5-2017

Construct Validity of the Teate Depression Inventory and the State-Trait Inventory for Cognitive and Somatic Anxiety

Eleanor M. Crouse

Construct Validity of the Teate Depression Inventory and the State-Trait Inventory for Cognitive and Somatic Anxiety

An Honors Thesis submitted to the

Department of Psychology
College of Sciences
And
Honors College
Eastern Illinois University

In partial fulfillment of the requirements
Of the Departmental Honors Program
For the degree of

BACHELOR OF ARTS

In the Department of Psychology

May, 2017

By

Eleanor M. Crouse

Successfully defended this 25th day of April, 2017

Thesis Advisor:

Committee Member:

Honors Coordinator:

Department Chair:

Construct Validity of the Teate Depression Inventory and the State-Trait Inventory for Cognitive and Somatic Anxiety

Eleanor Crouse

Honors Thesis, Spring 2017

Dr. Canivez

Eastern Illinois University

5/2/17

Abstract

The current study replicated previous research by examining convergent and discriminant validity of the Teate Depression Inventory (TDI), the General Behavior Inventory (GBI), and the State-Trait Inventory of Cognitive and Somatic Anxiety (STICSA). Results supported convergent validity between the TDI Total score and the GBI Depression subscale and discriminant validity was supported by the correlations between the TDI Total and GBI Hypomania-Biphasic subscale and the STICSA State- and Trait-Somatic subscales. Interestingly, results supported convergent validity between the TDI Total score and the STICSA State- and Trait-Cognitive subscale. These results are discussed.

Construct Validity of the Teate Depression Inventory and the State-Trait Inventory of Cognitive and Somatic Anxiety

Major depressive disorder, major depression, depression, dysthymia, or bipolar/manic depression are mood disorders with high prevalence rates, a variety of symptoms, and a large impact on individuals and society. Lifetime prevalence rates have been estimated anywhere between 9.5% (Robins, 1991) to 20% (Birmaher, 1996). According to Birmaher (1996), the specific prevalence rate in children is between 0.4%-2.5% for both males and females, and in adolescents, it is between 0.4%-8.3%, with base rates being slightly higher in females. In young adults, aged 18-24, the prevalence rates for females and males are 19.2% and 13.5%, respectively. General adult prevalence rates are estimated as being somewhere between 9.5% (Robins, 1991) and 14.4% (Angst, 1995).

There are a variety of symptoms associated with major depressive disorder (MDD). The following is a list of symptoms that are observed in young adults and adults: sadness, irritability, lack of energy, difficulty making decisions, changes in sleep patterns, hopelessness, feelings of failure, feelings of punishment, changes in health concerns, suicidal ideation, changes in appetite, changes in self-image, difficulty working, loss of interest in sex, self-criticism, and crying. Not everyone diagnosed with MDD presents with each of these symptoms; likewise, presentation in males and females may differ. For example, females are more likely to report changes in self-image, loss of interest in sex, sadness, self-criticism, and crying; whereas males are more likely to report hopelessness, feelings of failure, feelings of punishment, suicidal ideation, changes in appetite, and changes in health concerns (Lopez Molina, 2014).

Major depressive disorder can have a variety of impacts on individuals and in turn, society. At the individual level, MDD influences school performance, relationships with others,

and overall health. Specifically in adolescents, young adults, and adults, MDD may precede the onset of substance or alcohol abuse, it may also increase suicidal behaviors, and it may increase risk of early pregnancy (Birmaher, 1996). At a societal/global level, MDD has an impact on medical resources and loss of productivity. Productivity losses are a result of absenteeism, early retirement, and premature mortality (Berto, 2000); it is also possible that decreased motivation plays a role in the decrease of productivity in the workplace. According to data presented by Berto (2000) the economic burden of depression is estimated to be in the millions and even billions of dollars in the U.S. and the U.K. in both direct and indirect costs. Furthermore, to increase the emphasis on the impact depression has as a significant disease; it was ranked sixth as an economic burden and third by prevalence (Berto, 2000).

Based on the evidence above for the enormous impact depression has on individuals and society, it is imperative that mental health professionals have the tools to effectively screen for, and measure, depression. Over the years various measures have been created based on different theories of depression. Typically these measures are self-report inventories, in which an individual rates their agreement with a statement or question regarding their affective state. Most notable are the Beck Hopelessness Scale (BHS; Beck & Steer, 1988) and the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), both of which were created prior to the year 2000. Another measure is the General Behavior Inventory (GBI; Depue, Slater, Wolfstetter-Kausch, Klein, Goplerud, & Farr, 1981), which was developed in the 1980s to measure symptoms of bipolar depression or manic depression. Finally, a new measure that has been developed and that shows promise to effectively measure and determine severity of depression is the Teate Depression Inventory (TDI; Balsamo & Saggino, 2013). Another newer measure that will be discussed is the State-Trait Inventory of Cognitive and Somatic Anxiety

(STICSA; Ree et al., 2008). The STICSA was developed primarily to improve on the psychometric weaknesses of the State-Trait Anxiety Inventory (STAI; Spielberger, 1983), namely its lack of unidimensionality. Both the STAI and STICSA are based on the theory of two types of anxiety: state and trait. State anxiety is an emotional state that changes over time and varies in intensity while trait anxiety is a relatively stable susceptibility to experiencing state anxiety (Grös et al., 2007). With the emergence of new measures for psychological constructs is the need to assess the psychometric properties of these measures, such as reliability, validity, and diagnostic utility.

The current study focused on the convergent and discriminant evidence for construct validity of the TDI using the STICSA and GBI as comparison measures. Pearson product-moment correlation coefficients were used to assess the convergent and discriminant validity. It was expected that the TDI Total and GBI depression subscale would have high correlations. Additionally, it was expected that the TDI Total and GBI hypomania-biphasic subscale and both STICSA subscales (Trait and State anxiety) would have relatively lower correlations. However, before the specifics of the current study are discussed, a literature review of the three measures, the TDI, the STICSA, and the GBI, is provided.

A brief outline of what will be covered for the TDI is as follows: the TDI was developed using the Rasch measurement model to avoid the psychometric weaknesses associated with Classical test theory that most scales are based on, the TDI has primarily been evaluated using clinical and nonclinical Italian adult samples, to date the TDI has shown divergent/discriminant validity with the STICSA, there is currently only one study where the TDI was used in the U.S. with a nonclinical young adult sample following translation into English, the TDI has shown good internal consistency and reliability, and lastly, the TDI has support for a three-factor

bifactor structure but should be interpreted as unidimensional through the total score. An outline for the STICSA includes: the STICSA has shown better divergent/discriminant validity with depression measures than the STAI, it has been evaluated using nonclinical and clinical samples, it has shown good internal consistency and reliability, and finally, the STICSA has support for a hierarchical model and a four-factor model. The outline for the GBI is as follows: the GBI was initially developed to identify unipolar affective disorders then was modified to identify bipolar forms of affective disorders, the GBI has high positive predictive power and high negative predictive power, it has been evaluated primarily with Caucasian samples, the GBI is considered to have the most robust psychometric properties of any self-report inventory, and the GBI has support for a two-factor model.

