1-1-2019

Validation of a brief screening instrument for psychopathology in adults

Adam Sattler

VALIDATION OF A BRIEF SCREENING INSTRUMENT FOR PSYCHOPATHOLOGY IN ADULTS

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
The University of Mississippi

by

ADAM F. SATTLER

May 2019

ABSTRACT

Structured diagnostic interviews are generally considered to be the gold standard method of assessing mental health conditions; however, clinicians rarely use such techniques in clinical practice. A major factor contributing to the infrequent use of structured interviews is the amount of time required for administration and interpretation. As such, there is a need to develop assessment instruments that balance the objectivity and structure of evidence-based assessment techniques against personnel time and cost. To address this need, the current study sought to validate a brief screening instrument for assessing 9 common mental health conditions using a well-researched structured clinical interview to establish criterion diagnoses. Data were collected in the emergency department at University of Mississippi Medical Center in Jackson, Mississippi. Participants included 259 adult patients between the ages of 18 and 91 (M = 47.5). One hundred thirty-seven (52.9%) of the patients were women; 149 (57.5%) were African American, 109 (42.1%) were White, and 1(.4%) was Hispanic. Estimates for sensitivity, specificity, negative predictive value, positive predictive value and kappa generally supported the validity of the screening instrument for assessing mental health conditions within the context of an emergency department. The screening instrument may thus represent a promising solution to conducting efficient and accurate mental health assessment in a variety of applied contexts.

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor, Dr. John Young, and committee members, Drs. Todd Smitherman, John Bentley, and Danielle Maack. I could not have completed this dissertation project without their support and guidance.

I would also like to thank Brooklee Tynes, Lauren Weathers, and Elise Elligett for assisting with data collection, as well as the SITH/ADEPT lab members for their feedback and support in completing this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Accurate assessment is the cornerstone to the application of effective treatment techniques. An appropriate, evidence-based treatment cannot be selected and applied if the condition being treated is not accurately defined (Chorpita, Brown, & Barlow, 1998). Fortunately, various instruments have been developed to conduct broadband assessment of psychopathology in adults, including self-report measures (Kroenke, Spitzer, & Williams, 2001) and structured diagnostic interviews (Sheenan et al., 1998). In particular, structured interviews carry a number of advantages and are thus considered the gold standard in clinical assessment (Dawes, Faust & Meehl, 1989; Ehlert, 2007; Grove & Meehl, 1996; Joiner et al., 2005; Meehl, 1954; Schneider & Döpfner, 2004; Silverman & Ollendick, 2005). Structured interviews yield objective data (i.e., have evidence to support their reliability and validity) and help clarify, facilitate, and standardize the diagnostic process. This is especially salient in an era where classification manuals are becoming increasingly complex. Moreover, structured interviews provide a systematic method to assess a given client's symptoms very broadly, which enables evaluation of not only colloquially defined presenting problems but a full range of disorders and potential comorbidities (Basco et al., 2000; Zimmerman & Mattia, 1999).

The vast majority of practitioners in clinical settings, however, do not employ these assessment strategies (Garland, Kruse & Aarons, 2003; Hatfield & Ogles, 2004; Sattler et al., 2016; Whiteside, Sattler, Hathaway & Douglas, 2016). Rather, practicing clinicians rely on intuition, idiographic approaches, or other non-standardized, poorly supported methods of discerning a given individual's symptoms (Garb, 1998; Hatfield & Ogles, 2004). Unsurprisingly, these approaches are not known to be particularly reliable, and thus do not result in optimal conceptualization to guide treatment (Dawes et al., 1989; Grove & Meehl, 1996, Meehl, 1954). Related to infrequent use of structured interviews among clinicians are issues concerning the burden associated with implementing such procedures. For example, gold standard interviews like the World Health Organization's Composite International Diagnostic Interview (CIDI; Andrews & Peters, 1998) can take 2 or more hours to administer, which is a luxury most clinical settings do not have (Garland, Kruse, Aarons, 2003). Thus, there is a need to develop clinically feasible, standardized assessment procedures that demonstrate efficiency without sacrificing the quality of the diagnostic information that is produced.

Evidence-Based Assessment

The rise of the evidence-based practice (EBP) movement has become arguably the most influential paradigm within the field of mental health in the past two decades. Formally defined by the American Psychological Association Division 12 Task Force in 2006 (Anderson, 2006), the EBP model contains two complementary branches: evidence-based treatment (EBT) and evidence-based assessment (EBA). Until recently, however, the literature concerning EBP focused primarily on treatment applications and placed little emphasis on EBA. This is unfortunate, particularly because the successful delivery of EBTs depends so heavily on the implementation of assessment practices that yield accurate diagnoses (i.e., EBA; Chorpita, Brown, & Barlow, 1998). Moreover, lack of emphasis on EBA may make it difficult for clinicians to communicate with one another and/or understand and make use of research findings in their clinical work (Jensen & Weisz, 2002). Fortunately, the EBA literature has undergone considerable expansion in recent years (e.g., Hunsley & Mash, 2007; Mash & Hunsley, 2005).

Hunsley and Mash (2007) have been among the most influential researchers in terms of defining and describing the EBA approach. According to Hunsley and Mash, the EBA approach comprises three primary processes: psychiatric screening, diagnosis, and outcome measurement. Each process is supported by well-validated EBA instruments, which consist primarily of various structured interviews and self-report measures. The appropriate application of these instruments depends on the specific construct of interest and the purpose of the assessment. For example, within the context of an initial diagnostic evaluation, the EBA approach necessitates the use of an appropriately normed diagnostic instrument (i.e., a structured interview) as opposed to a tool validated for some other purpose (e.g., a rating scale designed to measure treatment outcomes). Of the three EBA processes (psychiatric screening, diagnosis, outcome measurement), practicing clinicians tend to consider diagnosis a particularly important component of assessment (Camara, Nathan, & Puente, 2000). Despite this importance, however, the diagnostic process has consistently been demonstrated to be unreliable in typical practice settings.

The issue concerning the reliability of unstructured psychiatric diagnoses has been the subject of examination for decades (e.g., Meehl, 1954). For example, in their seminal paper Dawes, Faust, and Meehl (1989) reviewed evidence from over 100 studies comparing the accuracy of clinical judgment (i.e., human intuition applied in the absence of structured or supported diagnostic practices) to that of actuarial (i.e., mathematical, rule-based) decisions and found that the latter demonstrated superiority or equivalence in nearly all cases. These findings were true even in instances where they were compared to each other (i.e., purely intuitive vs. purely actuarial) as well as when adapted hybrid models were examined (i.e., purely actuarial vs.

a combination of intuitive and actuarial approaches), thus demonstrating the importance of mathematical, rule-based methods in the process of psychiatric assessment.

The authors also described the factors underlying the superiority of actuarial methods by explaining that human judgment is subject to limitations such as fatigue, the bias introduced by recent experience, and random fluctuations in observed phenomena. More importantly, actuarial methods guarantee that the variables included in the decision making process contribute to conclusions according to their predictive power and their relationship with the criterion of interest (whereas clinical judgment processes frequently rely on invalid variables). For example, decisions based on multiple regression equations are based entirely on predictive variables (non-predictive variables are excluded), which are weighted according to their unique contribution to an accurate outcome.

Dawes et al. (1989), also describe the issue pertaining to practicing clinicians' limited access to information regarding the accuracy of their decisions. This lack of feedback hinders clinicians' ability to recognize and correct diagnostic errors, ultimately perpetuating inaccurate beliefs about the diagnostic process in general as well as the criteria underlying specific disorders. These limitations and observable errors in diagnostic decision-making were cited as foundational to the need to develop and apply actuarial methods of clinical assessment.

Structured interviews represent one contemporary solution to implementing actuarial methodology in psychiatric assessment. These tools generally contain specific, closed-ended questions (i.e., encourage responses of "yes"/"no") that are closely linked to diagnostic criteria. Given such standardization, structured interviews help to eliminate many aforementioned sources of bias during the diagnostic process. For example, administering the same set of questions for every patient encourages clinicians to assess *all* the relevant symptoms of a disorder, not simply those conferred by bias or anecdotal experience as most reflective of diagnosis. This more thorough process, when correctly applied, reduces the likelihood that initial clinical hypotheses (otherwise termed "guesses") will bias diagnostic outcome (Croskerry, 2003). Similarly, structured interviews guide clinicians in obtaining valid diagnostic information as diagnoses are made according to valid symptomatic criteria that are weighted appropriately in terms of importance, as opposed to clinicians' idiosyncratic understanding of disorders. Standardized structured interviews also reduce clinicians' tendency to perceive psychopathology over normal behavior and minimize clinician-held stereotypes based on ethnicity, age, and/or gender (Garb, 1998). Because the delivery of closed-ended questions is systematic, the information that is yielded via patient response is also more likely to be interpreted in a standardized manner (Andrews & Peters, 1998).

In addition to the benefits of standardization, structured interviews contain algorithms that guide clinicians through the assessment process so that they do not have to rely on their memory of complex diagnostic rules (Di Nardo, Brown, & Barlow, 1994). Structured interviews thus help the clinician combine information to identify superordinate diagnoses (yielding a more accurate and/or parsimonious clinical conceptualization), as well as symptoms that are accounted for by medical conditions or drug use (Andrews & Peters, 1998). Related to this point, structured interviews typically cover a broad range of diagnoses, thus aiding the clinician in identify comorbid disorders (Hunsley & Mash, 2007). Finally, because structured interviews facilitate documentation of symptoms, clinicians are able to refer back to information that has been collected and determine when/where the diagnostic process broke down when errors are made (Andrews & Peters, 1998).

An ever-expanding number of structured interviews have been (and continue to be) developed, with well-established versions of each demonstrating high levels of diagnostic reliability and validity (e.g., Di Nardo, Brown, & Barlow, 1994; First, Gibbon, Hilsenroth, & Segal, 2004; Kaufman et al., 1997; Sheehan et al., 1998; Silverman & Albano, 1996). As a result, these instruments yield superior diagnostic accuracy when compared to unstructured interview approaches (e.g., Basco et al., 2000; Denys, van Megan, & Westenberg, 2003; Ghanizadeh, Mohammadi, & Yazdanshenas, 2006; Jensen & Weisz, 2002; Jensen-Doss, Youngstrom, Youngstrom, Feeny & Findling, 2014; Jewell, Handwerk, Almquist, & Lucas, 2004; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). Structured interviews therefore represent an essential component of effective clinical practice, particularly in an era where evidence-based treatments are designed for use with specific disorders and symptom clusters. Unless standardized procedures such as structured interviews are used, it is unlikely that treatment manuals will be selected and implemented in an accurate fashion (Jensen-Doss & Weisz, 2008; Pogge et al., 2001; Weisz & Addis, 2006).

Despite the well-known shortcomings of clinical judgment and the availability of measurably superior strategies in structured interviews, such tools continue to demonstrate low rates of implementation in clinical practice. For example, Bruchmuller, Katrin, Margraf & Schneider (2012) surveyed clinicians about their use of structured interviews and found that less than half reported ever using them. Further, these results demonstrated that clinicians who did report any use of structured formats indicated doing so with only 14.8% of patients. These observed low rates of use appear to be widespread, as studies have demonstrated similar patterns of limited EBA among clinicians in numerous contexts (e.g., Anderson & Paulosky, 2004;

Belter, & Keller, 1998; Cashel, 2002; Garland, Kruse & Aarons, 2003; Hatfield & Ogles, 2004; Piotrowski, Gilbody, House, & Sheldon, 2002).

Although there are numerous factors that contribute to low rates of EBA (Bruchmuller et al., 2012; Garland et al., 2003; Gilbody et al., 2002; Hatfield & Ogles, 2007; Palmiter, 2004), a lack of tangible access to the resources necessary for implementation has been posited as a key impediment (Jensen-Doss & Hawley, 2010). For example, many structured interviews can take over two hours to administer and require advanced training to deliver proficiently (e.g., Kessler & Ustun, 2004; Spitzer, Williams, Gibbon & First, 1992). Unfortunately, such practices are likely to be prohibitively resource intensive as clinical settings typically contend with high workload, poor financial compensation, limited time, and intense demand for resources (Nunno, 2006). This issue has become particularly relevant as providers have come under increasing pressure from third-party payers (e.g., insurance companies, managed care providers) to enhance the cost-effectiveness of psychotherapy (Chorpita, Yim, & Tracey, 2002; Christensen & Jacobson, 1994; Richardson & Austad, 1991). Issues concerning the feasibility of EBA in clinical practice are therefore frequently cited as among the primary reasons clinicians do not implement such tools (Jensen-Doss & Hawley, 2010). Given the pressure to deliver effective services in a timely manner, there is an apparent need to develop clinically feasible EBA practices that reduce the cost, time, and resource requirements associated with implementation (Bumbarger & Campbell, 2012; Beidas et al., 2014).

Current solutions aimed at reducing the burden of EBA

Self-report measures

One potential method of reducing assessment burden is through the use of screening instruments (e.g., ratings scales, self-report measures) to guide subsequent interview administration. In comparison to structured interviews, self-report measures are generally brief and concise, and thus place relatively little burden on clinicians and patients (Aboraya, Rankin, France, El-Missiry & John, 2006; Beidas, Cross, & Dorsey, 2014). Such instruments can also be distributed to multiple informants (e.g., within the context of child assessment: parents, teachers, and the patient) to obtain a broad, yet detailed understanding of the patient's functioning (Velting, Setzer & Albano, 2004; Silverman et al., 2005). In addition, numerous instruments have been developed for a wide range problems and purposes (e.g., symptoms screening, outcome tracking), many of which are available free of charge (Beidas et al., 2014). In terms of limitations, self-report measures include the respondents' idiosyncratic interpretations of questions, the inability to clarify misinterpretations, as well as the inability to assess the validity of the information that is provided (Smith, Klein, & Benjamin, 2003). Overall, however, self-report measures represent an integral component of the EBA approach.

