (2014). Crime and Psychiatric Disorders Among Youth in the US Population: An Analysis of the National Comorbidity Survey-Adolescent Supplement. *Journal of the American Academy of Child & Adolescent Psychiatry, 53, 8,* 888-898.e2.

Costello, J.E., Mustillo, S., Erkanli, A., Keeler, G., and Angold, A. (2003). Prevalence and Development of Psychiatric Disorders in Childhood and Adolescence. *Archives of General Psychiatry*, *60*, 837–844.

Costello, J.E., and Maughan, B. (2015). Annual Research Review: Optimal Outcomes of Child and Adolescent Mental Illness. *Journal of Child Psychology and Psychiatry, 56,* 324-341.

Ebesutani, C., Bernstein, A., Chorpita, B., and Weisz, J. (2012). A transportable assessment protocol for prescribing youth psychosocial treatments in real-world settings: Reducing assessment burden via self-report scales. *Psychological Assessment*, *24*, 141-155.

Eyberg, S.M., and Ross, A.W. (1978). Assessment of Child Behavior Problems: The Validation of a New Inventory. *Journal of Clinical Child Psychology,* 113-116.

Husky, M.M., Sheridan, M., McGuire, L., and Olfson, M. (2011). Mental Health Screening and Follow-up Care in Public High Schools. *Journal of the American Academy of Child & Adolescent Psychiatry, 50,* 881-891.

Gardner, W., Murphy, M., Childs, G., Kelleher, K., Pagano, M., Jellinek, M., Mclnerny, T.K., Wasserman, R., Nutting, P., and Chiappetta, L. (1999). The PSC-17: A Brief Pediatric Symptoms Checklist with Psychosocial Problem Subscales. A Report from PROS and ASPN. *Ambulatory Child Health, 5,* 225-236.

Gardner, W., Kelleher, K.J., Pajer, K.A., and Campo, J.V. (2003). Primary Care Clinicians' Use of Standardized Tools to Assess Child Psychosocial Problems. *Ambulatory Pediatrics, 3,* 191-195.

Gardner, W., Kelleher, K.J., Pajer, K.A., and Campo, J.V. (2004). Primary Care Clinicians' Use of Standardized Psychiatric Diagnoses. *Child: Care, Health and Development, 30,* 401-412.

Goodman, R. (1994). A Modified Version of the Rutter Parent Questionnaire Including Extra Items on Children's Strengths: A Research Note. *Journal of Child Psychology and Psychiatry, 35, 1483*-1494.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, 38,* 581-586.

Goodman, R. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and consequent Burden. *Journal of Child Psychology and Psychiatry, 40,* 791-799.

Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40, 1337*-1345.

Goodman, R., Ford, T., Simmons, H., Gatward, R., and Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to Screen for Child Psychiatric Disorders in a Community Sample. *British Journal of Psychiatry, 177,* 534-539.

Goodman, R., Meltzer, H., and Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A Pilot Study on the Validity of the Self-Report Version. *European Society of Child & Adolescent Psychiatry, 7,* 125-130.

Goodman, R., Renfrew, D., and Mullick, M. (2000). Predicting Type of Psychiatric Disorder from Strengths and Difficulties Questionnaire (SDQ) scores in Child Mental Health Clinics in London and Dhaka. *European Society of Child & Adolescent Psychiatry, 9,* 129-134.

Goodman, R., and Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is Small Beautiful? *Journal of Abnormal Child Psychology, 27,* 17-24.

Grayson, P., and Carlson, G.A. (1991). The Utility of a DSM-III-R Based Checklist in Screening Child Psychiatric Patients. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 669-673.

Green, J.G., McLaughlin, K.A., Alegria, M., Costello, J.E., Gruber, M.J., Hoagwood, K., Leaf, P.J., Olin, S., Sampson, N.A., and Kessler, R.C. (2013). School Mental Health Resources

and Adolescent Mental Health Service Use. *Journal of the American Academy of Child & Adolescent Psychiatry, 52, 5,* 501-510.

Hartung, C.M., and Lefler, E.K. (2010). Preliminary Examination of a New Mental Health Screener in a Pediatric Sample. *Journal of Pediatric Health Care, 24,* 168-175.

Haynes, S.N., Richard, D.C.S., and Kubany, E.S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment, 7,* 238-247.