Research on the TDI

The Teate Depression Inventory (TDI; Balsamo & Saggino, 2013) is a newer depression inventory that was developed in Italy. It was developed using the Rasch measurement model rather than classical test theory (CTT), which has several psychometric weaknesses mainly arising from theoretical assumptions (Balsamo, Giampaglia, & Saggino, 2014). The central limitations of the CTT are: the method of scoring, the comparison of scores across varying samples, the total score method, and last is the scoring method. The method of scoring weighs each item equally even though not all items are representative of the same level of psychiatric severity. The comparison of scores across samples is inappropriate because items may or may not be equally effective across different samples. The total score method assumes each symptom that is measured on the scale is equally related to the construct they are proposed to measure; however, unfortunately unidimensionality is not the norm for most depression instruments. The scoring method assumes that the distance between ratings on items are of equal intervals when

they actually are not. Furthermore, some depression measures have been criticized for being long, labor-intensive, and exhausting for clients (Balsamo et al., 2014). The Rasch measurement model avoids some of the weaknesses of the CTT-based methods by allowing for the measurement of the performance of each individual item on the inventory. Additionally, it provides the ability to identify a core group of items with the best psychometric properties. Overall, the key strength of the Rasch model over the CTT is its capacity to provide a transformation of raw ordinal scale data into an interval scale. The Rasch model also provides for unidimensionality, which improves precision in measurement and interpretation (Balsamo et al., 2014).

Balsamo et al. (2014) conducted a study using Rasch analysis to select items for the TDI. The process began by administering a large pool of items to clinical and nonclinical samples with the purpose of finding items with strong psychometric properties. The pool of items was created using four steps. First was the generation of the preliminary item set using the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV-TR; American Psychiatric Association, 2000) criteria by five experts, psychiatrists and psychotherapists with a mean clinical experience of 26.02 ± 7.4 years. These preliminary items were negatively (representing presence of depression) and positively (representing absence of depression) worded, and they included simple and direct statements. This first set included 152 items. Second came the rating of the first set of 152 items by a different group of five independent clinicians with a mean clinical experience of 21.57 ± 4.8 years. This group of clinicians was asked to determine the correspondence between each item and the DSM-IV-TR diagnostic criterion using a five-point Likert scale ranging from 0 "not at all corresponding" to 4 "extremely corresponding." Items with a mean score of 2.5 or higher were retained for further rating and 41 items were then

deleted. The second step was completed by adding instructions and item response format where individuals rated each item on a five-point Likert scale from 0 "always" to 4 "never" to measure how much the symptoms were present over the past 14 days to establish consistency with the DSM-IV-TR criteria. The third step was a refinement of the initial pool of items. This time five psychometricians, with a mean clinical experience of 16.02 ± 5.6 years, independently rated the 111 items that were retained in terms of degree of clarity and unambiguity in representation of depressive symptoms and adherence to the proposed response format. Items with a mean score of 2.5 on a five-point Likert scale, ranging from 0 "not at all adequate" to 4 "extremely adequate," were retained and an additional 57 items were dropped. Finally, the remaining 54 items were randomly ordered for presentation on the assessment form. The fourth, and final, step in item selection was an examination of comprehensibility. Twenty nonclinical subjects (50% females; mean age 33.14 ± 10.58 years) and 20 outpatients (50% females; mean age 34.35 ± 5.25 years) with various psychiatric diagnoses read each item and evaluated its comprehensibility. Based on their evaluations, three items were deleted and four reformulated. The final item pool included 51 items, 36 negatively and 15 positively worded, with at least five assessing each DSM-IV-TR diagnostic criterion (Balsamo et al, 2014).

Following the creation of the 51 items they then were administered to a sample of 529 clinical and nonclinical participants. After this, the items were analyzed for best measurement properties to compose a brief, homogeneous, and unidimensional scale of depression (Balsamo et al, 2014). The first step in this procedure was to apply selection criteria, overall model fit, and individual item fit. Individual item fit was checked using χ^2 statistics and standardized residuals. Basically, this meant that the items were tested for the extent to which a set conformed to a single trait in the sample. Good fit would produce an overall model probability value that is

nonsignificant (with a Bonferroni correction). The next step was an assessment of whether the selected items conformed to Rasch model expectations (Balsamo et al, 2014). Several expectations were evaluated, starting with the Person Separation Index (PSI), consistent with Cronbach's α, whereby a minimum coefficient of 0.85 was considered good for clinical or individual use. Next came the category threshold parameter values, and if disordered thresholds were found then they might be rescored; then there were tests for local independence and unidimensionality, and if correlations exceed 0.3 they were assumed to indicate dependency. After the independence and unidimensionality tests, item bias (differential item functioning [DIF]) was evaluated to determine if groups with different characteristics responded differently to an item; and lastly, person location distribution was used to examine differences in levels of severity of depression.

Results of measures of expectations showed poor fit to model expectations. The overall model with Bonferroni correction for item-trait interaction did not exceed 0.05/51, PSI results were high, and item fit residuals ranged between +6.61 and -3.99 with eighteen items outside the acceptable range. Using the selection criteria for fit mentioned above, 30 items were removed due to poor fit, leaving a final set of 21 items that became the content of the Teate Depression Inventory. Based on the remaining items, all DSM-IV-TR criteria were represented with the exception of appetite disturbance and sleep disturbance, both of which are somatic/physiological. A total of 13 somatic-related items were removed, most likely because of their lack of unidimensionality. Finally, 10 of the 21 TDI items were positively worded (Balsamo et al., 2014).

The main purpose of the Balsamo et al. (2014) study was to apply and extend the Rasch model to a case of more than two ordered categories in order to select items in developing a new

self-report depression inventory. One of the analyses was to guarantee content validity, which is often lacking. The 21-item scale that fit the Rasch model was considered an "objective measure," and it is considered to be "sample free" and "test free" (Balsamo et al., 2014, p. 160). The Rasch-based TDI allows for the generation of a total summed score, expressed in logit units, as a true index of a person's severity of depression, which is easier to interpret and understand. The Rasch model also allows for the production of true interval scale data, this in turn allows for mathematical operations, statistical indicators, and quantitative comparisons between and within subjects. Several items with greater weight, such as suicidal ideation and self-shame, are stronger indicators of more severe depression. Because the TDI omits somatic items it is considered a unidimensional screening tool, and unbiased. More studies are required to examine how the TDI performs in different populations. It should be noted that the TDI is to be used to screen for the presence of depression and assess the severity of it; it should not be solely used to specify a diagnosis. This study did not specify cut off scores, therefore it was judged incomplete and recommend it should not be used by clinicians to identify patients as having depression. In conclusion, while Rasch analysis with the TDI requires more effort, it should be encouraged.