Guidelines for incorporating EBA into clinical practice recommend leveraging and combining the strengths of self-report measures and structured interviews by using results from the former to guide subsequent administration of the latter (e.g., Christon, McLeod, & Jensen-Doss, 2015). In other words, a clinician might administer a broad self-report measure to identify main problem areas, which are then assessed further via specific structured interview modules (as opposed to the entire interview). This gated process thus gains more detail where it is actuarially warranted and simultaneously reduces the burden associated with administration of the entire clinical interview. Specific applications of this approach have been described for pediatric bipolar disorder (Youngstrom, Jenkins, Jensen-Doss & Youngstrom, 2012), attention deficit disorder (Frazier & Youngstrom, 2006), and antisocial personality disorder (Guy, Poythress, Douglas, Skeem, & Edens, 2008). Chorpita & Nakamura (2008) also described a similar approach that utilizes an innovative, dynamic strategy for incorporating data from various self-report measures and a subsequent structured interview to diagnose a range of psychopathology in children. Overall, results demonstrated that this approach reduced administration time and yielded improved accuracy over standard procedures. However, this study represented a preliminary feasibility demonstration and its application in real-world practice has yet to be seen. As such, the studies cited above describe applications of EBA screening procedures in well-resourced research settings, which are not representative of the typical clinical practice setting (as outlined above).

Despite the availability and utility of self-report measures, rates of use among clinicians in practice settings are similar to those of structured interviews in that they are rarely implemented (Bickman et al., 2003; Camara et al., 2000; Cashel, 2002; Garland et al., 2003; Hatfield & Ogles, 2007; Jensen-Doss & Hawley, 2010; Lueger et al., 2001). Also similar to the implementation of structured interviews, the factors contributing to limited rates of use are varied and complex. For example, Garland et al (2003) explored barriers to the use of standardized outcome measures and found that clinicians frequently cite concerns regarding the clinical feasibility (e.g., burden of paperwork and scoring/interpretation), relevance (e.g., applying nomothetic measurement to unique individuals), and incremental validity of such practices. A survey conducted by Hatfield and Ogles (2007) also reflected these results by demonstrating clinicians' negative beliefs regarding the practicality and utility of outcome measurement. Similarly, Gilbody et al (2002) found that reasons for not implementing standardized assessment practices include beliefs that such measures do not adequately capture clinical problems, as well as doubts concerning the psychometric properties and practicality of such measures. Finally, Jensen-Doss and Hawley (2010) surveyed 1442 therapists about their use of EBA and found that concerns regarding the clinical feasibility of such techniques were the strongest and only independent predictor of use.

In sum, these studies describe a number of potential barriers to implementing self-report measures, thus demonstrating the complexity of the issues surrounding EBA dissemination and implementation. However, this review of the literature also reveals that clinicians' concerns regarding the practicality of such techniques are among the most consistent and frequently cited issues. Importantly, these concerns persist despite use of self-report measures being a relatively low burden EBA technique. The consistency with which concerns of practicality continue to be voiced thus highlights the need for further research in conducting efficient diagnostic screening (e.g., Kahana, Youngstrom, Findling, & Calabrese, 2003; Warnick, Weersing, Scahill, & Woolston, 2009). For example, given specific concerns regarding excessive paperwork and scoring/interpretation difficulties (e.g., Garland et al., 2003), methods aimed at reducing such sources of burden may enhance the implementation potential of self-report measures in particular and in EBA. Specifically, implementing EBA within computerized (i.e., automatic) formats may represent a promising solution to achieving widespread clinical use.

Computerized Assessment

Computerized administration formats may facilitate greater use of EBA in clinical settings because they tend to promote significant savings in clinician time (Ebesutani et al., 2012; Garb, 2007). Moreover, multiple studies have demonstrated good reliability and validity for computer-administered interviews more generally (Lewis, 1994; Jewell, Handwerk, Almquist, & Lucas, 2004; Reilly-Harrington, et al., 2010), as well as the patient acceptability of such methods (Bachman, 2003; Dignon, 1996; Hoyer, Ruhl, Scholz, & Wittchen, 2006; Petrie, & Abell, 1994; Rosenman, Levings, & Kosten, 1997; Shakeshaft, Bowman, & Sanson-Fisher, 1998).

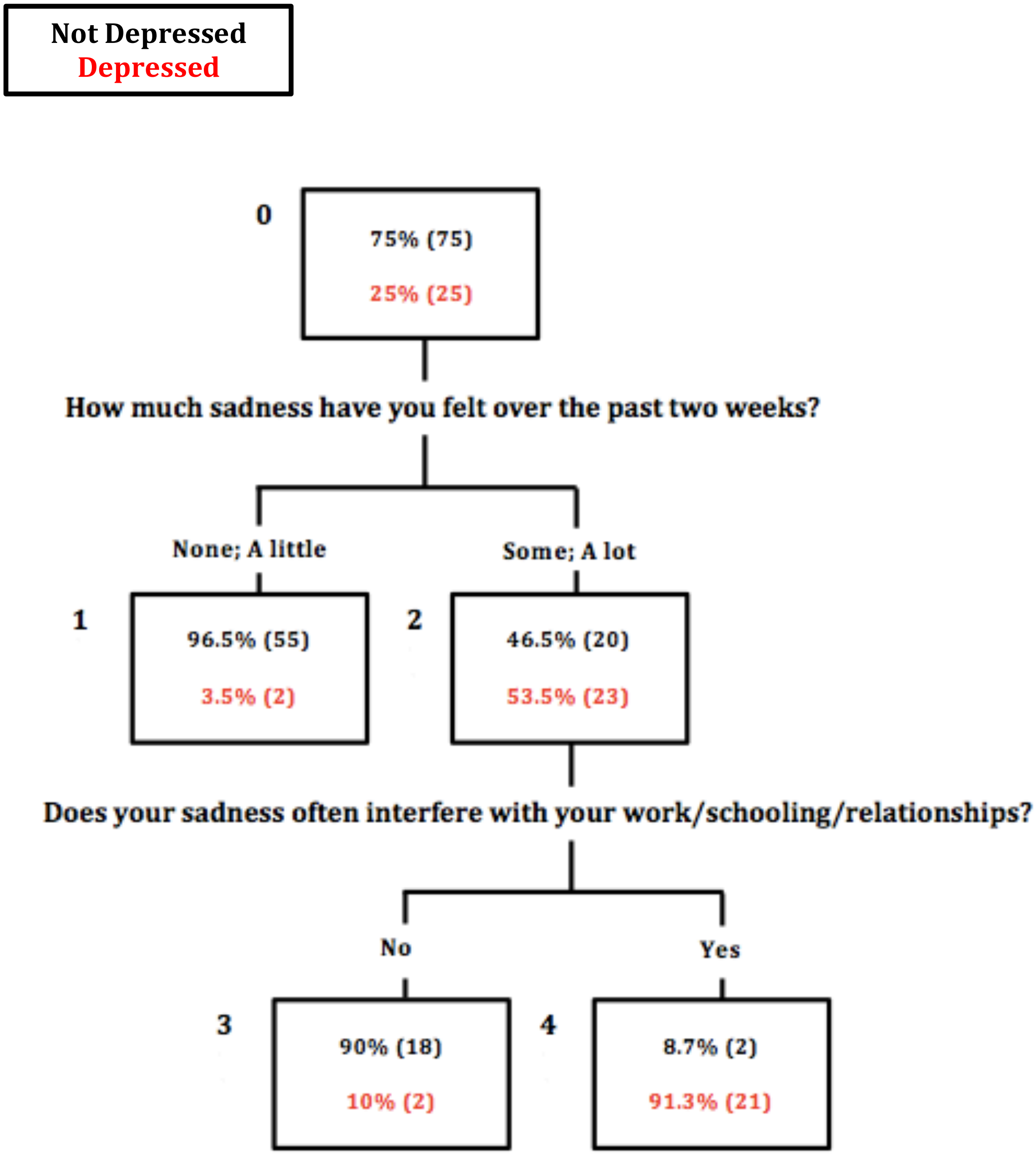
Computerized administration also eliminates the need to record and score data manually, thereby negating data entry and scoring errors. In addition, the benefits of computerized administration may be particularly apparent if the results of the assessment can be displayed graphically in a format that is easy for clinicians to interpret (Chorpita, Bernstein & Daleiden, 2008). Finally, computerization enables novel, efficiency-focused approaches to assessment that are both computationally demanding and dynamic, and consequently difficult/impossible to implement within a paper-pencil format. As such, computerized formats represent a solution to improving the clinical feasibility of EBA, particularly when combined with the convenience inherent in self-report formats and when they are used to enact novel and efficient means of assessment.

Decision-tree analysis

Decision-tree analysis offers a method to integrate the advantages of EBA reviewed above in an efficient manner. Specifically, it is a procedure that can be used to develop predictive models (i.e., algorithms) that dramatically reduce the amount of information needed to achieve accurate prediction of an outcome (in this case diagnosis) without sacrificing accuracy. The analysis employs statistical procedures (typically implemented via computer software) designed to examine large sets of variables and identify smaller subsets of variables to determine the ways they interact in predicting a dichotomous outcome (e.g., depressed/not depressed) as efficiently and accurately as possible. Importantly, the process is completely automated; meaning a priori knowledge of the complex relationships among variables is not required to run the analysis successfully. The analysis may also be particularly useful in applied contexts (e.g., clinical decision-making situations) as the yielded algorithms can be depicted graphically within a decision-tree format that is relatively straightforward and easy to implement.

What follows is a hypothetical application of decision-tree analysis to illustrate the process of algorithm development. In this example, a researcher is working to reduce the burden associated with assessment by developing a brief decision-tree algorithm for diagnosing depression. Prior to developing the algorithm, the researcher uses gold-standard methodology (including a structured interview and multiple self-report measures) to assess 100 individuals and determines that 25 meet diagnostic criteria for depression. This process produces a large amount of clinical data for each individual (e.g., their emotional state, sleeping patterns, appetite, etc.), which provides the original set of information that will be subject to reduction via decision-tree analysis. Ultimately, the researcher will use decision-tree analysis to identify the most important pieces of clinical information (thus omitting less meaningful information), and examine how this information can be combined to diagnose depression as efficiently and accurately as possible. The resulting algorithm is presented in Figure 1 below.

Figure 1. Example decision-tree structure



A decision-tree algorithm is typically depicted using a flowchart-like diagram similar to Figure 1. Essentially, the decision tree is represented as a series of *nodes* (pictured above as black boxes) depicting various samples with specific proportions of depressed and non-depressed individuals. Nodes consist of two types: *parent nodes* (i.e., nodes that are "split" to create two sub-nodes) and *child nodes* (i.e., the sub-nodes that result from the splitting of a single parent node). Note that a single node can be a parent and child node simultaneously (e.g., node 2 above). The various nodes in a decision tree are connected by "branches" (the black lines running between each of the boxes), which represent decisional options leading to differential nodal outcomes. *Terminal nodes* represent the end of a decision sequence, where the within-node group majority (whichever group holds a node proportion >50%) determines the classification of all individuals contained within that node. The node at the top of the tree structure is referred to as the *root node* (node 0), which depicts the proportions of the original sample of interest prior to splitting.

As previously mentioned, the goal of decision-tree analysis is to develop an algorithm that classifies the sample as accurately and efficiently as possible. In accomplishing this goal, a statistical procedure examines the sample of depressed and non-depressed individuals and analyzes ways in which the sample might be split to produce optimally homogenous subsamples. To do this, the process employs statistical tests to examine each piece of diagnostic information to identify a single variable containing a "split point" that can be used to divide the clinical sample into subsamples that contain minimal variability in regard to depressed and non-depressed individuals. A split point is any point on a particular variable that can be used to divide a sample (i.e., node) into subsamples (i.e., sub nodes) depending on the measurements of the sample with respect to that node. Information regarding the specific statistical tests that are used to identify predictor variables and split points are discussed in greater detail in the results section below.

In the current example, information concerning level of sadness over the previous two weeks is identified as the optimal splitting variable from among all the other pieces of diagnostic information. Specifically, of individuals who responded "none" or "a little" to the item "How much sadness have you felt over the past two weeks?," just 3.5% ended up meeting criteria for depression. This subsample of individuals is depicted in terminal node 1, which is termed non-depressed given the low base rate of misclassification.

Alternatively, of the individuals providing a response of "some" or "a lot" to the first item, 53.5% met criteria for depression (represented in node 2). Node 2 is therefore much less homogenous than node 1, and a classification decision made here would result in a much higher degree of error (i.e., classifying all individuals in node 2 as depressed would result in 20 misclassified non-depressed individuals). Given that within-node homogeneity is low, the statistical procedure re-examines all of the clinical information input to identify another variable that produces a locally optimal split at node 2. This re-examination determines that information regarding impairment due to sadness maximizes discrimination between depressed and non-depressed individuals within the subsample in node 2. Specifically, of the individuals who indicated "some" or "a lot" of sadness over the past two weeks *and* provided a response of "yes" to the question "Does your sadness often interfere with your work/schooling/relationships?" (node 4), 91.3% met criteria for depression. Due to the high probability of a depression diagnosis, all individuals in terminal node 4 are classified as meeting criteria for the disorder. However, of the individuals who indicated "no" interference, just 10% met criteria for depression (node 3). Hence, all individuals in terminal node 3 would be classified as non-depressed. The overall model, when interpreted as a whole, results in a highly efficient and accurate method of assessment (i.e., 6% error rate).