Jee, S.H., Halterman, J.S., Szilagyi, M., Conn, A., Alpert-Gillis, L., and Szilagyi, P.G. (2011). Use of a Brief Standardized Screening Instrument in a Primary Care Setting to Enhance Detection of Social-Emotional Problems among Youth in Foster Care. *Academic Pediatrics, 11,* 409-413.

Jellinek, M., Little, M., Murphy, J.M., and Pagano, M. (1995). The Pediatric Symptom Checklist: Support for a Role in a Managed Care Environment. *Archive of Pediatric and Adolescent Medicine, 149,* 740-746.

Jellinek, M.S., and Murphy, M. (1988). Screening for Psychosocial Disorders in Pediatric Practice. *American Journal of Diseases of Child, 142,* 1153-1157.

Jellinek, M.S., Murphy, M., and Burns, B.J. (1986). Brief Psychosocial Screening in Outpatient Pediatric Practice. *Journal of Pediatrics, 109,* 371-378.

Jellinek, M.S., Murphy, M.J., Robinson, J., Feins, A., Lamb, S., and Fenton, T. (1988). Pediatric Symptom Checklist: Screening School-Age Children for Psychosocial Dysfunction. *Journal of Pediatrics, 112,* 201-209.

Johnson, J.G., Harris, E.S., Spitzer, R.L., and Williams, J.B.W. (2002). The Patient Health Questionnaire for Adolescents: Validation of an Instrument for the Assessment of Mental

Disorders among Adolescent Primary Care Patients. *Journal of Adolescent Health, 30,* 196-204.

Jutte, D.P., Burgos, A., Mendoza, F., Ford, C.B., and Huffman, L.C. (2003). Use of the Pediatric Symptom Checklist in a Low-Income, Mexican American Population. *Archives of Pediatrics & Adolescent Medicine,157,* 1169-1176.

Kessler, R.C., Avenevoli, S., Costello, E.J., Green, J.G., Gruber, M.J., Heeringa, S., Merikangas, K.R., Pennell, B., Sampson, N.A., and Zaslavsky, A.M. (2009). The National Comorbidity Survey Adolescent Supplement (NCS-A): II. Overview and Design. Journal of the American *Academy of Child and Adolescent Psychiatry, 48,* 380-385.

Kessler, R.C., Berglund, P., Chiu, W.T., Demler, O., Heeringa, S., Hiripi, E., Jin, R., Pennell, B.E., Walters, E.E., Zaslavsky, A., and Zheng, H. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and Field Procedures. *International Journal of Methods in Psychiatric Research, 13,* 69-92.

Kessler, R.C., and Merikangas, K.R. (2004). The National Comorbidity Survey Replication (NCS- R): Background and Aims. *International Journal of Methods in Psychiatric Research, 13,* 60-68.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.U., and Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the united states: Results from the national comorbidity survey. *Archives of General Psychiatry, 51*(1), 8-19.

Kessler, R.C., and Ustun, T.B. (2004). The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research, 13,* 93-121.

Kessler, R.C., Chiu, W.T., Demler, O., and Walters, E.E. (2005). Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Archive of General Psychiatry, 62,* 617-627.

Kessler, R.C., Avenevoli, S., Costello, J.E., Georgiades, K., Green, J.G., Gruber, M.J., He, J.P., Koretz, D., McLaughlin, K.A., Petukhova, M., Sampson, N.A., Zaslavsky, A.M., and Merikangas, K.R. (2012). Prevalence, Persistence, and Sociodemographic Correlates of *DSM-IV* Disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Archive of General Psychiatry, 69*, 4, 372-380.

Kim-Cohen, J., Caspi, A., Moffitt, T.E., Harrington, H., Milne, B.J., and Poulton, R. (2003). Prior Juvenile Diagnoses in Adults with Mental Disorder: Developmental Follow-Back of a Prospective-Longitudinal Cohort. *Archive of General Psychiatry, 60*, 709-717.

King, M.W., and Resick, P.A. (2014). Data mining in psychological treatment research: A primer on classification and regression trees. *Journal of Consulting and Clinical Psychology*, *82*, 895-905.

Lavigne, J.V., Meyers, K.M., and Feldman, M. (2016). Systematic Review: Classification Accuracy of Behavioral Screening Measures for Use in Integrated Primary Care Settings. *Journal of Pediatric Psychology, 41,* 1091-1109.

Loh, WY., He, X., and Man, M. (2015). A regression Tree Approach to Identifying Subgroups with Differential Treatment Effects. *Statistics in Medicine, 34,* 1818-1833.