According to Balsamo, Romanelli, Innamorati, Ciccarese, Carlucci, & Saggino, (2013) self-report measures of anxiety and depression lack discriminant validity, possibly because of a lack of unidimensionality of the constructs. It has been theorized that self-report measures of anxiety and depression lack unidimensionality and instead measure general negative affect common to both. Despite these problems, many authors believe it is important to have a valid measure of anxiety. A commonly used measure is the State-Trait Anxiety Inventory (STAI; Spielberger, 1983). The STAI is composed of two subscales, a 20-item trait-anxiety scale and a 20-item state-anxiety scale. Balsamo et al. (2013) only examined the trait-anxiety subscale

(STAI-T). While during its development some items that measured depression were eliminated, others were retained under the theory that people with high trait anxiety are more discontent with themselves. Based on other studies and theories, support for the idea that the STAI-T measures depression and well-being along with anxiety has emerged, along with a need to clarify what exactly it actually measures.

Balsamo et al (2013) compared the TDI to the State-Trait Anxiety Inventory-trait scale (STAI-T) to determine whether the STAI-T was unidimensional or multidimensional. This was done by performing confirmatory factor analysis (CFA) and assessing internal consistency and convergent and discriminant/divergent validity. Balsamo et al (2013) included a total of 1,124 participants and each completed the three paper-and-pencil questionnaires in counterbalanced order to avoid order effects. These questionnaires included the STAI-T, the Beck Depression Inventory-II (BDI-II; Beck et al., 1996), and the TDI. Participants included both clinical and nonclinical samples. Overall, the STAI-T total score correlated most strongly with the depression measures (*r*s = .70-.76) than with another measure of anxiety (*r*s = .55-.61), the Beck Anxiety Inventory (BAI; Beck & Steer, 1993). Fitness of several structural models was assessed, including a one-factor model, a model with two correlated factors, a one-construct two-method model, a two uncorrelated-factors model, and a bifactor model. Several measures for fit of each model were examined. After models were determined for best-fit, their individual psychometric properties were assessed using Cronbach's alphas indices for internal consistency. Correlation coefficients were used to determine convergent and discriminant validity of the STAI-T and its factor subscales.

Results showed there was a statistical lack of fit for all models tested; most likely because of the large sample size and sensitivity of X^2 with large samples. However, based on the

goodness-of-fit indices, the one-construct two-method and the bifactor models were preferred because they showed stable fit across both the clinical and nonclinical samples. Based on the bifactor model, Cronbach's alpha for the "Anxiety" factor was .86 in the clinical sample and .87 in the nonclinical sample. Likewise, Cronbach's alpha for the "Depression" factor was .91 in the clinical sample and .89 in the nonclinical sample. Based on the one-construct two-factor model, Cronbach's alphas for Positive Polarity were .89 for the nonclinical sample and .88 for the clinical sample, and were .88 for the nonclinical and .89 for the clinical samples for Negative Polarity. The homogeneity of the separate factors was higher than that of the general factor. Correlations among the dimensions of the two best-fitting models and the STAI-T total score were high, despite the higher correlation between the STAI-T total score and the "Depression" factor. The STAI-T also correlated strongly with the concurrent measures of depression (*r* between .70 and .76) showing a lack of discriminant validity. The clinical and nonclinical samples differed significantly on all measures and factors from the CFA with large effect sizes, except for the "Anxiety" factor where the effect size was moderate. Overall, the clinical sample obtained higher scores compared to the nonclinical sample, as expected. The clinical sample was further divided into three diagnostic groups: those with anxiety disorders, those with mood disorders, and those with other specific DSM-IV-TR disorders. Based on mean differences between these three clinical groups on several factors, it appeared that factors believed to consistently measure depression were better able to discriminate between anxiety and depressive disorders (Balsamo et al, 2013).

Balsamo et al (2013) results suggested that studies challenging the unidimensionality of the STAI-T are correct in that it should be considered multidimensional, measuring two separate but complex correlated aspects of negative affect. In general, the STAI-T factors do not measure

anxiety or depression in the strict sense because of a lack of discriminant validity. Despite these less-than-positive results, the STAI-T was able to discriminate between the clinical and nonclinical samples. In conclusion, the STAI-T appears to measure a general vulnerability to psychological disorders but should not be considered an accurate assessment of anxiety as separate from depression. This study along with others that called into question the validity of the STAI (complete version) instigated the creation of a new measure of state and trait anxiety named the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA; Ree, French, MacLeod, & Locke, 2008) that will be discussed following the TDI.

Balsamo and Saggino (2014) examined the diagnostic utility of the TDI by identifying cut-off scores in order to differentiate various levels of depression severity. It has long been debated whether dimensional assessments truly reflect the nature of mental disorders; however, it is not so much a question of "either-or" as it is of when dimensional assessments are appropriate. This brings up the issue of the "cut-off point dilemma." Typically, cut-off scores are not ideal for instruments constructed using Rasch measurement models because they were created using traditional scoring methods. However, the TDI has recently been found to be a more accurate measure of depression than other, more commonly used measures. Balsamo et al. (unpublished data, 2014) conducted three studies on large nonclinical and clinical samples. In the first, internal consistency was high, with a Cronbach's alpha of 0.92. In the second, Cronbach's alphas were 0.94 and 0.92 for clinical and nonclinical samples, respectively; and correlation with the BDI-II was 0.73 for both groups. In the third, the TDI was administered to a sample of middle-aged and older adult population, which produced a Cronbach's alpha of 0.88 and a significant correlation with the Geriatric Depression Scale (GDS) ($r = 0.56, p < 0.01$). Correlations between the TDI and the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA; Ree et al., 2008), the

trait and state factors, were weaker than the correlations between the TDI and the GDS displaying better discriminant validity. Balsamo and Saggino (2014) is possibly the first study, or at least one of the first studies, to illustrate discriminant and convergent validity for the STICSA compared to the STAI and a measure of depression, namely the TDI.

Balsamo and Saggino (2014) sought to determine cut-off scores on the TDI that could differentiate varying levels of depression in a group of clinically diagnosed individuals. Each individual went through a structured clinical interview for DSM-IV-TR Axis I disorders (SCID-I; Mazzi et al., 2000) to assess the DSM-IV-TR diagnostic criteria. Following this interview the TDI was explained and administered. Several procedures were completed in order to develop a set of cut-off scores. First, based on the diagnosis from the SCID-I, patients were classified into four groups: (1) mildly depressed, (2) moderately depressed, (3) severely depressed, and (4) nondepressed. Second, ideal cut-off scores were developed using Receiver Operating Characteristic (ROC; Swets, 1988; Treat & Viken, 2012) curves. Three ROC curves were constructed: (1) nondepressed vs. mildly depressed; (2) nondepressed and mildly vs. moderately depressed; and (3) nondepressed, mildly, and moderately vs. severely depressed. The main outcome variable is the area under the ROC curve (AUC), which is a direct representation of the overall accuracy of the TDI in screening for depression. For this study, maximizing sensitivity, which is the probability of correctly classifying an individual as depressed or more depressed, and maximizing specificity, which is the probability of correctly classifying an individual as not depressed or less depressed, is equally important. The results of the AUCs with 95% confidence intervals suggested the TDI is a good scale for discriminating nondepressed from mild depressed patients and non- depressed and mildly depressed from moderately depressed. The third AUC

showed excellent predictive accuracy for discriminating the non- depressed, mildly depressed, and moderately depressed from the severely depressed.