Briefly, there are three primary approaches to developing decision-tree algorithms like the one described above. The oldest approach, Classification and Regression Trees (CART), recursively splits samples in a binary fashion, thus producing a *pair* of child nodes each time a parent node is split. The resultant decision-tree is consequently binary in that each split is limited to two outcomes, thus producing two decisional branches per split (like the algorithm produced in the example above). The second approach to developing decision trees is referred to as Chi-Square Automatic Interaction Detector (CHAID). CHAID is conceptually similar to its predecessor; however, unlike CART, CHAID can produce non-binary (i.e., multi-way) splits, thus resulting in decision-tree structures that contain numerous outcomes per split. For example, splitting a single parent node could produce four branches leading to four separate child nodes (note that there is no particular advantage offered through multi-way splits). Finally, the most recently developed approach, Quick, Unbiased Estimation Statistic (QUEST), is similar to the two previous methods. Its primary advantage compared to CART and CHAID, however, is that it does not conduct an exhaustive search of potential split points on predictor variables (thereby reducing computation time). Moreover, unlike the other two approaches, which tend to favor the selection of variables with greater numbers of potential split points, QUEST is unbiased in terms of its selection of predictor variables. It is worth mentioning, however, that no approach is generally superior to the others – each has its strengths and weaknesses, depending on the application.

Breiman and colleagues were the first to describe decision tree methods in 1984 (Breiman, Friedman, Stone, & Olshen, 1984). The analysis is an example of an approach to machine learning, which is a subfield in computer science. Briefly, machine learning specializes in the creation of algorithms that "learn" from existing data to make predictions about future events. Specifically, these algorithms function by constructing models from existing input data, which then form the basis for data-driven predictions and decisions about outcomes that are not yet known. Since its inception, decision-tree methodology has since been applied in diverse areas of research, including forensics (Appavu & Rajaram, 2008), chemistry (Chen, Rusinko, & Young, 1998), and astronomy (Owens, Griffiths, & Ratnatunga, 1996). However, the best-documented and arguably most popular applications are described within biomedical research, where classification of data points (e.g., affected versus healthy patients) is a central issue (Bris, Majernik, Pancerz, & Zaitseva, 2015). Given the relevance that the task of classifying individuals according to their (mental) health has within EBA, the applicability of decision-tree methods within clinical psychology is perhaps easy to recognize.

Despite the potential the methodology offers in terms developing efficient assessment instruments, decision-tree analysis has rarely been used for such purposes. However, an exhaustive search of the relevant literature did reveal applications of decision-tree methods in identifying risk factors for various problem areas. In general, these studies leverage decision-tree methods to conduct exploratory analyses and (automatically) identify non-obvious relationships among predictors in terms of accurately classifying individuals according to a specific outcome. For example, such methods have been used to explore data and determine risk factors for depression over the adult life span (Batterham et al., 2009), as well as risk factors for depression specifically in older adults (Schoevers et al., 2006; Smits et al., 2008). Similar research has also identified non-obvious relationships among temperamental variables such as neuroticism and self-esteem in predicting depression (Schmitz, Kugler, & Rollnik, 2003), as well as the importance of depression in determining quality of life in multiple sclerosis patients (D'Alisa et al., 2006). Decision-tree methods have also been used to examine risk factors in the development of suicidal thoughts and behaviors in general (Batterham et al., 2012), as well as within the context of substance use disorders (Buri et al., 2009; Tiet et al., 2006) and depression (Iigen et al., 2009; Mann et al., 2008). Finally, the analysis has been used to develop models for predicting violent reoffending behavior in prison populations (Liu et al., 2011; Stalans, Yarnold, Seng, Olson, & Repp, 2004) as well as individuals with psychotic illnesses (Thomas, 2015). In sum, these studies demonstrate the applicability of decision-tree methodology to psychiatrically relevant research questions, particularly in terms of identifying hidden risk factors to better understand and classify behavioral problems.

In terms of applications more directly related to the assessment of psychological constructs, decision-tree methods have been utilized in the development of a computer-based research instrument, referred to as the Copernicus system (Jachyra, Pancerz, & Gomula, 2013). The system, which is based off the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway, McKinley & MMPI Restandardization Committee, 1989), is designed to help psychological researchers assess study participants according to their MMPI profiles. Copernicus applies a "knowledge base" of previously developed algorithms (based on decision-tree methodology as well as other machine learning approaches) to incoming MMPI data. The system thus provides multiple classification outcomes for individuals according to various subscales (e.g., criminality, psychoticism, sociopathy) and depicts these results in an output, which researchers can use to assess study participants.

Although the Copernicus system represents a novel approach to analyzing MMPI data, it is primarily a research instrument and is thus not likely to be adopted by clinicians. For example, the system requires exhaustive administration of the MMPI (which contains 567 items and can take up to 1.5 hours to administer), thereby limiting its clinical feasibility in terms of requisite resources. In addition, although the personality classes represented in the Copernicus profile correspond to the MMPI, most of these classes do not clearly map onto diagnoses in the Diagnostic and Statistical Manual (DSM; American Psychiatric Association, 2013), thereby limiting its potential as a clinical diagnostic tool. Finally, although Copernicus does appear to be relatively user-friendly for researchers, the vast number of functions it offers in terms of classification methods and indexes (among many other esoteric analytic options) are not likely to be useful (and potentially overwhelming) to clinicians. The Copernicus system thus represents an application of decision-tree methodology that enables researchers to examine MMPI profiles in novel ways, potentially yielding greater theoretical understanding of certain psychological constructs. However, the system (as well as the standard MMPI instrument itself) lacks clinical utility and feasibility, and is thus unlikely to be adopted for clinical use.

Further review of the literature revealed just two studies describing implementation of decision-tree methodology for the specific purpose of developing clinically feasible EBA instruments. Ebesutani et al. (2012) published the first of these. In their study, the authors used decision-tree analysis to integrate information from two child and two parent self-report measures in the development of an assessment protocol to predict whether or not treatment was warranted (for anxiety, depression, attention-deficit/hyperactivity disorder (ADHD), and disruptive behavior problems). Three decision-tree structures were developed to address questions concerning whether treatment was warranted, what general type of treatment (internalizing or externalizing) was needed, and which disorder-specific protocol should be used. The algorithm-based approach was compared with criterion decisions established by expert assessors who integrated information from parent and child structured interviews as well as self-report measures to determine the need for treatment. In terms of determining whether treatment was needed, the algorithm-based approach demonstrated excellent performance (94% accuracy rate). In addition, the algorithm-based approach achieved good accuracy (83%) in terms of identifying appropriate internalizing versus externalizing treatment. Finally, the algorithm yielded fair accuracy (79%) in identifying disorder-specific treatment. Ebesutani et al (2012) thus demonstrated the feasibility of developing low burden, accurate, efficient decision-making protocols from more intensive and costly procedures. In doing so, the researchers designed a method that not only reduced assessment burden, but also illuminated important factors that predict need for treatment in general as well as need for specific types of treatment.

Stewart et al. (2015) also applied decision-tree analysis for the specific purpose of developing a clinically feasible assessment instrument. Their effort involved an application of decision-tree analysis in the development of an abbreviated assessment protocol based on the Clinician Administered PTSD Scale (Blake et al., 1995). In its original form, the CAPS is a 30-item structured interview that demonstrates excellent psychometric properties for diagnosing PTSD in a variety of populations (Weathers, Keane, & Davidson, 2001). Stewart et al (2015) successfully reduced the number of CAPS items administered by over 75% while retaining 92% overall diagnostic accuracy. Thus, decision-tree analysis enables the omission of items that do not optimally differentiate between individuals with and without psychiatric diagnoses, thereby drastically reducing the time required for assessment and potentially increasing the likelihood of use in applied clinical activities.

Present Study

The findings in the literature reviewed converge to demonstrate the need for efficient, evidence-based diagnostic instruments. On the basis of this research, it also appears that self-report instruments, computerized administration, and decision-tree analysis can be logically combined to formulate a parsimonious strategy to meet this need. Thus far, however, application of such methodology has been restricted to the development of diagnostic algorithms for relatively circumscribed problem areas (i.e., various categories of child psychopathology and

PTSD in adults). Moreover, such algorithms have not yet been implemented within a deliverable format for the end user, much less a standalone, fully automated, self-report instrument.

Therefore, the present study seeks to explore the validity of a previously developed assessment protocol designed to efficiently assess a range of common psychiatric disorders. Formulation of this measure involved application of decision-tree algorithms to a large data set for the purposes of deriving a computer-administered self-report assessment tool. The current investigation describes a two-part study, the first of which has already been conducted in preparation for the second (which constitutes the majority of this dissertation project). It is important to note that the first study was archival and is still open to suggestion or revision as necessary, given continued accessibility of a large data set previously examined.

Briefly, the development phase involved the derivation of abbreviated decision-tree algorithms for 10 common psychiatric disorders based on clinical interview data contained within a large, publicly available data set (Kessler & Merikangas, 2004). The general results from this study were that question sets for specific disorders were greatly reduced in length (often by more than 90%) with limited impact on overall diagnostic accuracy. The more prospective aspect of study will evaluate the algorithms derived in the development phase in terms of their correspondence to diagnostic profiles produced by gold-standard assessment procedures. This is an important part of programmatic research in EBA (and this instrument in particular) given that the predictive accuracy and generalizability of these algorithms has not yet been established beyond the sample and format in which they were developed (this point is discussed in greater detail below; Rokach & Maimon, 2007).

CHAPTER 2

STUDY 1 METHODOLOGY

Access to the National Institute of Mental Health *Collaborative Psychiatric Epidemiology Surveys* (CPES) database was obtained through the Inter-university Consortium for Political and Social Research (ICPSR) website. The CPES provides data concerning base rates of mental health diagnoses in adults by joining together three nationally representative surveys, which include the National Comorbidity Survey Replication (NCS-R), the National Survey of American Life (NSAL), and the National Latino and Asian American Study (NLAAS). These studies took an epidemiological approach to understanding psychiatric diagnoses in the country at large, which required thorough interview of over 20,000 participants with diverse demographic backgrounds. This initiative also required administration of a gold-standard diagnostic interview, the World Health Organization's Composite International Diagnostic Interview (CIDI Andrews & Peters, 1998). Thus, the CPES database contains item-level data from a structured diagnostic interview that was administered to a large, nationally representative sample. Demographics and characteristics of the CPES sample are presented in Table 1 below.

Table 1
Sample Characteristics

Age – Mean (SD)	43.38 (16.72)
Female – No. (%)	11,463 (57.3%)
Race/ancestry – No. (%)	
"Other"	284 (1.4%)
Filipino	508 (2.5%)
Puerto Rican	495 (2.5%)
Vietnamese	520 (2.6%)
Cuban	577 (2.9%)
Chinese	600 (3.0%)
"Other" Asian	656 (3.3%)
"Other" Hispanic	1,106 (5.5%)
Mexican	1,442 (7.2%)
Afro-Caribbean	1,492 (7.5%)
African American	4,746 (23.7%)
Non-Latino White	7,587 (37.9%)
Marital status – No. (%)	
Married/cohabitating	10,726 (53.6%)
Divorced/separated/widowed	4,523 (22.6%)
Never married	4,754 (23.8%)
Missing	10 (0.0%)
Employment status – No. (%)	
Employed	13,123 (65.6%)
Unemployed	1,690 (8.4%)
Not in labor force	5,135 (25.7%)
Household income – \$Mean (SD)	\$50,182 (4,601)

Measures

As stated above, the CPES initiative required administration of a gold-standard diagnostic interview: the CIDI 3.0 (Andrews & Peters, 1998). The CIDI is a fully structured, comprehensive interview that can be used to diagnose lifetime, 12-month, and current diagnoses according to DSM-IV (American Psychological Association, 1994) and ICD-10 (World Health Organization, 1993) criteria. Administration of the CIDI begins with an initial screening module. The screening module contains “gate” items that assess primary features of disorders, thus providing an initial indication as to whether further assessment is warranted for specific diagnostic categories. Depending on which gate items are endorsed, the interview proceeds with exhaustive administration of appropriate subsequent modules to determine the presence (or absence) of symptoms. For example, endorsing a gate item pertaining to “a period of sadness lasting several days or longer” would initiate subsequent administration of the major depressive disorder module. A diagnosis of major depressive disorder would then be made if sufficient criteria are met following exhaustive administration of this module.

The CIDI is the most widely used structured interview for large-scale epidemiological studies (Robins & Regier, 1991; Kessler et al., 1994), and is updated regularly to maintain consistency with current diagnostic systems. Multiple versions of the CIDI have demonstrated strong psychometric properties. In terms of inter-rater reliability, large international field trials have shown that most diagnostic categories demonstrate Kappa coefficients 0.9 or greater (Cottler et al., 1991; Wittchen et al., 1991). Examinations of the CIDI’s test-retest reliability reveal good to excellent Kappa ratings for all but two sections (GAD and panic disorder) over a 1-6 day interval (Semler et al., 1987; Wacker, Battegay, Mullejans, & Schlosser, 1990). The CIDI also demonstrates adequate validity when compared to diagnoses made over a period of time by experts who reach consensus diagnoses after examining all available data (i.e., the LEAD procedure; Spitzer, 1983). Specifically, validation studies indicate that the CIDI detects 88% of LEAD standardized diagnoses (Peters & Andrews, 1995).