Lucas, C.P., Zhang, H., Fisher, P.W., Shaffer, D., Regier, D.A., Narrow, W.E., Bourdon, K., Dulcan, M.K., Canino, G., Rubio-Stipec, M., Lahey, B.B., and Friman, P. (2001). The DISC Predictive Scales (DPS): Efficiently Screening for Diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 443-449.

Mayne, S.L., Ross, M.E., Song, L., McCarn, B., Steffes, J., Liu, W., Margolis, B., Azuine, R., Gotlieb, E., Grundmeier, R.W., Leslie, L.K., Localio, R., Wasserman, R., and Fiks, A.G. (2016). Variations in Mental Health Diagnosis and Prescribing Across Pediatric Primary Care Practices. *Pediatrics, 137,* 1-10.

McGee, R., Prior, M., Williams, S., Smart, D., and Sanson, A. (2002). The Long-Term Significance of Teacher-Rated Hyperactivity and Reading Ability in Childhood: Findings from two Longitudinal Studies. *Journal of Child Psychology and Psychiatry, 43,* 8, 1004-1017.

Merikangas, K., Avenevoli, S., Costello, J.E., Koretz, D., and Kessler, R.C. (2009). National comorbidity survey replication adolescent supplement (NCS-A): I. Background and measures. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*, 367-379.

Merikangas, K., He, J., Burstein, M., Swanson, S., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., and 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 & Adolescent Psychiatry*, *49*, 980–989.

Merikangas, K.R., He, J., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., Georgiades, K., Heaton, L., Swanson, S., and Olfson, M. (2011). Service Utilization for Lifetime Mental Disorders in U.S. Adolescents: Results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*, 32–45.

Murphy, J.M., Arnett, H.L., Bishop, S.J., Jellinek, M.S., and Reede, J.Y. (1992). Screening for Psychosocial Dysfunction in Pediatric Practice. A Naturalistic Study of the Pediatric Symptom Checklist. *Clinical Pediatrics, 31,* 660-670.

Murphy, J.M., Reede, J.Y., Jellinek, M.S., and Bishop, S.J. (1992). Screening for Psychosocial Dysfunction in Inner-City Children: Further Validation of the Pediatric Symptom Checklist. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*, 1105–1111.

Nock, M.K., Borges, G., Bromet, E.J., Cha, C.B., Kessler, R.C., and Lee, S. (2008) Suicide and Suicidal Behavior. *Epidemiologic Reviews, 30,* 133-154.

Nock, M.K., Green, J.G., Hwang, I., McLaughlin, K.A., Sampson, N.A., Zaslavsky, A.M., and Kessler, R.C. (2013). Prevalence, Correlates, and Treatment of Lifetime Suicidal Behavior Among Adolescents. *Journal of American Medical Association of Psychiatry, 70, 3,* 300-310.

Onwuegbuzie, A., and Collins, K. (2010). An innovative method for stress and coping researchers for analyzing themes in mixed research: Introducing chi-square automatic interaction detection (CHAID). In K.M. Collins, A.J. Onwuegbuzie, & G. Jio (Eds.), *Toward a Broader Understanding of Stress and Coping: Mixed Methods Approaches* (pp. 287-301). Charlotee, NC: Information Age.

Perou, R., Bitsko, R.H., Blumberg, S.J., Pastor, P., Ghandour, R.M., Gfroerer, J.C., Hedden, S.L., Crosby, A.E., Visser, S.N., Schieve, L.A., Parks, S.E., Hall, J.E., Brody, D., Simile, C.M., Thompson, W.W., Baio, J., Avenevoli, S., Kogan, M.D., and Huang, L.N; Centers for Disease Control and Prevention (CDC). Mental health Surveillance among Children-United States, 2005-2011. *MMWR* 2013;62(Suppl; May 16, 2013): 1-35.

Polanczyk, G.V., Salum, G.A., Sugaya, L.S., Caye, A., and Rohde, L.A. (2015). Annual Research Review: A Meta-Analysis of the Worldwide Prevalence of Mental Disorders in Children and Adolescents. *Journal of Child Psychology and Psychiatry, 56,* 345-365.

Roberts, R.E., Roberts, C.R., and Xing, Y. (2007). Rates of DSM-IV Psychiatric Disorders among Adolescents in a Large Metropolitan Area. *Journal of Psychiatric Research, 41,* 959-967.