Ideal cut-off values were chosen by analyzing the coordinates of each ROC curse. For the first ROC curve, the cut-off score of 21, produced sensitivity of 0.86, specificity of 0.94, and overall classification accuracy of 0.90. For the second ROC curve, the cut-off score of 35.5, produced sensitivity of 0.82, specificity of 0.98, and overall classification accuracy of 0.90. For the third ROC curve, the cut-off score of 49.5, produced sensitivity of 0.81, specificity of 0.94, and overall classification accuracy of 0.88. These cut-off scores created the guidelines for patients with major depression: scores of 0-21, depression is considered "minimal"; scores of 22-36, depression is considered "mild"; scores of 37-50, depression is considered "moderate"; and scores of 51-84 (84 is the highest possible score), depression is considered "severe" (Balsamo & Saggino, 2014).

According to Balsamo and Saggino (2014) sensitivity and specificity are the most evident markers as a method of determination for cut-off scores. Typically sensitivity and specificity are given equal weight because the costs of false negatives and false positives are equally serious. In this study, the weighing of both equally in determining cut-off scores was supported by the results. The cut-off scores may be used differently depending of the purpose of the TDI for a specific sample. It should be noted that further research with more diverse samples is needed, particularly for information on external validity.

Balsamo et al. (2015) evaluated the construct validity of the Other As Shamer (OAS) scale using CFA, and also investigated the psychometric properties of the Italian version of the OAS, examining the clinical significance of the resulting factors and correlations with the BDI-II and TDI. Additionally, the sensitivity and specificity of the OAS in determining if nonclinical

subjects met clinical thresholds for depression and ROC curves were assessed. This study consisted of 687 Italian participants from the general population, with a little over half being undergraduate psychology students. A subsample of 70 students was asked to complete questionnaires ten days later to evaluate test-retest reliability.

All participants were volunteers from a sample of convenience. All measures were administered in Italian. The items of the OAS specifically were translated into Italian according to the standard procedures of forward and back-translation. Each translation was performed by two independent bilingual professionals. Following these translations, the questionnaire was given to fifteen PhD students who commented on any items with awkward or ambiguous wording. Based on their assessments, some items were slightly modified without changing the original meaning. All the participants completed three paper-and-pencil scales administered in random order. The OAS includes items rated on a five-point Likert-type scale, ranging from 0 (never) to 4 (almost always), and the maximum score was 72, indicating great external shame. In this study, Cronbach's alpha for OAS was .87. The BDI-II uses a four-point Likert scale ranging from 0 to 3; and Cronbach's alpha in this study was .82. The TDI uses a five-point Likert-type scale, ranging from 0 (always) to 4 (never), and Cronbach's alpha was .92, which was considered excellent.

Previous studies have focused on using the total OAS score for analysis. Balsamo et al (2015) decided to test a different hierarchical factor model. A number of fit indices were used to determine goodness-of-fit for models. Based on these indices, Model 2 was retained, and included one second-order factor, the total OAS score, and three first-order factors, inferiority, mistakes, and emptiness. Internal consistency of the OAS total score with the original version was high, internal consistency of the subscales was good, and the test-retest coefficient for the

total score was very good. The subscales were significantly correlated (hence the second order structure) and they were correlated with the OAS total score. The pattern of correlations between the OAS total score and its subscales with the depression measures showed that all OAS scores were significantly correlated with the BDI-II and the TDI. The emptiness subscale was the most highly correlated with depression severity, particularly on the TDI. "The TDI cutoff criterion of 36 for both the OAS scale and its three subscales" (Balsamo et al, 2015, p. 85) was used along with ROC curves to compare the nondepressed and nonclinical depressed groups. Results indicated the OAS scale and its subscales were able to discriminate between the two groups.

This study used the OAS to measure external shame in relation to self-other perception and evaluate the association this shame has with depressive symptoms. The results showed there was a significant association between depression and external shame. Specifically, of the OAS's factors, emptiness had the strongest correlation with measures of depression. Additionally, the emptiness factor seemed to perform best at discriminating between depressed and nondepressed participants in this particular study (Balsamo et al., 2015).

Balsamo, Carlucci, Sergi, Murdock, and Saggino (2015) explored the role of schema domains in relation to co-rumination and symptoms of depression among nonclinical young adults. Their hypothesis was that the statistically significant relationship between concurrent depression and self-reported co-rumination would disappear when they controlled for the maladaptive cognitive schemas. This would show that co-rumination contributes to depression because it activates the maladaptive cognitive processes. They also hypothesized that there would be gender differences.

Balsamo et al. (2015) included 461 Italian participants, with slightly over half being female. All participants were given the Italian versions of the Co-Rumination Questionnaire

(CRQ; Rose, 2002), the TDI, and the Young Schema Questionnaire Long Form, Third Edition (YSQ-L3; Young, 2003). All questionnaires were given in paper format in a fixed order. They were all administered by three licensed psychologists who had received specific training for the study.

Zero-order correlation coefficients were calculated among all the measures. Absolute correlations of .30 or greater were considered salient given that the probability was influenced by the sample size. Baron and Kenny's causal-steps method was used to test the mediation hypotheses (Baron & Kenny, 1986). This method uses a series of regression analyses to go through the steps of mediation testing. Once a path is reduced in absolute size, the test of Sobel is used to indicate partial mediation (Sobel ME, 1982). Gender differences were examined separately for levels of co-rumination, depression, and salient cognitive schema domains.

Results indicated that the relationship between depression and co-rumination was significant. Among the YSQ-I domains, only two, Overvigilance/Inhibition and Other-Directedness, were discovered to have salient correlations with the CRQ score. Depression was also only correlated with the same two YSQ-I domains, Overvigilance/Inhibition and Other-Directedness. There were also gender differences in the strength of the correlations between depression and co-rumination, significant for females (r = .139) and nonsignificant for males (r = .027). The YSQ-I domain of Overvigilance/Inhibition was determined to completely mediate the relationship between depression and co-rumination, as was the domain of Other-Directedness. However, this mediation was only found in the female sample (Balsamo, Carlucci et al., 2015).