Data Analysis

Decision-tree analysis. Ten diagnostic algorithms (Table 4 below) were developed using the Classification Tree function in SPSS 21.0. Decisions regarding which diagnoses to include in the analysis were based on epidemiological base rates (the 10 most prevalent disorders were selected for algorithm development). A specific diagnostic variable indicating the presence/absence of the disorder of interest was selected as the outcome for each algorithm. In terms of input data, the entire set of variables contained within the CIDI module corresponding to the outcome of interest was selected for potential inclusion in each algorithm. Given that the administration of CIDI modules depends on whether gate items are endorsed, data within each specific module are available only for individuals who endorsed the appropriate gate item. As such, decision-tree algorithms were developed within subsets of data from individuals who had endorsed gate items corresponding to the diagnosis of interest.

The QUEST algorithm was utilized in the development of each decision tree. This approach was selected due to variability in terms of the complexity and resolution of the scales on which the CIDI variables are measured (variables are either continuous, dichotomous, or ordinal). Had either of the biased approaches (i.e., CART, CHAID) been used, variables with greater numbers of possible split points would be favored for model inclusion due to their dimensionality alone (and not their predictive utility), thus increasing the potential for suboptimal algorithms. In selecting variables for model inclusion, QUEST performs F-tests for all continuous variables and chi-square tests for all categorical variables to determine their relative strength in regard to predicting the outcome. At any given step in model development, the variable with the smallest *p*-value is selected to make the split at the given node. To identify the optimal cut point on the selected predictor, the algorithm examines all potential cut points on the variable and uses the one that maximizes sub-node homogeneity with respect to the outcome. This process is then repeated until the user-specified growth limit is reached. For each of the 10 algorithms, a growth limit was established using the minimum number of cases rule. Consistent with Onwuegbuzie & Collins' (2010) recommendations, the minimum number of cases was set at 50 for parent nodes and 25 for child nodes.

As described earlier, classification decisions yielded by decision trees are typically based on majority group membership within terminal nodes. Stated differently, all data points within a particular node will be identified as belonging to whichever group holds the node majority within that node (i.e., >50%). Although this approach works well when each of the outcome categories is equally important to detect, it may not provide the desired level of power when one particular category is the target of the analysis. This is particularly true when handling low base rate phenomena such as psychiatric disorders. In such a situation, it may be beneficial to adjust the classification decision rules (i.e., node proportions required to make decisions) such that individuals with the disorder of interest are easier to detect.

Receiver operator characteristic analysis. Receiver operator characteristic (ROC) analysis provides a solution to examining optimal cut off scores to maximize the predictive utility of decision tree algorithms. Specifically, ROC analysis can calculate how well various node cut off scores perform within the context of a particular decision tree. To illustrate this concept, Figure 2 depicts the results of the ROC analysis corresponding to the mania/hypomania algorithm derived in the current study (Figure 3). Figure 2 plots the decision tree's sensitivity on the Y-axis against its specificity on the X- axis. In the present example, sensitivity is the percentage of *manic/hypomanic* individuals that are identified correctly (i.e., true positive rate) and specificity is the percentage of *non-manic/hypomanic* individuals that are correctly identified (i.e., true negative rate). On this particular graph, higher values on the Y-axis indicate greater sensitivity, whereas lower values on the X-axis (i.e., 1 – specificity) indicate greater specificity. The blue line represents the "ROC curve", which plots the range of possible node cut off scores according to their associated levels of sensitivity and specificity. As such, a (large) ROC curve maximizing the area between itself and the green reference line (which represents zero predictive utility), is indicative of an algorithm with high overall performance. The proportion of the graph contained within the ROC curve is represented as an index: area under the ROC curve (AUC). AUC values range from .5 (no predictive utility, represented by the green reference line) to 1.0 (perfect discrimination). In reference to Table 2 below, the AUC associated with the ROC analysis for mania/hypomania is .924, thus indicating that the algorithm exhibits excellent discrimination (Youngstrom, 2014).

Figure 2. ROC analysis for mania/hypomania algorithm

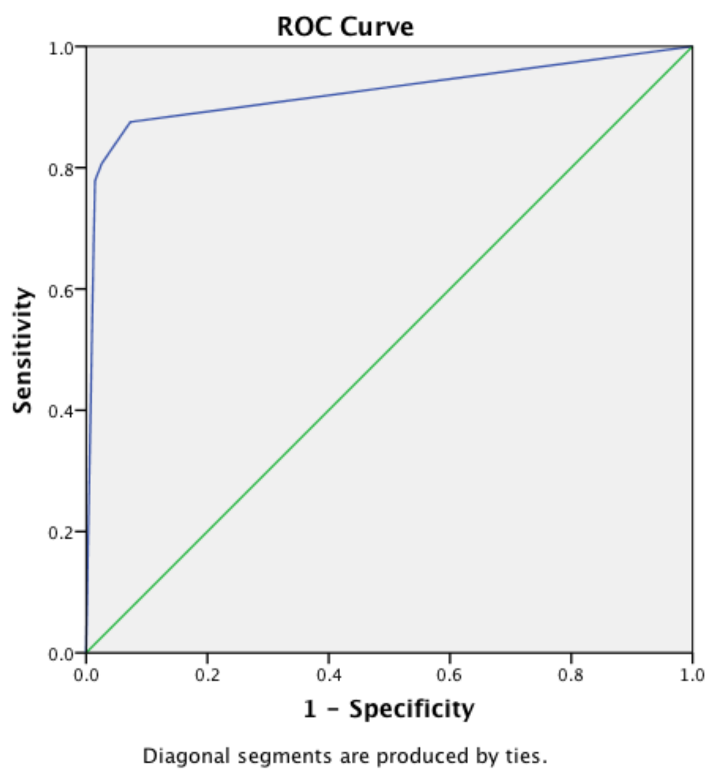


Table 2
Example ROC AUC results for mania/hypomania algorithm
Area Under the Curve

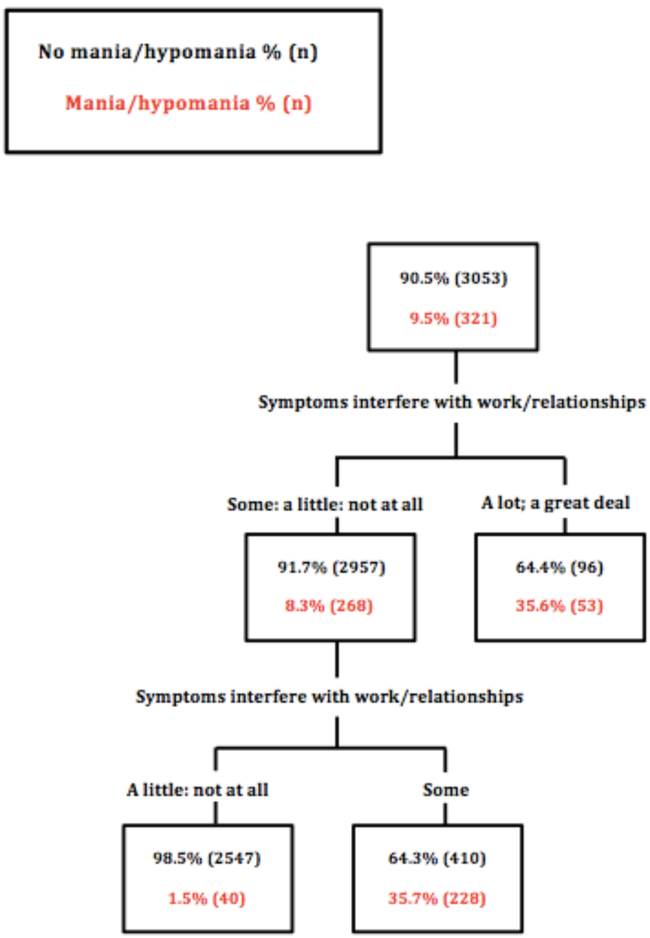
Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.924	.011	.000	.902	.945

As previously mentioned, ROC analysis can also be used to identify optimal cut off scores. Each of the potential cut off scores plotted on the ROC curve is presented in Table 3 below, in addition to their associated levels of sensitivity and specificity. Selecting the optimal cut off score is relatively straightforward: the researcher chooses the score that provides the best balance of sensitivity and specificity within the context of the current application. As a general rule of thumb, the optimal cut off score is represented as the point on the ROC curve that is closest to the top left corner of the ROC graph. Accordingly, a node cut off score of .0392 (i.e., proportion of 3.92%) was selected from Table 3 below for the mania/hypomania algorithm. Hence, any individual residing in a terminal node containing at least 3.92% positive manic/hypomanic cases will be classified as having the disorder. This cut off score yields a sensitivity rating of .875 (i.e., 87.5% of manic/hypomanic individuals are correctly identified) and a specificity rating of .927 (i.e., 1 - .073; 92.7% of non-manic/hypomanic individuals are classified correctly). Importantly, had a standard probability cutoff of 50% been used, the algorithm would have yielded a sensitivity rating of .000 (0% of manic/hypomanic individuals would have been identified). Thus, ROC analysis represents an important compliment to decision-tree analysis as it enables the identification of optimal node cut offs, thereby enhancing overall diagnostic utility.

Table 3
Example ROC cutoffs and corresponding sensitivity and specificity for mania/hypomania algorithm

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0392	.875	.073
.0844	.807	.026
.3967	.779	.015
.7014	.642	.012
1.0000	.000	.000

Figure 3. Decision-tree structure for mania/hypomania algorithm



CHAPTER 3

STUDY 1 RESULTS

The performance and efficiency of each algorithm are presented in Table 4 below. Overall, algorithms yielded AUCs ranging from good to excellent (Ferdinand, 2008). In terms of sensitivity, all algorithms ranged from good to excellent, with the exception of the GAD algorithm (.717; fair). In addition, specificity ratings ranged from good to perfect, with the exception of the specific phobia algorithm (.742; fair). Finally, the decision-tree algorithms were highly efficient relative to their corresponding CIDI modules as the number of items administered ranged from 1 to 5, yielding an item reduction of 83-97% on average.

Table 4
Algorithm performance in development phase

	AUC	Cutoff	Sens.	Spec.	# items
GAD	.868	.338	.717	.996	2-5
Panic	.985	.495	.968	.999	2-3
PTSD	.997	.635	.994	1.000	1-4
Social Phobia	.950	.148	.910	.976	1-2
Specific Phobia	.992	.076	.995	.742	1-2
Major Depressive Episode	.997	.654	.988	1.000	2-3
Mania/Hypomania	.924	.0392	.875	.927	2-4
Alcohol Abuse	.955	.465	.889	.995	1-4
Drug Abuse	.891	.020	.824	.816	1-3
ADD	.897	.145	.822	.940	2-3

After the development phase, the diagnostic algorithms were entered into Qualtrics, an online survey tool. Using Qualtrics enabled implementation of the algorithms within a deliverable format that can be accessed via any electronic device with an internet connection (e.g., iPad, smart phone, laptop). This instrument will be the subject of validation study 2, which is the basis for this dissertation project.

CHAPTER 4

STUDY 2 METHODOLOGY

A major issue with decision-tree methods is that algorithms tend to overestimate their own classification accuracy. This is because when decision-trees are grown, their structure develops in such a way that is optimally fitted to the specific input data. As a result, the decision-tree's accuracy will often be reduced when applied to external data due to differences in the subjects that are contained within the new sample. It is thus critical to test and validate decision-tree algorithms for accuracy and generalizability beyond the sample from which they were developed. The importance of validation is particularly significant within the context of the current instrument, as it will be implemented within a format/modality that differs from its origin (i.e., self-report versus structured interview).

Sample

Two-hundred fifty-nine adult patients presenting in the emergency department at the University of Mississippi Medical Center in Jackson, MS served as the sample for the current study. Patients were between the ages of 18 and 91, with a mean age of 47.5. One-hundred thirty-seven (52.9%) of the patients were women. One-hundred forty-nine (57.5%) were African American, 109 (42.1%) were White, and 1 was (.4%) Hispanic.

Procedure

After providing consent for participation in the study, patients were assessed with the MINI and SCID to establish a criterion diagnostic profile (Sheehan et al., 2002; First, Spitzer, Robert, Gibbon, Miriam, & Williams, 2007). The SCID was used to establish diagnoses for specific phobia; all other diagnoses were established with the MINI. Three psychology graduate students were trained in the delivery of the instruments and they conducted the structured interviews under the supervision of a PhD clinical psychologist. All interviews were conducted while patients were waiting to be seen by the emergency department personnel following triage. The screening instrument was administered immediately following completion of the structured interview. The algorithms for Generalized Anxiety Disorder, alcohol use, and drug use were edited midway though data collection to address patients' confusion with the items' initial wording. The sample size used to validate these three algorithms is thus somewhat smaller than what was used to validate the other algorithms.

Measures

The Mini International Neuropsychiatric Interview 5.0 (MINI; Sheehan & Lecrubier, 2002). The MINI is a brief structured interview that assesses 17 disorders according to DSM diagnostic criteria. In terms of administration format, the MINI employs one or two screening questions for each disorder that rule out diagnoses when not endorsed. If screening questions are endorsed, relevant symptomatology is assessed further by investigation of other diagnostic criteria. The interview focuses on current experience of symptoms and probes for severity, disability, and potential medical explanations for observed difficulties. Psychometric studies have examined the MINI's inter-rater and test-retest reliability, as well as its validity versus a number of structured clinical interviews, including the CIDI.