Sattler, A.F., Bentley, J.P., and Young, J. (2019). Validation of a Brief Screening Instrument for Adult Psychopathology. Professional Psychology: Research and Practice, Advance online publication. Http://dx.doi.org/10.1037/pro0000252

Sattler, A.F., Whiteside, S.P.H., Bentley, J.P., and Young, J. (2018). Development and Validation of a Brief Screening Procedure for Pediatric Obsessive-Compulsive Disorder Derived from the Spence Children's Anxiety Scale. *Journal of Obsessive-Compulsive and Related Disorders, 16,* 29-35.

Simonian, S.J., and Tarnowshi, K.J. (2001). Utility of the Pediatric Symptom Checklist for Behavioral Screening of Disadvantaged Children. *Journal of Child Psychiatry and Human Development, 31,* 269-278.

Spitzer, R.L., Kroenke, K., and Williams, J.B.W. (1999). Validation and Utility of a Self-Report Version of PRIME-MD: The PHQ Primary Care Study. *Journal of American Medical Association*, 282, 1737-1744.

Stewart, R., Tuerk, P., Metzger, I., Davidson, T., and Young, J. (2015). A Decision-Tree Approach to the Assessment of Posttraumatic Stress Disorder: Engineering Empirically Rigorous and Ecologically Valid Assessment Measures. *Psychological Services*, *13*, 1–9.

Suryavanshi, M.S., and Yang, Y. Centers for Disease Control and Prevention (CDC). Clinical and Economic Burden of Mental Disorders among Children with Chronic Physical Conditions, United States, 2008-2013. *Preventing Chronic Disease.* 2013;13: 1-10

Vogels, A.G.C., Crone, M.R., Hoekstra, F., and Reijneveld, S.A. (2009). Comparing Three Short Questionnaires to Detect Psychosocial Dysfunction among Primary School Children: A Randomized Method. *BMC Public Health, 9*, 1-12.

Walker, W.O., LaGrone, R.G., and Atkinson, A.W. (1989). Psychosocial Screening in Pediatric Practice: Identifying High-Risk Children. *Journal of Developmental and Behavioral Pediatrics, 10,* 134-138.

Wilens, T.E., Biederman, J., Brown, S., Tanguay, S., Monuteaux, M.C., Blake, C., and Spencer, T.J. (2002). Psychiatric Comorbidity and Functioning in Clinically Referred Preschool Children and School-Age Youths with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 262–268.

Wimberger, R.C., and Gregory, R.J. (1968). A Behavior Checklist for use in Child Psychiatry Clinics. *Journal of the American Academy of Child Psychiatry*, *7*, 677–681.

Wissow, L.S., Brown, J., Fothergill, K.E., Gadomski, A., Hacker, K., Salmon, P., and Zelkowitz, R. (2013). Universal Mental Health Screening in Pediatric Primary Care: A Systematic Review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 1134–1147.

Wittchen, H.U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research*, *28*, 57–84.

LIST OF APPENDICES

***Table 1***. Demographic Characteristics of the Sample

| Socio-demographics | Category Variables | *N* = 10,123 |
|---|---|---|
| Sex | Male | 4,953 |
| | Female | 5,170 |
| Age | 13 – 14 | 3,780 |
| | 15 | 1,887 |
| | 16 | 2,010 |
| | 17 – 18 | 2,356 |
| Race/Ethnicity | White | 5,634 |
| | Black | 1,953 |
| | Hispanic | 1,914 |
| | Other | 622 |
| Parent's education level | < High School | 1,684 |
| | High School | 3,081 |
| | Little College | 1,998 |
| | College (graduate) | 3,360 |
| Parent's marital status | Married or cohabiting | 4,602 |
| | Divorced/widowed | 1,009 |
| | Never married | 308 |
| | Other/unknown | 4,204 |
| Urbanicity | Metro | 4,508 |
| | Suburb | 3,304 |
| | Rural | 2,311 |

[a] Calculated poverty index ratio data were not included in this Table

***Table 2.*** CART Classification of Diagnostic Status

| Sample | Observed | Predicted | | |
|---|---|---|---|---|
| | | Absent | Present | Percent Correct |
| Training | Absent | 3007 | 1184 | 71.7% |
| | Present | 814 | 2098 | 72.0% |
| | Overall Percentage | 53.8% | 46.2% | 71.9% |
| Test | Absent | 1189 | 594 | 66.7% |
| | Present | 352 | 909 | 72.1% |
| | Overall Percentage | 50.6% | 49.4% | 68.9% |

*Note.* The split sample validation consisted of 70% ($N$ = 7,103) for the Training sample and 30% ($N$ = 3044) for Test sample. The Observed and Predicted categories contain the true negative, false negative, false positive and true positive rates. The Percent Correct category contains the specificity, sensitivity, and overall accuracy percentages.