Balsamo, Carlucci et al. (2015) seem to be the first to investigate the role cognitive schema domains have in the depression-co-rumination relationship. The results indicated that all five schemas on the YSQ-I were correlated with co-rumination, with Overvigilance/Inhibition

and Other-Directedness having the highest. Further model analysis showed these two domains fully mediated the relationship between depressive symptoms and co-rumination in non-clinical young adults. Combined, these two YSQ-I schemas focus on meeting others' needs and suppressing one's own emotional expression. The results of this study indicated that people who engage in co-rumination may be more vulnerable to maladaptive schemas that in turn elevate levels of depressive symptoms. Furthermore, the results suggest co-rumination is more prevalent and may have a strong role with depression in women.

Ruan, Pendergast, Dixon, Liao, Jones, and von der Embse (2016) utilized a sample of American college students to evaluate the factor structure of the TDI. This evaluation was done using EFA and CFA among nonclinical samples. This study was the first in the United States and with an English version of the TDI. Ruan et al. (2016) hypothesized that the TDI scores would support its technical adequacy and show a moderate correlation with another measure of depression as a display of concurrent validity. A total of 409 young adults from a variety of undergraduate courses served as the sample and was randomly split into an EFA sample and a CFA sample.

The original TDI was translated into English according to a translate-retranslate protocol. First the original items from the Italian version were translated into English by a bilingual psychologist. Second, a proficient bilingual speaker with a background in mental health translated the English version back into Italian. Third, the authors of the TDI reviewed the quality of the back translation and compared it with the original version. Lastly, the authors of the TDI, who are bilingual, and a team of psychologists, evaluated the English version.

All participants were administered the TDI. Cronbach's alphas were .942 (EFA sample) and .945 (CFA sample). Participants were also given the Center for Epidemiologic Studies

Depression Scale-Revised (CESD-R) to examine concurrent validity. Data collected for this particular study was collected as part of Project METS, which was a larger project. All responses were collected anonymously through an online survey tool.

The EFA examined the factor structure of the TDI using multiple statistical analyses to determine the adequacy of the sample, the sufficiency of the correlation matrix for factor analysis, and factor retention. An exploratory bifactor analysis was also conducted. Reliability of the general and second-order factors was also calculated. Following the exploratory bifactor analysis, CFA was conducted with the designated CFA sample using only the items retained from the EFA. A one-, two-, and three-factor bifactor models were examined. Model fit was also evaluated using several goodness-of-fit indices.

Results of Ruan et al. (2016) indicated the sample was adequate and there were sufficient correlations for factor analysis. Based on factor extraction, a two- or three-factor structure was possible, thus both were examined. Following these examinations, the three-factor structure was further evaluated. According to the Omega-hierarchical coefficient, the total score, rather than the subscale scores should be utilized for interpretation. The three-factor structure conflicts with previous research that suggested the TDI had a "unidimensional enough" structure to render interpretation of the total score appropriate. However, some modifications had to be made regarding correlated residuals so the three-factor bifactor model would fit. According to a bivariate Pearson correlation, the concurrent criterion-related validity of the TDI total scores compared to the CESD-R total scores was significant ($r = .81$, $p < .001$).

Ruan et al. (2016) examined the psychometric qualities of the TDI in American young adults. Results supported a three-factor bifactor structure, showed excellent reliability, and gave initial support for concurrent validity in a non-clinical sample. The total score of the TDI showed

excellent internal consistency, which was consistent with previous research; additionally, its subscales showed acceptable to excellent inter-item reliability. Overall, the TDI appears to be a precise measure when using the total score for interpretation. Ruan et al. (2016) also extended the applicability of the TDI by translating it into English and examining it with a different statistical framework, such as classical test theory. The use of another framework allowed for the display of TDI's robustness; it focuses on affective and cognitive symptoms as opposed to somatic symptoms, which may allow for the application of it in populations with chronic health conditions. This study was also the first to evaluate the psychometric qualities of the TDI in a U.S. general population. The greatest strength of doing so was the representation of gender and an ethnically diverse group. Lastly, by adapting the TDI for use in the U.S. it could facilitate cross-cultural comparison (Ruan et al., 2016).

Research on the STICSA

A distinction between two different types of anxiety was first introduced by Cattell (1966), and was later expanded on by Spielberger (1983) who said that state anxiety was an emotional state that varies in intensity and changes over time. Contrastingly, trait anxiety is a stable susceptibility to frequently experience state anxiety (Grös, Antony, Simms, & McCabe, 2007). Spielberger (1966, 1972) elaborated on his model and described how an anxious state is initiated by an external stressor or internal cue. The anxious state in characterized by thoughts of impending doom and physiological arousal. As a means of investigating his model, Spielberger (1983) created the State-Trait Anxiety Inventory (STAI). The STAI has two 20-item self-report measures, one for state anxiety and one for trait anxiety. While the STAI has shown generally positive psychometric qualities, it has been criticized for its inability to discriminate anxiety from

depression, for a lack of positive psychometric qualities in younger, undereducated populations, and the two-factor structure of anxiety-absent and anxiety-present.

The above criticisms; in particular the inability to discriminate anxiety and depression from each other, led to the development of a newer state and trait anxiety measure. Ree, MacLeod, French, and Locke (2000) developed a newer anxiety measure: the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). This new measure is based on Spielberger's theoretical foundations of state and trait anxiety and has the same format as the STAI in terms of separate scales for State and Trait anxiety. The STICSA State is a 21-item self-report scale that assesses how a respondent feels at that moment, while the STICSA Trait (also 21-items) assesses how a respondent feels in general. Both scales use a 4-point Likert scale, ranging from 1 (not at all) to 4 (very much so) (Grös et al., 2007).

While the STAI and STICSA share many features, the STICSA was constructed to improve on the shortcomings of the STAI. The main improvement was of the structure; instead of distinguishing between anxiety-absent and anxiety-present, the STICSA distinguishes between cognitive and somatic anxiety symptoms. This distinction is not a new theory and has previously been adapted to other measures. Additionally, the STICSA was created to better discriminate between anxiety and depression by providing a more accurate measure of pure anxiety by favoring symptoms unique to anxiety and avoiding nonspecific symptoms.

Ree, French, MacLeod, and Locke (2008) conducted a series of studies in developing the STICSA and validating it. These studies were conducted several years before they were published (see Ree, 2000). A fair amount of attention has been given to the distinct dimensions of state and trait anxiety; however, while attention has been given to the distinct dimensions of somatic and cognitive anxiety, not much attention has been given to including these dimensions

in assessment. The somatic factor includes symptoms of hyperventilation, trembling, sweating, palpitations, muscle tension, and stiffness. The cognitive factor includes symptoms of worry, intrusive thoughts, lack of concentration, and any other symptoms associated with thought processes. Because of a lack of inclusion of these two dimensions in assessment there is little data for the validity of these distinctions within trait anxiety. If these dimensions are found to be valid they could enhance prediction of an individual's state anxiety response. Ree et al. (2008) explicitly tested the validity of the distinction between trait cognitive and somatic anxiety; and investigated whether these dimensions predicted anxiety responses.