Studies comparing the MINI to the CIDI as the gold-standard diagnostic criterion (Lecrubier et al., 1997) have indicated that kappa coefficients for concordance were typically in the good or very good range. Some exceptions occurred in diagnostic categories of simple phobia (.43) and GAD (.36), but kappas for all other diagnoses were sufficiently high that they were considered as convergent evidence of the much briefer MINI's utility. The MINI's specificity was high for all diagnoses (.72-.97) as was sensitivity for all but the following categories: panic disorder (.67), agoraphobia (.59), simple phobia (.46), and bulimia (.63). Negative predictive values were high for all diagnoses and positive predictive values were also generally strong with the exception of GAD (.34), bulimia (.52), current manic episode (.56), and social phobia (.55). Finally, the MINI's inter-rater reliability was very high (kappa coefficients ranging from .80 to 1.0) and test-retest reliability was acceptable (kappa coefficients ranging from .76 to .93). In terms of establishing test-retest reliability, the MINI was re-administered within two days of the first administration.

Structured Clinical Interview for DSM-IV-TR (SCID; First, Spitzer, Robert, Gibbon, Miriam, & Williams, 2007). The SCID is a gold-standard structured interview that is used to diagnose DSM Axis I disorders. It is often used as the criterion against which newer structured interviews (e.g., the MINI) are validated. The SCID shows generally strong agreement with other structured interviews and LEAD diagnoses (e.g., Sheehan et al., 1997; Haro et al., 2006, Basco et al., 2000).

Analytic plan

The validity of the abbreviated instrument implemented in this study was examined using the MINI and SCID (for specific phobia) to discern the sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and Kappa. SPSS 21 was used to conduct all statistical analyses. For all more detailed descriptions that follow, the MINI and SCID (for specific phobia) were used to provide data for the criterion variable (i.e., the accurate outcome, which will be used to gauge the performance of the abbreviated instrument).

Sensitivity. As described above, sensitivity is a metric that describes the ability of a test to correctly identify individuals with the condition as having the condition. Specifically, sensitivity is the proportion of individuals with the condition that are correctly identified as having the condition. Possible values range from 0.0 to 1.0, with larger values indicating greater sensitivity.

Specificity. Also described above, specificity is a metric that described the ability of a test to correctly identify individuals without a condition as not having the condition. Specifically, it is the proportion of individuals without the condition that are correctly identified as not having the condition. Like sensitivity, possible values range from 0.0 to 1.0 with larger values indicating greater specificity.

PPV. PPV describes the probability that an individual identified as having a condition truly has the condition. It is calculated by dividing the number of truly positive cases by the total number of cases identified as having the condition (true positives/ true positives + false positives). Possible values range from 0.0 to 1.0, with higher values indicating greater positive predictive value.

NPV. NPV describes the probability that an individual identified as not having a condition truly does not have the condition. It is calculated by dividing the number of truly negative cases by the total number of cases identified by the test as not having the condition (true negatives/ true negatives + false negatives).

Cohen's kappa coefficient. Kappa is a statistic that can be used to measure concordance rates between tests with categorical outcomes. In general, it is thought to be a more stringent statistic than simple percent agreement as it accounts for agreement occurring by chance (Landis & Koch, 1977). Values range from 0.0 to 1.0, with higher values indicating greater concordance. Interpretation for the Cohen's kappa coefficient is as follows: < 0 = less than chance agreement; 0.01-0.20 = slight agreement; 0.21-0.40 = fair agreement; 0.41-0.60 = moderate agreement; 0.61-0.80 = substantial agreement; 0.81-0.99 = almost perfect agreement (Viera & Garrett, 2005).

CHAPTER 5

STUDY 2 RESULTS

The performance of each screening algorithm is presented Table 5. The patient sample sizes for the various algorithms were: 183 for alcohol and drug use, 133 for GAD, and 259 for all the other algorithms. The operating characteristics of the screening algorithms were generally strong. Sensitivity was between 75% and 100% for all areas of the algorithms except specific phobia (64%) and problematic alcohol use (50%). All of the algorithms demonstrated sensitivity and NPVs that were greater than 79%. PPVs were above 75% for major depression/dysthymia and alcohol use problems; above 50% for panic, post-traumatic stress disorder (PTSD), GAD, and drug use problems; and above 25% for social anxiety, mania/hypomania, and specific phobia. Kappa estimates were above .60 for major depression/dysthymia, PTSD, GAD, drug use, and alcohol use (i.e., "substantial agreement"); above .40 for panic and social anxiety (i.e., "moderate agreement"); and above .20 for mania/hypomania and specific phobia (i.e., "fair agreement"; Viera & Garrett, 2005).

Table 5
Algorithm performance in validation phase

Disorder Category	**n (%)**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Kappa**
MDD/Dysthymia	96 (37)	82	87	79	89	.68
Panic	53 (21)	87	79	51	96	.52
PTSD	38 (15)	87	91	64	98	.68
Social Anxiety	26 (10)	89	86	40	99	.48
Mania/Hypomania	20 (8)	75	85	29	98	.35
GAD	17 (13)	94	88	53	99	.62
Specific Phobia	11(8)	64	84	27	96	.29
Drug Use	8 (4)	75	98	67	99	.69
Alcohol Use	8 (4)	50	100	100	98	.66

* *GAD = Generalized Anxiety Disorder; MDD = Major Depressive Disorder; PTSD = Post-Traumatic Stress Disorder*

CHAPTER 6

DISCUSSION

Collectively, study 1 and 2 sought to develop and validate a clinically feasible (i.e., brief, yet accurate) screening procedure to facilitate the identification of mental health disorders in clinical practice settings. In study 1, decision-tree analysis reduced the number of CIDI items needed to identify various mental health disorders to question subsets containing one to five items (depending on the disorder and the respondents' answers to questions). Furthermore, study 1 indicated that the abbreviated algorithms retained classification accuracy relative to administration of the full CIDI diagnostic modules, within a small, quantifiable degree of error. The results of study 1 thus demonstrated the utility of decision-tree methods with respect to developing efficient algorithms for identifying specific mental disorders.

Following the initial development phase, the algorithms were incorporated into an electronic, self-report screening instrument that was used to prospectively test the algorithms in an applied, ecologically valid context (i.e., patients presenting to an emergency department at a large medical center). Performance estimates for sensitivity, specificity, PPV, NPV, and Kappa were generally supportive of the screening algorithms' validity; however, the sensitivity estimate for the alcohol abuse algorithm was relatively low. It is possible that sensitivity for this condition was lower due to participants' unwillingness to endorse problematic use in a self-report format as opposed to a structured interview. The literature concerning the effects of data collection modes on assessing substance abuse is inconsistent; however, at least two studies have demonstrated substantially lower rates of endorsed problematic consumption on questionnaires as opposed to an interview format (Rehm & Arminger, 1996; Rehm & Spuler, 1993). Regardless, the performance of the alcohol abuse algorithm likely represents an improvement over the subjective assessment procedures typically employed in applied settings given the inaccuracy such methods consistently demonstrate (e.g., Ustun & Sartorius, 1995; Dawes et al., 1989). Moreover, all the NPVs, including the estimate for the alcohol use algorithm, were very strong, thus indicating the algorithms' ability to avoid false negatives. Strong performance in this domain demonstrates the utility of the instrument as a first-pass screening procedure because it enables clinicians to be confident that any individual receiving a negative screen does not likely meet criteria for the given disorder. As opposed to the potential need for additional follow up when cases screen positive, this enables treatment planning that does not require more testing to verify negative screens.

Beyond demonstrating support for the validity of the screening instrument, the results of the current study and the process of data collection more generally revealed a number of important issues regarding the prevalence and handling of mental illness in ED settings. First, a number of conditions (e.g., depression, PTSD, panic, suicidality) were far more common among ED patients than prior studies examining billing data for ED admissions suggest (e.g., Owens, Mutter, & Stocks, 2010). This discrepancy is likely due to the methodology employed by such studies, which relies on the ability of ED personnel to identify mental conditions, generally without the aid of psychometric instruments. Because ED personnel usually do not detect a psychiatric condition unless it is extremely severe or the primary reason for admission, much of the prior research may underestimate the prevalence and impact of such disorders (La Vonne, Zun, & Burke, 2012; Zun, 2016). It is worth noting that most of the participants in the current study with a mental condition were probably not receiving mental health treatment, and would not have been identified as needing such care through standard ED triage procedures. Individuals with untreated mental illness frequently utilize the ED and generally cost about 2.5 to 3.5 times more to treat, meaning the current lack of integrated mental health services in ED settings is a major source of inefficiency within the health care system (Melek, Norris, and Paulus, 2014). The screening instrument could thus facilitate current efforts aimed at stemming wasteful spending by providing ED personnel with a means to identify patients in need of mental health treatment.

There is no guarantee that enhancing recognition of mental illness would necessarily lead to appropriate treatment or referral, however. For example, during data collection it was unclear whether notifying ED personnel of psychiatric diagnoses had any impact on the delivery of care. From what could be observed, ED personnel generally did not conduct further psychological assessment, offer treatment, or make referrals based on the information that was provided to them. One exception is when patients endorsed active suicidality. In such cases, ED personnel usually contacted a supervising physician or psychiatric resident to conduct an assessment, although this process sometimes took hours to initiate and complete. During the interim, patients were placed on a medical hold and could not legally leave on their own accord. Furthermore, if intervention was deemed necessary, a patient might be "boarded" for any number of additional hours until an inpatient bed could be secured. This practice of boarding is a major contributing factor to the widespread overcrowding currently observed in EDs across the US, and is associated with lower-quality care and poorer outcomes (Baillargeon et al., 2008; Downey, Zun, & Gonzales, 2009; Hackman et al., 2006; Salinsky & Loftis, 2007; Tang et al., 2007; Chakravarthy, Tenny, Anderson, Rajeev, Istanbouli, & Lotfipour, 2013). Given these issues, improving ED patients' access to mental health treatment and lowering the cost of care more broadly will likely require efforts beyond instating regular screening, including systemic changes that increase the number of resources available for treating mental illness. Nonetheless, providing personnel with a means of efficient and accurate assessment represents an important first step towards increasing access to mental health treatment in ED settings.

The base rates observed in the current study are not unique to the ED—they reflect a pattern that indicates a high prevalence of mental illness among patients within the health care system more broadly. In particular, mental health concerns are common in primary care settings as one-third of all appointments involve a direct psychiatric component (Ustun et al., 1995; Linden, Maier et al., World Health Organization, 1992). Furthermore, the vast majority of Americans utilize primary care as their *only* source of mental health services (Kathol et al., 2005; Regier, 1993; Wang et al., 2006; Wang et al., 2005). Despite serving as the de facto standard for mental health services, primary care physicians generally lack the time and expertise to employ EBA practices and thus fail to recognize more than 50% of affected individuals (Ustun et al., 1995). Even when physicians do recognize mental illness, less than half of such cases ever receive first-line, evidence-based treatment (Ustun et al., 1995). Although there are a number of barriers that prevent access to quality mental health treatment in primary care, one key factor is the fragmentation of behavioral and medical services in the recognition and treatment of psychiatric disorders.

The past decade has thus witnessed increased efforts to bridge the gap between medical and behavioral healthcare. A prime example is the growing trend towards Behavioral Health Integration (BHI), which engenders collaboration among primary care providers and mental health specialists in the provision of comprehensive healthcare. A number of BHI models are still evolving and much work is needed to refine implementation in general; however, studies have shown that BHI can improve population health, increase patient satisfaction, and reduce the overall cost of healthcare (Archer et al., 2012; Butler et al., 2011; Reed, Shore, & Tice, 2016; Reiss-Brennan et al., 2016). One element that is consistent across all BHI initiatives is the need to identify mental illness in general medical settings (Collins, Hewson, Munger, & Wade, 2010; Geritty, 2016). The US Preventive Services Task Force thus recommends mental health screening for all primary care appointments (Siu, 2016) via the nine-item Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002). Although the PHQ-9 is the most frequently used mental health questionnaire in medical settings, it is somewhat limited in that it is specific to depression and does not assess other conditions that frequently present in primary care (e.g., anxiety, trauma-related distress, substance abuse). As such, the current screening instrument may facilitate progress toward BHI by providing medical professionals with a means to conduct psychological assessment.

Whether implemented in a primary care clinic, ED, or even a community mental health center, psychological screening procedures are not likely to achieve sustained use unless they are efficient and easy to use. Fortunately, the current screening instrument has at least two features intended to meet such demands. First, it can be administered automatically via any electronic device with an internet connection (e.g., tablet, computer, smartphone) in under 10 minutes. This provides considerable flexibility as patients can access the instrument from a variety of locations, such as their home environment or a clinic waiting room. Furthermore, responses are scored automatically and clinicians can access an electronic patient profile as soon as administration is completed. Not only does this feature eliminate human error in terms of tabulating scores, but also provides the practitioner with screening results that are straightforward and informative for guiding clinical decision making.