***Table 3.*** CIDI Items Generated by CART Analysis

| Item | Description of predictor |
|---|---|
| 1 | Has this irritable episode ended or is still going on |
| 2 | During fear-sick to stomach |
| 3 | Recently had problem because of drug use |
| 4 | Avoidance of situation interfered with work/social life/relations |
| 5 | How recently did you use weight control strategy |
| 6 | In most severe fear-feel surrounding unreal/dreamlike |
| 7 | In most severe fear-fear losing control/going crazy/pass out |
| 8 | Recently had problem because of drug use |
| 9 | Avoid situations caused physical sensation in past 12 months |
| 10 | How recently had alcohol problems |
| 11 | Receive professional treatment for irritability in past year |
| 12 | Social fear situation-chills/hot flushes |
| 13 | Social fear situation-feel things unreal/dreamlike |
| 14 | Worst month in past year-still enjoyed things you used to |
| 15 | Social fear situation-feel distant, not really there |
| 16 | Worst month past year-felt emotionally distant |
| 17 | Worst month in past year-laugh/see funny side |
| 18 | Receive prof treatment for eating/weight problem in past year |
| 19 | During worst event-felt numb |
| 20 | Worst month past year-lose interest in enjoyable things |

*Note.* CIDI = Composite International Diagnostic Interview.