The first study in the Ree et al. (2008) series was the development of the new anxiety measure. It was initially supposed to be a single set of items used to distinguish cognitive and somatic anxiety on a trait scale and state scale with the distinction being made in the instructions. A starting pool of 131 items considered to reflect the dimensions was created by clinical psychologists, clinical graduate students, and research psychologists. This pool was then inspected by three clinical psychology graduate students who were given definitions of cognitive and somatic anxiety. Another eight clinical graduate students rated each items in terms of how clearly and unambiguously it represented the two dimensions. From this, 62 items were judged to be distinctive and these were incorporated into the preliminary questionnaire. The preliminary questionnaire consisted of first person statements in order to enhance the description of the symptoms. Respondents rated each item on the trait scale from 1 (almost never) to 4 (almost always). Respondents rated each item on the state scale from 1 (not at all) to 4 (very much so). The two scales were counterbalanced. Participants came from a variety of samples, including senior high school students, businessmen, and university staff. Of the 2,500 distributed questionnaires, 576 were completed and returned. The demographics of these 576 responses are:

mean age of 34, range of 16-82 years. A little over 50% were female and 30 did not identify their gender.

After the administration of the preliminary questionnaire the items were statistically evaluated using the scores on the trait scale. To exclude items due to floor and ceiling effects items were retained if they had a mean score between 1.25 and 3.75. Also, to reduce item redundancy, those with a correlation above .45 on the same scale were examined to determine if the origin was because of highly similar item content, if so, the item with the closer to mid-point range was retained. This evaluation resulted in the retention of 26 items, 14 cognitive and 12 somatic.

Factor analysis of the trait scale showed a correlated two-factor model with the 26 items. A one-factor model did not acceptably fit the data. Items that cross-loaded on both factors were eliminated, resulting in the deletion of four cognitive items and one somatic item. The final 21-item data set fit well with the correlated two-factor model with no items cross-loading. Factor analysis of the state scale, using 21-items, showed a correlated two-factor model fit the data well with no cross-loading. This study displayed strong support for the validity of a distinction between the cognitive and somatic dimensions within state and trait anxiety.

The second study aimed to replicate the factor structure of the trait and state scales found in Study 1. In order to test generalizability the replication in this study was done with a population that differed from the one used in Study 1, and so because students were easily accessible, they were utilized (Ree et al., 2008). A total of 941 undergraduate psychology students participated; with 712 being administered the trait scale and 229 being administered the state scale.

The results of the STICSA trait scale CFA were obtained from a sample size of 687 following elimination of missing values. A correlated two-factor model fit the data well with no cross-loading of items. All loadings on the specified factor were significant at an alpha level of .01. A comparison of the correlated two-factor model with a one factor model and orthogonal model showed a significant chi-square difference between the correlated two-factor model and the latter two models in favor of the correlated two-factor model (Ree et al., 2008). The correlation between the cognitive and somatic factors of the trait scale was .59, and the coefficients of determination were 0.94 and 0.95 for the trait somatic and trait cognitive scales respectively, indicating high reliability (Ree et al., 2008). The correlation of .59 is actually more indicative of discriminant validity between the STICSA trait cognitive and somatic factors, not reliability of the scale.

The results of the STICSA state scale CFA were obtained from a sample size of 225 following elimination of missing values. A correlated two-factor model fit the data well with one item cross-loading. The same comparisons between the correlated two-factor model and a one factor model and orthogonal model were conducted and showed favor to the correlated two-factor model. The correlation between the cognitive and somatic factors of the state scale was also .59. The coefficients of determination were 0.94 and 0.92 for the state cognitive and state somatic scales respectively, indicating high reliability (Ree et al., 2008). As stated earlier, this correlation, also .59, is an indicator of discriminant validity; here it is between the STICSA state cognitive and somatic factors.

Correlations between the STICSA and measures of depression and anxiety were calculated to determine convergent and discriminant validity. The method used to do so was recommended by Meng, Rosenthal, and Rubin (1992), which involves Fisher Z transformations

of the correlation coefficients in order to compare them via a *t*-test. Based on this method it was shown that the STICSA-state scale correlated higher with the STAI-state scores than the BDI-II scores. The STICSA-trait scale correlated higher with the STAI-trait scores than the BDI-II scores. This is evidence for convergent and discriminant validity; the STICSA correlates more highly with another measure of anxiety and less so with a measure of depression.

This second study replicated the findings of the first study through CFAs on the trait and state scales of the STICSA in support of two factors within both scales: a cognitive dimension and a somatic dimension. These dimensions allow for a better explanation of the data when present (Ree et al., 2008). This replication further confirms the structural validity, not reliability as it was referred to as before, of the two dimensions at both state and trait levels of anxiety.

The third study further determined the nature of the construct validity of the state and trait scales of the STICSA by answering two questions: first, do the state scales detect increases in state anxiety when participants are in a situation known to increase anxiety?; second, does the initial level of trait cognitive or trait somatic anxiety predict later changes in state cognitive and state somatic anxiety following exposure to a stressful situation? The obvious hypothesis is the trait cognitive would strongly predict the state cognitive while the trait somatic would strongly predict the state somatic (Ree et al., 2008).

For this third study, 129 undergraduate psychology students completed the STICSA trait and state scales during a neutral time period and again during a stressful time period, during end of the year examinations. The neutral time period was considered a baseline period and the end-of-the-year examinations was considered the exam period and will be hereafter referred to as such.

A three-way repeated measures ANOVA with assessment period, anxiety dimension, and questionnaire form as within-subject factors revealed a significant main effect of anxiety dimension, indicating cognitive scores were typically higher than somatic scores. The main effect of assessment period was also significant, with examination stress scores being higher than the baseline scores. There was also a significant two-way interaction between assessment period and questionnaire form. Simple effects analyses indicated that mean state scores at exam time were higher than baseline scores. There was no significant difference between trait scores at exam time and baseline. The next part was to see if the baseline trait scores could predict later state scores form baseline to exam time. Sequential multiple linear regression analysis was performed to determine this. Because baseline scores were important, and so they did not hinder the interpretation of the variance based on the trait scales, they were always entered during the first step of the regression. A total of four multiple regression analyses were performed. The trait cognitive score at baseline predicted 5.2% of the variance in state cognitive scores during exam time; these scores also predicted an additional 5.4% of the variance in the state somatic scores during exam stress. The trait somatic score did not account for any significant amount of variance for either of the state scale scores during exam time.

In conclusion, the trait cognitive scale scores at baseline predicted increases in both state cognitive and somatic scores during exam time. Trait somatic scale scores at baseline did not predict any increases and may be a poor predictor of state anxiety responses. However, it could be the trait scales of the STICSA only predict anxiety responses in specific stressful situations. It is also possible the trait cognitive scale predicts general state anxiety responses to cognitive stressors while the trait somatic scale predicts general state anxiety responses to somatic stressors (Ree et al., 2008).