It is important to note, however, that despite its convenience, the screening instrument is not intended to replace gold-standard EBA practices (i.e., administration of structured interviews, multi-method assessment). Considering the observed PPVs, a significant proportion of positive screens result in false positives, thus highlighting the need for additional assessment to determine whether a formal diagnosis is warranted. In a medical setting, results from the screener would thus ideally prompt further psychological assessment from a qualified specialist. If implemented within a mental health setting, results from the screening instrument would most appropriately be used to guide the administration of specific structured interview modules for the indicated conditions. Of course, such use represents the ideal—even if the instrument were not followed-up with additional assessment, it may still provide better results as compared to the poorly supported techniques that are typically used in applied settings (Garland, Kruse & Aarons, 2003; Hatfield & Ogles, 2004; Sattler et al., 2016; Whiteside, Sattler, Hathaway & Douglas, 2016; Dawes et al., 1989).

Given the screening instrument's potential to facilitate EBA use in diverse applied environments, additional work is needed to examine and address factors likely to influence the success of dissemination and implementation (D&I) efforts. For example, successful D&I will require close attention to end users' attitudes regarding the utility and clinical feasibility of the instrument (Chorpita & Regan, 2009; Rogers, 2003; Stirman, Crits-Christoph, & DeRubeis, 2004). Messaging aimed at achieving uptake among clinicians might thus focus primarily on the instrument's potential to complement existing practices, as opposed to an academic presentation of its psychometric viability. Clinicians might also be more likely to adopt the instrument if they are informed of a recent proposal by Centers for Medicare and Medicaid Services (CMS) to instate a new Current Procedural Terminology (CPT) code that enables reimbursement for the administration of standardized psychological assessments (Centers for Medicare & Medicaid Services, 2016). This proposal represents a major opportunity as it provides clinicians with a financial incentive to conduct regular mental health screening. It also signifies that CMS is serious about moving towards BHI as the programs do not often expand billing options. D&I efforts may also target clinicians' knowledge of EBA in general, as training/education is predictive of the propensity to adopt such practices (Jensen-Doss et al., 2010; Jensen-Doss et al., 2011; Whiteside et al., 2016). Perhaps D&I initiatives could accomplish this by developing training programs that are geared towards clinicians in emphasizing the basic benefits of data-based decision making within a format that does not require in-depth knowledge of psychometrics or statistics (e.g., Lyon, Charlesworth-Attie, Vander Stoep, & McCauley, 2011; Lyon, Dorsey, Pullmann, Silbaugh-Cowdin, & Berliner, 2015).

In addition to clinician-level issues, D&I efforts might do well to target organizational factors likely to foster sustained adoption (Rogers, 2003). Namely, organizations might be encouraged to support personnel by offering resources such as regular question and answer sessions and/or listservs to provide a channel through which to address concerns regarding use of the instrument (Berliner, Dorsey, Merchant, Jungbluth & Sedlar, 2013; Lyon et al., 2015). Other organizational supports might include point of contact reminders and measurement feedback systems, which can be used to prompt personnel to administer the instrument and automatically integrate results into the delivery of treatment, respectively (Bickman, Kelley, & Athay, 2012; Shojania and Grimshaw, 2005). Finally, D&I efforts might also do well by attending to patients' experience with the instrument to ensure its clinical acceptability as clinicians may be less likely to use the instrument if they believe it is disruptive to developing/maintaining therapeutic rapport (Bruchmuller, Margraf, Suppiger, & Schneider, 2011; Suppiger, In-Albon, Hendriksen, Hermann, Margraf, & Schneider, 2009).

Further work is also needed to extend the results of the current study, as well as address its limitations. First, future research should examine the test-retest reliability of the screening instrument as logistical constraints prevented collecting such data during the present study. If such studies demonstrated temporal stability of the instrument, this could provide a foundation for additional research using the tool for tracking clinical outcomes longitudinally. Upcoming research should also capture information regarding respondents' educational and socioeconomic background to determine whether such variables influence the instrument's validity. Given that the current study utilized a sample of patients residing in an urban area located in one of the most economically impoverished and least educated US states (DeNavas-Walt, Proctor, & Smith, 2015), it is possible that the instrument might yield stronger performance if examined in a different location. Future research might also examine whether specific items, or the instrument as a whole, function differentially depending on respondents' ethnicity, gender, age, or other demographic variables (i.e., measurement invariance; Brown, 2014; Vandenberg & Lance, 2000). For example, prior research indicates that ethnic minorities endorse greater somatization in the expression of mood disorders (e.g., Ferrari et al., 2015) and several prevalent measures of emotional symptoms have been shown to be non-invariant (e.g. Trent et al., 2012; Raykov, 2004). Thus, it is possible that items placing more/less emphasis on physical symptoms would improve overall performance depending on the ethnic characteristics of the sample. Additional work is also needed to examine the effect that various methods of administration might have in terms of impacting validity. For example, the instrument may yield differential performance depending on whether it is implemented orally versus in a self-report format; administered via smartphone versus tablet; or accessed in a patient's home environment versus a clinical setting. Future research should also seek to examine the performance of the screening algorithms as they relate to newer diagnostic systems (i.e., these algorithms were developed and validated based on the DSM-IV). Finally, additional work should examine methods of ranking comorbid conditions ordinally to improve the utility of the instrument in terms of making hierarchical diagnoses. Potential models for enabling such an undertaking might involve classification trees for ordinal outcomes (Frank & Hall, 2001), applications of item response theory (Embretson & Reise, 2013), or random forest/ensemble methods for ordinal classification (Breiman, 2001).

REFERENCES

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The reliability of psychiatric diagnosis revisited. *Psychiatry, 3*(1), 41-50.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Anderson, D., & Paulosky, C. (2004). Psychological assessment of eating disorders and related features. *Handbook of Eating Disorders and Obesity,* , 112-129.

Anderson, N. B. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271-285.

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

Appavu, S., Rajaram, R., Muthupandian, M., & Athippan, G. (2008). Automatic mining of threatening e-mail using ad infinitum algorithm. *International Journal of Information Technology, 14*(2), 81-108.

Archer, J., Bower, P., Gilbody, S., Lovell, K., Richards, D., Gask, L., ... & Coventry, P. (2012). Collaborative care for depression and anxiety problems. *The Cochrane Library*.

Bachman, J. W. (2003). The patient-computer interview: A neglected tool that can aid the clinician. Paper presented at the *Mayo Clinic Proceedings, , 78*(1) 67-78.

Baillargeon, J., Thomas, C. R., Williams, B., Begley, C. E., Sharma, S., Pollock, B. H., ... & Raimer, B. (2008). Medical emergency department utilization patterns among uninsured patients with psychiatric disorders. *Psychiatric Services*, *59*(7), 808-811.

Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*(10), 1599-1605.

Batterham, P. J., & Christensen, H. (2012). Longitudinal risk profiling for suicidal thoughts and behaviours in a community cohort using decision trees. *Journal of Affective Disorders, 142*(1), 306-314.

Batterham, P. J., Christensen, H., & Mackinnon, A. J. (2009). Modifiable risk factors predicting major depressive disorder at four year follow-up: A decision tree approach. *BMC Psychiatry, 9*(1), 1.

Becker, J., Kocalevent, R., Rose, M., Fliege, H., Walter, O. B., Frommer, J., & Klapp, B. F. (2006). Standardisierte diagnosefindung: Computergestützte (CIDI)-versus klinische diagnosestellungen an einer psychosomatischen stichprobe. *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie, 56*(01), 5-14.

Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don’t tell me: Behavioral rehearsal as a training and analogue fidelity tool. *Cognitive and Behavioral Practice, 21*(1), 1-11.

Berliner L, Dorsey S, Merchant L, Jungbluth N, Sedlar G. Practical Guide for EBP Implementation in Public Health. Washington State Division of Behavior Health and Recovery, Harborview Center for Sexual Assault and Traumatic Stress & University of Washington, School of Medicine, Public Behavior Health and Justice Policy; 2013.

Bickman, L., Kelley, S. D., & Athay, M. (2012). The technology of measurement feedback systems. *Couple and Family Psychology: Research and Practice*, *1*(4), 274.

Bickman, L., Smith, C. M., Lambert, E. W., & Andrade, A. R. (2003). Evaluation of a congressionally mandated wraparound demonstration. *Journal of Child and Family Studies, 12*(2), 135-156.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress, 8*(1), 75-90.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees* CRC press.

Bris, R., Majernik, J., Pancerz, K., & Zaitseva, E. (2016). *Applications of computational intelligence in biomedical technology* Springer.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Brown, T. A., Barlow, D. H., & Di Nardo, P. A. (1994). *Anxiety disorders interview schedule for DSM-IV (ADIS-IV): Client interview schedule/*. Graywind Publications Incorporated.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107*(2), 179.

Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? overdiagnosis and influence of client gender on diagnosis. *Journal of Consulting and Clinical Psychology, 80*(1), 128.

Bumbarger, B. K., & Campbell, E. M. (2012). A state agency–university partnership for translational research and the dissemination of evidence-based prevention and intervention. *Administration and Policy in Mental Health and Mental Health Services Research, 39*(4), 268-277.

Buri, C., von Bonin, B., Strik, W., & Moggi, F. (2009). Predictors of attempted suicide among swiss patients with alcohol-use disorders. *Journal of Studies on Alcohol and Drugs, 70*(5), 668-674.

Butler, M., Kane, R. L., McAlpine, D., Kathol, R., Fu, S. S., Hagedorn, H., & Wilt, T. (2011). Does integrated care improve treatment for depression?: a systematic review. *The Journal of ambulatory care management*, *34*(2), 113-125.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*(2), 141.

Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice, 33*(5), 446.

Centers for Medicare & Medicaid Services (2016). Medicare Program; Revisions to Payment Policies under the Physician Fee Schedule and Other Revisions to Part B for CY 2017

Chakravarthy, B., Tenny, M., Anderson, C. L., Rajeev, S., Istanbouli, T., & Lotfipour, S. (2013). Analysis of Mental Health Substance Abuse–Related Emergency Department Visits from 2002 to 2008. *Substance abuse*, *34*(3), 292-297.

Chen, X., Rusinko, A., & Young, S. S. (1998). Recursive partitioning analysis of a large structure-activity data set using three-dimensional descriptors 1. *Journal of Chemical Information and Computer Sciences, 38*(6), 1054-1062.

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health and Mental Health Services Research, 35*(1-2), 114-123.

Chorpita, B. F., & Nakamura, B. J. (2008). Dynamic structure in diagnostic structured interviewing: A comparative test of accuracy and efficiency. *Journal of Psychopathology and Behavioral Assessment, 30*(1), 52-60.

Chorpita, B. F., & Regan, J. (2009). Dissemination of effective mental health treatment procedures: Maximizing the return on a significant investment. *Behaviour Research and Therapy*, *47*(11), 990-993.

Chorpita, B. F., Yim, L. M., & Tracey, S. A. (2002). Feasibility of a simplified and dynamic bayesian system for use in structured diagnostic interviews. *Journal of Psychopathology and Behavioral Assessment, 24*(1), 13-23.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5*(1), 8-14.

Christon, L. M., McLeod, B. D., & Jensen-Doss, A. (2015). Evidence-based assessment meets evidence-based treatment: An approach to science-informed case conceptualization. *Cognitive and Behavioral Practice, 22*(1), 36-48.

Collins, C., Hewson, D. L., Munger, R., & Wade, T. (2010). *Evolving models of behavioral health in- tegration in primary care*. New York, NY: Mil- bank Memorial Fund.

Cottler, L. B., Robins, L. N., Grant, B. F., Blaine, J., Towle, L. H., Wittchen, H. U., & Sartorius, N. (1991). The CIDI-core substance abuse and dependence questions: Cross-cultural and nosological issues. the WHO/ADAMHA field trial. *The British Journal of Psychiatry : The Journal of Mental Science, 159*, 653-658.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775-780.

D'Alisa, S., Miscio, G., Baudo, S., Simone, A., Tesio, L., & Mauro, A. (2006). Depression is the main determinant of quality of life in multiple sclerosis: A classification-regression (CART) study. *Disability and Rehabilitation, 28*(5), 307-314.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science (New York, N.Y.), 243*(4899), 1668-1674.

DeNavas-Walt, C., Proctor, B. D., & Smith, J. C. (2015). *Income and poverty in the United States: 2014* (p. 22). United States Census Bureau.

Denys, D., van der Wee, N., van Megen, H. J., & Westenberg, H. G. (2003). A double blind comparison of venlafaxine and paroxetine in obsessive-compulsive disorder. *Journal of Clinical Psychopharmacology, 23*(6), 568-575.

Dignon, A. (1996). Acceptability of a computer-administered psychiatric interview. *Computers in Human Behavior, 12*(2), 177-191.

Downey, L. V., Zun, L. S., & Gonzales, S. J. (2009). Utilization of emergency department by psychiatric patients. *Primary Psychiatry*, *16*(4), 60-64.

Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., . . . Weisz, J. R. (2012). The revised child anxiety and depression scale-short version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment, 24*(4), 833.

Ehlert, U. (2007). Eine psychotherapie ist immer nur so gut wie ihre diagnostik. *Verhaltenstherapie, 17*(2), 81-82.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Ferrari, S., Poloni, N., Diefenbacher, A., Barbosa, A., & Cosci, F. (2015). From hysteria to somatic symptom disorders: searching for a common psychopathological ground. *Journal of Psychopathology*, *21*, 372-379.

First, M. B., & Gibbon, M. (2004). The structured clinical interview for DSM-IV axis I disorders (SCID-I) and the structured clinical interview for DSM-IV axis II disorders (SCID-II).