***Figure 1***. CART Decision-Tree Method

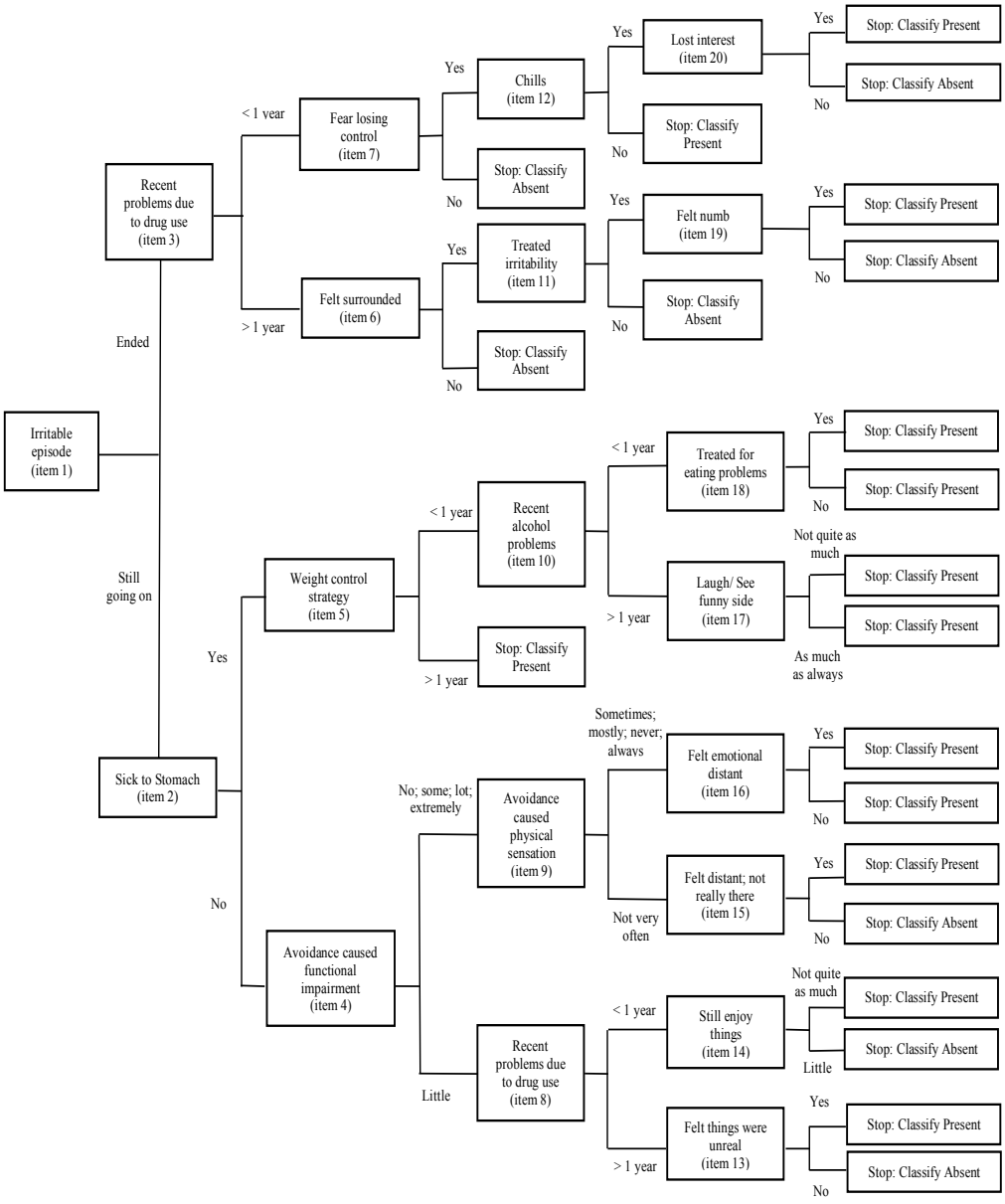


Figure 1. Decision-tree structure displays one of the 20 potential CIDI items used to determine if any particular respondent had any diagnosis of any kind

CURRICULUM VITA

# Steven M. Bishop, M.A.

CURRICULUM VITAE

## EDUCATION

**M.A.** **University of Mississippi** (Oxford, MS) December 2019
Clinical Psychology
Thesis: "Youth Screening Protocol for General Psychopathology"
Advisor: John N. Young, Ph.D.

**M.A.** **Reformed Theological Seminary** (Jackson, MS) December 2000
Marriage and Family Therapy
Advisor: Jim Hurley, Ph.D.

**B.A.** **Union University** (Jackson, TN) May 1995
Majors in Psychology and Religion-Philosophy
Minor in History

## CLINICAL EXPERIENCE

**Communicare** July 2018-present
*Therapist*
Supervisors: Dixie Church, M.A., LMFT, John Young, Ph.D., & Dani Maack, Ph.D.
Provide evidence-based supported treatment interventions to outpatient adults and children in a rural community mental health clinic with diverse presenting problems including those with serious mental illnesses.

**North Mississippi Regional Center** July 2017-June 2018
*Psychological and Behavioral Services Intern*
Supervisors: Giovanni Biffle, M.A., BCBA & Kim Sallis, Ph.D.
Provided individual interventions for inpatient clients with intellectual and developmental disabilities including Prader-Willi Syndrome and Lesch-Nyhan Syndrome. Planned and implemented individually tailored interventions using applied behavior analysis. Conducted formal full-battery assessments with children and adults suspected of having developmental, intellectual, and behavioral disorders.

**University of Mississippi Psychological Services Center** August 2017-present
*Graduate Therapist*
Supervisors: Stefan Schulenberg, Ph.D., John Young, Ph.D., Dani Maack, Ph.D.
Provided individual cognitive-behavioral therapy (CBT) and evidence-based treatments (EBTs) for university and community outpatient adults and children with DSM-5 disorders.

## RESEARCH EXPERIENCE

**Graduate Research Assistant, University of Mississippi** 2016-present
Scientific Infusion That Helps (SITH) Laboratory
Advisor: John N. Young, Ph.D.

## PROFESSIONAL EXPERIENCE

**Therapist** February 2014-July 2016
**Canopy Children's Solutions, formerly Mississippi Children's Home Services**
Provided evidence-based interventions for community outpatient children and their families with a broad range of DSM-5 disorders including those with serious emotional disturbance.

**Clinical Coordinator** September 2010-February 2014
**Three Oaks Behavior Health, LLC**
Provided interdisciplinary treatment planning for outpatient adults in a partial hospital setting with serious mental illnesses including individual and group psychotherapy, supervision of sessions, staff in-services, and direct support to Program Director.

## LEADERSHIP EXPERIENCE & SERVICE

| | |
|---|---|
| **Mississippi Association for Marriage & Family Therapists** | 2003-2014 |
| Past-President | 2012-2014 |
| President | 2010-2012 |
| President-Elect | 2008-2010 |
| Chairman, Legislative Committee | 2003-2008 |

## RELEVANT COURSEWORK

Advanced Evidence Based Treatments Seminar Fall 2019
Instructor: John Young, Ph.D.

Couples' Therapy Seminar — Fall 2018
Instructor: Dani Maack, Ph.D.

Evidence Based Treatments Seminar — Fall 2017
Instructor: John Young, Ph.D.

**REFERENCES**

References available upon request.