The fourth, and final study of the series, investigated the ability of the STICSA-trait scales to predict state anxiety responses to cognitive and somatic stressors. The design separates these two types of stressors in order to determine if it predicts the stressor being measured, cognitive or somatic. The cognitive stressor in this study was impending university exams while the somatic stressor was inhalation of CO_2-enriched air. The sample consisted of 42 undergraduate psychology students. They were individually tested for the laboratory aspect of the study, which was the administration of CO_2-enriched air. Each participant was also given the option to complete the questionnaire-based aspect of the study, which was to complete the STICSA under neutral and stressor conditions. A total of 32 participants completed both aspects of the study.

The state and trait scales for the laboratory aspect were completed with paper and pencil while for the questionnaire-based aspect (cognitive stressor) they were completed over the internet. For the somatic component, medical air (control) and 5% CO_2-enriched medical air were administered via a tube with a mouthpiece from a cylinder that was kept out of sight. The testing room was small with a few pieces of furniture and a one-way mirror that allowed the participants to be monitored without the experimenter being in their presence. For the cognitive stressor component, the participants completed the scales over the internet during a neutral period and then several weeks later during their final exam period. The somatic component was completed over two weeks during the neutral period of the cognitive stressor component. The participants completed the trait scale before being administered the CO_2-enriched air, then following the administration they completed the state scale.

Results of a two-way repeated measure ANOVA on STICSA state scores with time and anxiety dimension as factors showed a significant main effect of time. The scores obtained

during the exam period were higher than those obtained during the baseline. The main effect of anxiety dimension was also significant, with state cognitive scores being higher than state somatic. Following this, it was determined through sequential multiple linear regression analyses that the STICSA-trait cognitive scores at baseline predicted a significant amount of variance in state somatic scores at exam time. Trait cognitive scores at baseline also predicted much variance (37%) in state cognitive scores at exam time. Contrastingly, trait somatic scores did not predict any significant amount of variance for cognitive or somatic scores at exam time (Ree et al., 2008).

Results of a two-way, repeated measure ANOVA of state scores with manipulation phase and anxiety dimension as factors showed a significant main effect of manipulation phase. The scores obtained post CO_2 exposure were higher than those obtained at baseline. The interaction was also significant, revealing that cognitive and somatic scales changed from baseline to post CO_2. This increase in somatic scores was significant. Following this, it was determined through sequential multiple linear regression analyses that the STICSA-trait somatic scores at baseline predicted a significant amount of variance in state somatic scores post CO_2-inhalation. Trait somatic scores at baseline also predicted much variance in state cognitive scores post CO_2-inhalation. Contrastingly, trait cognitive scores at baseline did not predict any significant amount of variance for cognitive or somatic scores post CO_2-inhalation.

To summarize, this fourth study discovered that baseline trait cognitive scores of the STICSA predicted a significant amount of variance in both cognitive and somatic state anxiety responses under exam stress. Also, baseline trait somatic scores predicted a significant amount of variance in both cognitive and somatic state anxiety response after inhalation of CO_2-enriched air.

Overall, the first two studies in this series describe the development of the STICSA and suggested it produced reliable and valid scores of anxiety in a nonclinical population. Additional CFAs confirmed the factor structure of the trait and state scales in a clinical population. The second two studies suggested that trait cognitive and somatic anxiety represent the type of stressor an individual will show elevated state anxiety in. Lastly, the trait cognitive scale predicted state anxiety responses to cognitive stressor while the trait somatic scale predicted state anxiety responses to somatic stressors (Ree et al., 2008).

Grös et al. (2007) compared the construct validity of the STICSA to the STAI, with emphasis on discriminant validity. Internal consistency, convergent and divergent validity, and the factor structure in a patient sample was evaluated by Grös et al. (2007). Furthermore, the ability of the STICSA to discriminate between individuals with heightened chronic anxiety and less severe anxiety was also investigated by comparing the scores of the patient sample to a comparison sample. Lastly, the STICSA was compared to the STAI to investigate whether it had greater discriminant validity by comparing the scores of both measures with the two subscales of the Depression Anxiety Stress Scales (DASS; P. F. Lovibond & S. H. Lovibond, 1995). It was hypothesized that the STICSA Trait would have more positive correlations with the Anxiety scale, and less positive correlations with the Depression scale, of the DASS.

The Grös et al. (2007) sample consisted of 567 psychiatric outpatient participants from an anxiety treatment and research facility. A modified version of the SCID-IV was administered to establish diagnoses. A sample of 311 undergraduate psychology students was used to create a nonclinical comparison group. Initially 709 patients completed the numerous scales used in this study but those with substantial missing data (17% of participants) were not included, thus resulting in a total of 567 patients. The main difference between those who were included and

those who were not is the latter had significantly higher state anxiety. At the time of this study the STICSA was unpublished. Also, each of the subscales consists of 21 items rated on a 4-point Likert scale.

As part of replicating previous studies on the psychometric qualities of the STICSA, CFAs were conducted to confirm the factor structure of the STICSA's subscales; additionally, Cronbach's alpha coefficients were calculated. The ability of the STICSA subscales to discriminate the patients from the comparison group was tested through a series of one-way ANOVAs, followed by Tukey-corrected post hoc tests. The convergent and divergent validity of the STICA was evaluated through correlations between the STICSA, STAI, and DASS.

Results of the CFAs from both samples: patient and comparison, showed that of the four models tested, the four-factor model (State-Cognitive, State-Somatic, Trait-Cognitive, Trait-Somatic) fit best, as implied by the item pool and instructions. All factor loadings were moderate to high, and factor intercorrelations were statistically significant and consistent with the predicted pattern; the two trait factors were highly correlated, the two cognitive factors were highly correlated, the two somatic factors were highly correlated, and the two state factors were highly correlated. All subscales had excellent internal consistency, with the coefficients being slightly weaker for the comparison sample. Convergent and divergent validity was only investigated in the patient sample. The STICSA showed strong correlations to the corresponding scales of the STAI and the DASS scales. The STICSA Trait was more highly correlated with the DASS-A than the STAI Trait, and the STAI Trait was more highly correlated with the DASS-D. Similar patterns were found for the State subscales of both inventories. In sum, the STICSA had stronger correlations with another measure of anxiety while the STAI had stronger correlations with a measure of depression. Results of the discrimination tests showed the four patient groups

(OCD, PD, SP, and another anxiety or mood disorder) scored significantly higher on both the STICSA Trait and State than the comparison group. This indicated the STICSA was sensitive to the differences and may be able to detect distinctions within various anxiety disorders (Grös et al., 2007).