Frank, E., & Hall, M. (2001, September). A simple approach to ordinal classification. In *European Conference on Machine Learning* (pp. 145-156). Springer Berlin Heidelberg.

Frazier, T. W., & Youngstrom, E. A. (2006). Evidence-based assessment of attention-deficit/hyperactivity disorder: Using multiple sources of information. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(5), 614-620. doi:10.1097/01.chi.0000196597.09103.25 [doi]

Garb, H. N. (1998). *Case formulation*. American Psychological Association.

Garb, H. N. (2007). Computer-administered interviews and rating scales. *Psychological Assessment, 19*(1), 4.

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services & Research, 30*(4), 393-405.

Gerrity, M. (2016). Evolving models of behavioral health integration: evidence update 2010-2015.

Ghanizadeh, A., Mohammadi, M. R., & Yazdanshenas, A. (2006). Psychometric properties of the farsi translation of the kiddie schedule for affective disorders and schizophrenia-present and lifetime version. *BMC Psychiatry, 6*(1), 1.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Outcomes research in mental health. systematic review. *The British Journal of Psychiatry : The Journal of Mental Science, 181*, 8-16.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293.

Guy, L. S., Poythress, N. G., Douglas, K. S., Skeem, J. L., & Edens, J. F. (2008). Correspondence between self-report and interview-based assessments of antisocial personality disorder. *Psychological Assessment, 20*(1), 47.

Hackman, A. L., Goldberg, R. W., Brown, C. H., Fang, L. J., Dickerson, F. B., Wohlheiter, K., ... & Dixon, L. (2006). Brief reports: use of emergency department services for somatic reasons by people with serious mental illness. *Psychiatric Services*, *57*(4), 563-566.

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., De Girolamo, G., Guyer, M. E., Jin, R., ... & Sampson, N. A. (2006). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *International journal of methods in psychiatric research*, *15*(4), 167-180.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35*(5), 485.

Hathaway, S. R., McKinley, J. C., & MMPI Restandardization Committee. (1989). *MMPI-2: Minnesota multiphasic personality inventory-2: Manual for administration and scoring* University of Minnesota Press.

Hoyer, J., Ruhl, U., Scholz, D., & Wittchen, H. (2006). Patients' feedback after computer-assisted diagnostic interviews for mental disorders. *Psychotherapy Research, 16*(03), 357-363.

Hunsley, J., & Mash, E. J. (2011). Evidence-based assessment. *Oxford Handbook of Clinical Psychology,* , 76-97.

Ilgen, M. A., Downing, K., Zivin, K., Hoggatt, K. J., Kim, H. M., Ganoczy, D., . . . Valenstein, M. (2009). Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide [CME]. *The Journal of Clinical Psychiatry, 70*(11), 1495-1500.

Jachyra, D., Pancerz, K., & Gomuła, J. (2013). Multiway classification of MMPI profiles. Paper presented at the *Digital Technologies (DT), 2013 International Conference on,* 90-98.

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70*(1), 158.

Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology, 39*(6), 885-896.

Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology, 76*(5), 711.

Jensen-Doss, A., Youngstrom, E. A., Youngstrom, J. K., Feeny, N. C., & Findling, R. L. (2014). Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of Consulting and Clinical Psychology, 82*(6), 1151.

Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the diagnostic interview schedule for children. *Journal of Clinical Child and Adolescent Psychology, 33*(3), 536-546.

Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the diagnostic interview schedule for children. *Journal of Clinical Child and Adolescent Psychology, 33*(3), 536-546.

Joiner Jr, T. E., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment, 17*(3), 267.

Kahana, S. Y., Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Employing parent, teacher, and youth self-report checklists in identifying pediatric bipolar spectrum disorders:

An examination of diagnostic accuracy and clinical utility. *Journal of Child and Adolescent Psychopharmacology, 13*(4), 471-488.

Kathol, R. G., & Clarke, D. (2005). Rethinking the place of the psyche in health: toward the integration of health care systems. *Australian and New Zealand Journal of Psychiatry*, *39*(9), 816-825.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(7), 980-988.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., . . . 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., & Merikangas, K. R. (2004). The national comorbidity survey replication (NCS-R): Background and aims. *International Journal of Methods in Psychiatric Research, 13*(2), 60-68.

Kessler, R. C., & Üstün, 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*(2), 93-121.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509-515.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of General Internal Medicine, 16*(9), 606-613.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* , 159-174.

La Vonne, A. D., Zun, L. S., & Burke, T. (2012). Undiagnosed mental illness in the emergency department. *The Journal of emergency medicine*, *43*(5), 876-882.

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., . . . Dunbar, G. C. (1997). The mini international neuropsychiatric interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry, 12*(5), 224-231.

Lewis, G. (1994). Assessing psychiatric disorder with a human interviewer or a computer. *Journal of Epidemiology and Community Health, 48*(2), 207-210.

Linden M, Maier W, Achberger M, et al. (1996). Psychische Erkrankungen und ihre Behandlun in Allgemeinarztpraxen in Deutschland. Ergebnisse auseiner Studie der Weltgesundheitsorganisation (WHO). *Nervenarzt ;67*. 205-215.

Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., & Coid, J. W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology, 27*(4), 547-573.

Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69*(2), 150.

Lyon, A. R., Charlesworth-Attie, S., Vander Stoep, A., & McCauley, E. (2011). Modular psychotherapy for youth with internalizing problems: Implementation with therapists in school-based health centers. *School Psychology Review*, *40*(4), 569.

Lyon, A. R., Dorsey, S., Pullmann, M., Silbaugh-Cowdin, J., & Berliner, L. (2015). Clinician use of standardized assessments following a common elements psychotherapy training and consultation program. *Administration and policy in mental health and mental health services research*, *42*(1), 47-60.

Mann, J. J., Ellis, S. P., Waternaux, C. M., Liu, X., Oquendo, M. A., Malone, K. M., . . . Currier, D. (2008). Classification trees distinguish suicide attempters in major psychiatric disorders: A model of clinical decision making. *The Journal of Clinical Psychiatry, 69*(1), 23-31.

Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 362-379.

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.

Melek, S. P., Norris, D. T., & Paulus, J. (2014). Economic impact of integrated medical-behavioral healthcare. *Milliman Am Psychiatr Assoc Rep*.

Nunno, M. (2006). The effects of the ARC organizational intervention on caseworker turnover, climate, and culture in children's services systems. *Child Abuse & Neglect, 30*(8), 849-854.

Onwuegbuzie, A., & Collins, K. (2010). An innovative method for stress and coping researchers for analyzing themes in mixed research: Introducing chi-square automatic interaction detection (CHAID). *Research on Stress and Coping in Education.Toward a Broader Understanding of Stress and Coping: Mixed Methods Approaches,* , 287-301.

Owens, E., Griffiths, R., & Ratnatunga, K. (1996). Using oblique decision trees for the morphological classification of galaxies. *Monthly Notices of the Royal Astronomical Society, 281*(1), 153-157.

Owens, P. L., Mutter, R., & Stocks, C. Mental health and substance abuse-related emergency department visits among adults, 2007. HCUP statistical brief no. 92. Rockville, MD: US Department of Health and Human Services, Agency for Healthcare Research and Quality; 2010.

Palmiter Jr, D. J. (2004). A survey of the assessment practices of child and adolescent clinicians. *American Journal of Orthopsychiatry, 74*(2), 122.

Petrie, K., & Abell, W. (1994). Responses of parasuicides to a computerized interview. *Computers in Human Behavior, 10*(4), 415-418.

Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of managed care on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment, 70*(3), 441-447.

Pogge, D. L., Wayland-Smith, D., Zaccario, M., Borgaro, S., Stokes, J., & Harvey, P. D. (2001). Diagnosis of manic episodes in adolescent inpatients: Structured diagnostic procedures compared to clinical chart diagnoses. *Psychiatry Research, 101*(1), 47-54.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, *35*(2), 299-331.

Reed, S. J., Shore, K. K., & Tice, J. A. (2016). Effectiveness and value of integrating behavioral health into primary care. *JAMA internal medicine*, *176*(5), 691-692.

Reiss-Brennan, B., Brunisholz, K. D., Dredge, C., Briot, P., Grazier, K., Wilcox, A., ... & James, B. (2016). Association of integrated team-based care with health care quality, utilization, and cost. *Jama*, *316*(8), 826-834.

Rehm, J., & Arminger, G. (1996). Alcohol consumption in Switzerland 1987-93: adjusting for differential effects of assessment techniques on the analysis of trends. *Addiction*, *91*(9), 1335-1344.

Rehm, J., & Spuhler, T. (1993). Measurement error in alcohol consumption: the Swiss Health Survey. *European journal of clinical nutrition*, *47*, S25-30.

Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. *Archives of general psychiatry*, *50*(2), 85-94.

Reilly-Harrington, N. A., DeBonis, D., Leon, A. C., Sylvia, L., Perlis, R., Lewis, D., & Sachs, G. S. (2010). The interactive computer interview for mania. *Bipolar Disorders, 12*(5), 521-527.

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*(3), 169-184.

Richardson, L. M., & Austad, C. S. (1991). Realities of mental health practice in managed-care settings. *Professional Psychology: Research and Practice, 22*(1), 52.

Robins, L. N., & Regier, D. A. (1991). *Psychiatric disorders in america: The epidemiologic catchment area study* Free Press.

Rogers, E. M. (2004). A prospective and retrospective look at the diffusion model. *Journal of health communication*, *9*(S1), 13-19.

Rokach, L., & Maimon, O. (2014). *Data mining with decision trees: Theory and applications* World scientific.

Salinsky, E., & Loftis, C. (2007, August). Shrinking inpatient psychiatric capacity: cause for celebration or concern. In *National Health Policy Forum Issue Brief* (No. 823, pp. 1-21).

Sattler, A. F., Ale, C. M., Nguyen, K., Gregg, M. S., Geske, J. R., & Whiteside, S. P. (2016). Use of evidence-based assessments for childhood anxiety disorders within a regional medical system.

Schmitz, N., Kugler, J., & Rollnik, J. (2003). On the relation between neuroticism, self-esteem, and depression: Results from the national comorbidity survey. *Comprehensive Psychiatry, 44*(3), 169-176.

Schneider, S., & Döpfner, M. (2004). Leitlinien zur diagnostik und psychotherapie von angst- und phobischen störungen im kindes-und jugendalter: Ein evidenzbasierter diskussionsvorschlag. *Kindheit Und Entwicklung, 13*(2), 80-96.

Schoevers, R. A., Smit, F., Deeg, D. J., Cuijpers, P., Dekker, J., van Tilburg, W., & Beekman, A. T. (2006). Prevention of late-life depression in primary care: Do we know where to begin? *American Journal of Psychiatry,*

Semler, G., Witichen, H., Joschke, K., Zaudig, M., Von Geiso, T., Kaiser, S., . . . Pfister, H. (1987). Test-retest reliability of a standardized psychiatric interview (DIS/CIDI). *European Archives of Psychiatry and Neurological Sciences, 236*(4), 214-222.

Shakeshaft, A. P., Bowman, J. A., & Sanson-Fisher, R. W. (1998). Computers in community-based drug and alcohol clinical settings: Are they acceptable to respondents? *Drug and Alcohol Dependence, 50*(2), 177-180.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry,* Sheehan, D., & Lecrubier, Y. (2002). MINI international neuropsychiatric interview for DSM-IV (english version 5.0. 0). *Tampa: University of South Florida,*

Shojania, K. G., & Grimshaw, J. M. (2005). Evidence-based quality improvement: the state of the science. *Health affairs*, *24*(1), 138-150.

Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 380-411.

Smith, T. L., Klein, M. H., & Benjamin, L. S. (2003). Validation of the wisconsin personality disorders inventory-IV with the SCID-II. *Journal of Personality Disorders, 17*(3), 173.

Smits, F., Smits, N., Schoevers, R., Deeg, D., Beekman, A., & Cuijpers, P. (2008). An epidemiological approach to depression prevention in old age. *The American Journal of Geriatric Psychiatry, 16*(6), 444-453.

Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry, 24*(5), 399-411.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III-R (SCID): I: History, rationale, and description. *Archives of General Psychiatry, 49*(8), 624-629.

Stalans, L. J., Yarnold, P. R., Seng, M., Olson, D. E., & Repp, M. (2004). Identifying three types of violent offenders and predicting violent recidivism while on probation: A classification tree analysis. *Law and Human Behavior, 28*(3), 253.

Stewart, R. W., Tuerk, P. W., Metzger, I. W., Davidson, T. M., & Young, J. (2016). A decision-tree approach to the assessment of posttraumatic stress disorder: Engineering empirically rigorous and ecologically valid assessment measures. *Psychological Services, 13*(1), 1.

Stirman, S. W., Crits-Christoph, P., & DeRubeis, R. J. (2004). Achieving successful dissemination of empirically supported psychotherapies: A synthesis of dissemination theory. *Clinical psychology: Science and practice*, *11*(4), 343-359.

Suppiger, A., In-Albon, T., Hendriksen, S., Hermann, E., Margraf, J., & Schneider, S. (2009). Acceptance of structured diagnostic interviews for mental disorders in clinical practice and research settings. *Behavior therapy*, *40*(3), 272-279.

Tang, N., Stein, J., Hsia, R. Y., Maselli, J. H., & Gonzales, R. (2010). Trends and characteristics of US emergency department visits, 1997-2007. *Jama*, *304*(6), 664-670.

Thomas, N. (2015). What's really wrong with cognitive behavioral therapy for psychosis? *Frontiers in Psychology, 6*, 323.

Tiet, Q. Q., Ilgen, M. A., Byrnes, H. F., & Moos, R. H. (2006). Suicide attempts among substance use disorder patients: An initial step toward a decision tree for suicide management. *Alcoholism: Clinical and Experimental Research, 30*(6), 998-1005.

Trent, L. R., Buchanan, E., Ebesutani, C., Ale, C. M., Heiden, L., Hight, T. L., ... & Young, J. (2013). A measurement invariance examination of the Revised Child Anxiety and Depression Scale in a southern sample: Differential item functioning between African American and Caucasian youth. *Assessment*, *20*(2), 175-187.

Üstün, T. B., & Sartorius, N. (1995). *Mental illness in general health care: an international study*. John Wiley & Sons.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.

Velting, O. N., Setzer, N. J., & Albano, A. M. (2004). Update on and advances in assessment and cognitive-behavioral treatment of anxiety disorders in children and adolescents. *Professional Psychology: Research and Practice, 35*(1), 42.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, *37*(5), 360-363.

Wacker, H., Battegay, R., Mullejans, R., & Schlosser, C. (1990). Using the CIDI-C in the general population. *Psychiatry: A World Perspective, 1*, 138-143.

Wang, P. S., Demler, O., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2006). Changing profiles of service sectors used for mental health care in the United States. *American Journal of Psychiatry*, *163*(7), 1187-1198.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of general psychiatry*, *62*(6), 629-640.

Warnick, E. M., Weersing, V. R., Scahill, L., & Woolston, J. L. (2009). Selecting measures for use in child mental health services: A scorecard approach. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(2), 112-122.

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-Administered PTSD scale: A review of the first ten years of research. *Depression and Anxiety, 13*(3), 132-156.

Weisz, J. R., & Addis, M. E. (2006). The research-practice tango and other choreographic challenges: Using and testing evidence-based psychotherapies in clinical care settings.

Whiteside, S. P., Sattler, A. F., Hathaway, J., & Douglas, K. V. (2016). Use of evidence-based assessment for childhood anxiety disorders in community practice. *Journal of Anxiety Disorders, 39*, 65-70.

Wittchen, H. (1994). Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): A critical review. *Journal of Psychiatric Research, 28*(1), 57-84.

World Health Organization. (1993). The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204-221. doi:10.1093/jpepsy/jst062 [doi]

Youngstrom, E. A., Jenkins, M. M., Jensen-Doss, A., & Youngstrom, J. K. (2012). Evidence-based assessment strategies for pediatric bipolar disorder. *The Israel Journal of Psychiatry and Related Sciences, 49*(1), 15-27.

Zimmerman, M., & Mattia, J. I. (1999). Axis I diagnostic comorbidity and borderline personality disorder. *Comprehensive Psychiatry, 40*(4), 245-252.

Zun, L. (2016). Care of psychiatric patients: The challenge to emergency physicians. *Western journal of emergency medicine*, *17*(2), 173.

VITA

Adam F. Sattler

2400 Anderson Road
Apartment 15
Oxford, MS 38655
asattler@go.olemiss.edu
701-391-4680

EDUCATION

Ph.D. in Clinical Psychology
University of Mississippi
Oxford, Mississippi
Advisor: John Young, Ph.D. L.P.
Degree Expected: 2019

Dissertation Title: *Validation of a Brief Screening Instrument for Psychopathology in Adults*
(Defended: May 2016)
Committee: John N. Young, Ph.D. L.P. (Chair)
John Bentley, Ph.D.
Danielle J. Maack, Ph.D.
Todd A. Smitherman, Ph.D.

M.A. in Clinical Psychology
Minnesota State University, Mankato
Mankato, Minnesota
Advisor: Jeffrey Buchanan, Ph.D.
Degree Awarded: May 2014

Thesis Title: *The Quality of Assessments for Childhood Anxiety Disorders within a Regional Medical System*
Committee: Stephen Whiteside, PhD, L.P. (Chair)
Jeffrey Buchanan, Ph.D.
Sarah Sifers, Ph.D., L.P.

B.A. in Psychology
Minor: Music
Saint John's University

Collegeville, Minnesota

LICENSES/CERTIFICATIONS

Provisionally Certified Mental Health Therapist (Jurisdiction: MS)
Examination for Professional Practice in Psychology (EPPP) *Passed at the Doctoral Level: October 2016*
Interdisciplinary Certificate in Applied Statistics (Expected May 2018)

CLINICAL EXPERIENCE

University of Mississippi: Graduate Therapist
Communicare: Community Mental Health Center
Water Valley, Mississippi
July 2017-present

Duties: Conducting intake assessments; developing treatment plans; reviewing and updating annual paperwork; providing individual and group therapy to underserved clients; consulting with interdisciplinary professionals.

Supervisors: Scott Gustafson, Ph.D., Dixie Church, LMFT

University of Mississippi: Graduate Assessor
Psychological Assessment Clinic
Oxford, Mississippi
May 2017-present

Duties: Providing comprehensive psychological evaluations to assess for learning disabilities, Attention-Deficit/Hyperactivity Disorder, mood/anxiety disorders, and personality disorders.

Supervisor: Scott Gustafson, Ph.D.

University of Mississippi: Graduate Therapist
Psychological Services Center
Oxford, MS
January 2015-present

Duties: Conducting intake assessments; developing treatment plans; providing individual therapy; preparing client process notes and reports.

Supervisors: Kelly G. Wilson, Ph.D., Danielle J. Maack, Ph.D., John N. Young, Ph.D.

University of Mississippi: Graduate Intern
Department of Education and Research, The Baddour Center
Senatobia, MS

July 2015-July 2016

Duties: Providing individual and group therapy; developing behavior plans; conducting assessments; consulting with interdisciplinary professionals.

Supervisors: Alan Gross, Ph. D., Shannon Hill, Ph. D.

Minnesota State University, Mankato: Graduate Assessor
The Assessment Clinic
Mankato, Minnesota
August 2013-May 2014

Duties: Providing comprehensive psychological evaluations to assess for learning disabilities and Attention-Deficit/Hyperactivity Disorder.

Supervisor: Sarah Sifers, Ph.D., L.P.

Chambers and Blohm Psychological Services Center: Assessment Intern
Assessment Services
Bismarck, North Dakota
August 2011-August 2012

Duties: Administering, scoring, and interpreting psychological tests for Attention-Deficit/Hyperactivity Disorder and learning disabilities.

Supervisor: Kathy Blohm, Ph.D., L.P.

RESEARCH EXPERIENCE

The University of Mississippi: Graduate Research Assistant
The Science Infusion That Helps Lab
Oxford, Mississippi
August 2014-present

Duties: Collecting, entering, and analyzing data; preparing manuscripts for publication; preparing research presentations; supervising various research projects; providing consultation on statistics and research design.

Supervisor: John N. Young, Ph.D., L.P.

The Mayo Clinic: Research Assistant
The Pediatric Anxiety Disorders Clinic
Rochester, Minnesota
June 2013-present

Duties: Accessing, managing, and analyzing data from health system's electronic medical record; preparing manuscripts for publication; preparing research conference presentations.

Supervisor: Stephen P. Whiteside, Ph.D., L.P.

PracticeWise: Literature Coding Consultant
Services and Product Development
Satellite Beach, Florida
September 2015-present

Duties: Coding randomized clinical trails for inclusion in the PracticeWise Evidence-Based Services Database of Clinical Trials.

Supervisor: Kimberly Becker, Ph.D.

Delta Autumn Consulting: Research Assistant
Research Division
Oxford, Mississippi
June 2015-present

Duties: Developing continuing education materials on various topics, including research design and statistics, telemental health, and ethics in psychology.

Supervisor: John N. Young, Ph.D., L.P., Danielle J. Maack, Ph.D.

Minnesota State University, Mankato: Data Analyst
Center for Excellence in Scholarship and Research
Mankato, Minnesota
August 2012-May 2014

Duties: Providing graduate students, staff, and faculty consultation on research design and statistics; provided consultation on various statistical packages, including R and SPSS; lead monthly workshops on research design and statistics; assisted in the development of an SPSS certification program for Graduate Assistants employed through the University.

Supervisor: In-Jae Kim, Ph.D.

Saint John's University: Undergraduate Research Assistant
Cognition and Emotion Lab
Collegeville, Minnesota
January 2010-May 2010

Duties: Assisting in study design and implementation; conducting literature review; administering and scoring psychometric instruments; collecting, entering, and analyzing data; poster development and presentation.

Supervisor: Michael G. Livingston, Ph.D.

Saint John's University: Undergraduate Research Assistant
Cognition and Perception Lab
Collegeville, Minnesota
January 2010-May 2010

Duties: Data collection; participant recruitment; data collection, entry, and analysis; poster development and presentation.

Supervisor: Benjamin Faber, Ph.D.

PROFESSIONAL PUBLICATIONS

Sattler, A. F., Whiteside, S. P., Bentley, J. P., & Young, J. (2017). 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*.

Sattler, A. F., Leffler J. M., Harrison, N. L., Bieber, E. D., Kosmach, J. K., Sim, L. A., & Whiteside, S. P. H. (in press). The Quality of Assessments for Childhood Psychopathology within a Regional Medical Center. *Psychological Services*.

Sattler, A.F., Tiede, M.S., Dammann, J.E., Jensen, A.H., Walton, G.A., & Whiteside, S.P. (2017). Utility of a Single-item Child Anxiety Rating for Use in Community Practice. *Professional Psychology: Research and Practice*, *48*(4), 259-266.

Whiteside, S.P., **Sattler, A.F**., Hathaway, J., & Douglas, K.V. (2016). Use of evidence-based assessment for childhood anxiety disorders in community practice. *Journal of anxiety disorders*, *39*, 65-70.

Sattler, A.F., Ale, C.M., Nguyen, K., Gregg, M.S., Geske, J.R., & Whiteside, S.P. (2016). Use of evidence-based assessments for childhood anxiety disorders within a regional medical system. *Psychological services*, *13*(4), 411.

Whiteside, S.P., **Sattler, A.F**, Ale, C.M., Young, B., Hillson Jensen, A., Gregg, M.S., & Geske, J.R. (2016). The use of exposure therapy for child anxiety disorders in a medical center. *Professional Psychology: Research and Practice*, *47*(3), 206.

Lombardi, N.J., Buchanan, J.A., Aflerbach, S., Campana, K., **Sattler, A.**, & Lai, D. (2014). Is Elderspeak Appropriate?: A Survey of Certified Nursing Assistants. *Journal of gerontological nursing*, *40*(11), 44.

MANUSCRIPTS IN PROGRESS/SUBMITTED

Sattler, A.F., Bentley J.P., Young J.N. Validation of a Brief Screening Instrument for Psychopathology in Adults.

SYMPOSIA/PRESENTATIONS

Sattler, A.F., & Young J.N. (2016, May) *Validation of a brief screening instrument for Psychopathology in Adults*. Presented at the 4th Annual University of Mississippi Conference on Psychological Science, Oxford, MS.

Sattler, A.F., & Young J.N. (2015, May) *Development of a brief screening instrument for Psychopathology in Adults*. Presented at the 3rd Annual University of Mississippi Conference on Psychological Science, Oxford, MS.

Sattler, A.F., & Tiede, M.S., Whiteside, S.P. (2015, November) *The quantity and quality of treatment for childhood anxiety disorders in clinical settings: Data from a large regional health system*. Symposium presentation at the 49th annual meeting of the Association for Cognitive and Behavioral Therapies, Chicago, IL.

Sattler, A.F., & Sifers, S. (2013, March). *Poverty status and utilization of the BackPack Food Program*. Minnesota Conference of Professional Psychology, Owatonna, MN.

Sattler, A.F., Lai, D., Afflerbach, S., Lombardi, N., & Buchanan J. (2013, March). *An examination of antecedent variables related to the use of elderspeak among nursing assistants*. Minnesota Conference of Professional Psychology, Owatonna, MN.

POSTER PRESENTATIONS

Sattler, A., & Livingston, M. (2010, April). *Experimental Disclosure in Adolescents:Effects on depression and loneliness*. Minnesota Undergraduate Psychology Conference. Northfield, MN.

Sattler, A., & Faber, B. (2010, April). *Failings of implicit theory of mind in adults*. Minnesota Undergraduate Psychology Conference. Northfield, MN.

AD-HOC REVIEWER

Behavior Therapy

TEACHING EXPERIENCE

University of Mississippi: Instructor of Record
PSY 201 *General Psychology*
2016-2017
Supervisor: John N. Young, Ph.D., L.P.

University of Mississippi: Teaching Assistant
PSY 202 *Statistics for Behavioral Sciences*
2014-2015
Supervisor: Scott A. Gustafson, Ph.D., L.P.

Saint John's University: Teaching Assistant
PSYCH 340 *Physiological Psychology*
2011

PROFESSIONAL REFERENCES

John N. Young, Ph.D., L.P.	jnyoung1@olemiss.edu University of Mississippi, Peabody Hall 309 University, MS 38677 662-915-1775
Stephen P. Whiteside, Ph.D.	Whiteside.Stephen@mayo.edu Mayo Clinic, West 11 Rochester, MN 55905 507-293-0089
Kelly G. Wilson, Ph.D.	kwilson@olemiss.edu University of Mississippi, Peabody Hall 311C University, MS 38677 662-816-5189
Todd A. Smitherman, Ph.D.	tasmithe@olemiss.edu University of Mississippi, Peabody Hall 310D University, MS 38677 662-915-1825