The findings of Grös et al. (2007) are replications of previous findings on the STICSA and its psychometric qualities, and extended them to a patient population which further supports its reliability and construct validity. The STICSA demonstrated better differentiation of convergent and discriminant validity than the STAI. Factor analysis of the STICSA (both subscales) supported previous findings of four factors: State-Cognitive, State-Somatic, Trait-Cognitive, and Trait-Somatic. The STICSA subscales displayed good internal consistency and significant correlations with the appropriate subscales on the STAI and DASS with the patient sample. Given that the STICSA was developed to improve on the STAI and to have a better pattern of convergent and discriminant validity, the results of Grös et al. (2007) support these purposes. A final note, the findings of Grös et al. (2007) could be discussed within the context of the tripartite model of anxiety and depression. Clark and Watson (1991) developed the model to explain the relationship between anxiety and depression. The model states that the two disorders share an unspecified component of general distress, called "negative affect," which partly explains the overlap. The model also considers two other components that are unique to depression or anxiety. Physiological hyperarousal is unique to anxiety while low positive affect, or anhedonia, is unique to depression (Grös et al., 2007). The components of this tripartite model could help explain how the STICSA was developed and provide a foundation for creating other tests.

Roberts, Hart, and Eastwood (2015) investigated the factor structure and reliability and convergent and divergent/discriminant evidence for validity of the STICSA in a sample of undergraduate students. The participants completed a number of measures online, including scales of depression, affect, and social desirability. At this time there were no standardized norms for the STICSA. Previous research on factor structure has indicated good fit for a two-factor somatic-cognitive model separately for the state and trait items. However, a four-factor state-trait somatic-cognitive model was suggested by Grös and Colleagues (2010), and analysis showed adequate fit with the data. Previous research on reliability indicated the subscales had adequate internal consistencies across various samples. Also, test-retest with a 2-month time lapse show adequate correlations for trait somatic and cognitive scores. Previous research has shown the STICSA to have good convergent and divergent validity based on correlations between the STICSA and the STAI, DASS, and BDI-II (Roberts et al., 2015).

Although previous research has shown favor to the STICSA, more was needed to further test the factor models and to provide more evidence for convergent and divergent validity across different samples. Roberts et al. (2015) aimed to validate test score interpretations of the STICSA and its subscales and to compare it to a more commonly used self-report measure of anxiety, the STAI.

A sample of 585 college students completed the study online. Participants with minimally completed data on the STICSA were deleted, a total of 25, leaving a total of 560 to be used. Those who were deleted did not significantly differ from those retained (complete data). Participants completed all of the following measures in addition to the STICSA: the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983), the Cognitive Somatic Anxiety Questionnaire (CSAQ; Schwartz et al., 1978), the Trimodal Anxiety Questionnaire (TAQ; Lehrer

& Woolfolk, 1982), the Anxiety Sensitivity Index-3 (ASI-3; Taylor et al., 2007), the Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991), the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996), the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977), the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), and the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1994). All measures were administered using Qualtrics software (Qualtrics Labs, Inc., 2009).

Results of the factor analyses showed the two-factor model for the state and trait items had a marginal fit to the data with all factor loadings being significant ($p < .001$). The four-factor and hierarchical models for the subscales combined fit the data well. Standardized factor loadings for both models were significant. Further analyses of the hierarchical model to determine if a global anxiety factor was uniquely correlated with other variables showed it was more highly correlated with scores on a measure of global anxiety as opposed to specific factors. These results support justification for a hierarchical model with a global anxiety factor. Scores on the state and trait versions, as well as the cognitive and somatic subscales of the two versions, had good internal consistencies. Also, total scores on the two versions had strong correlations, as did the cognitive and somatic subscales of the trait and state versions.

Pearson product-moment correlations displayed strong positive correlations between the state version of the STICSA and the state version of the STAI, and a similar pattern was found for the trait versions. The scores on the cognitive subscales of the STICSA had stronger correlations with both versions of the STAI than did the somatic subscales of the STICSA. The scores on the trait somatic subscale of the STICSA had stronger correlations with somatic subscales of the ASI-3, TAQ, CSAQ ($rs \geq .50$), and the anxious arousal subscale of the MASQ

than did the trait cognitive subscale. In terms of divergent/discriminant validity, Pearson product-moment correlations showed the expected patterns between scores on the STICSA and scores on depression scales. STICSA trait scores were moderately to strongly positively correlated with several measures of depression (*r*s = .44-.64). Cognitive scores were significantly more strongly positively correlated with depression scores than were somatic scores (*r*s = .49-.65). The STICSA trait scores were weakly correlated with a measure of social desirability (*r*s = -.08 to -.17). The STICSA state scores were strongly positively correlated with scores on the negative affect subscale of the state version of the PANAS and a similar pattern was found for the trait version of both measures (*r* = .62). There was a weak negative correlation between the positive affect subscale of the state and trait versions of the PANAS with the state and trait versions of the STICSA (*r*s = -.05 & -.15) (Roberts et al., 2015).

When comparing the state and trait versions of the STICSA and STAI with all the other measures, it was revealed that the STICSA had stronger positive correlations with other measures of anxiety than did the STAI. There were no differences between the trait versions of the STICSA and the STAI and the negative affect subscale of both versions of the PANAS. It was also revealed that scores on the STAI and its subscales were consistently more strongly correlated with scores on measures of depression than any scores on the STICSA, thus showing favor to the STICSA in terms of divergent/discriminant validity.

In conclusion, Roberts et al. (2015) provided additional evidence for the reliability and validity of test score interpretations of the STICSA. This study also provided support for a four-factor model of the STICSA with somatic and cognitive subscales on both the state and trait versions. There was also support found for a hierarchical model of the STICSA with a global factor and four specific factors. Given the theoretical nature of anxiety, the hierarchical model is

considered the best fit for the data. Overall the STICSA exhibited good internal consistencies, and the expected patterns of correlations with the appropriate measures to support convergent and divergent/discriminant validity. Specifically, in terms of divergent/discriminant validity between the STICSA and measures of depression, despite strong correlations between the trait version of the STICSA and measures of depression, the STICSA had relatively stronger correlations with other measures of anxiety, supporting convergent validity. When comparing the STICSA and the STAI, results indicated the STICSA to be a more valid measure of somatic anxiety. The STICSA consistently had better patterns of correlations in support of convergent and divergent/discriminant validity than did the STAI.

Research on the GBI

A large amount of psychopathology research relies on identification of affective conditions on a trait or lifetime basis, and given the low prevalence in nonclinical populations, inventory identification is typically used as a first-stage strategy. Because of this purpose, the inventory must be exceedingly specific to affective conditions and sensitive to severities in nonclinical populations. Resulting from a lack of such inventories, the General Behavior Inventory (GBI; Depue, Krauss, Spoont, & Arbsi, 1981; 1989) was initially developed in 1981 and revised in 1989. The original version was created to identify bipolar affective disorders, cyclothymic to bipolar I. The GBI was revised to include identification of unipolar affective disorders of varying intensities, from subsyndromal to full syndromal (Depue et al., 1989). Initial validation using several different populations indicated the original GBI had favorable psychometric properties, including specificity and sensitivity. Validation of the revised version showed 99% of patients with nonaffective disorders were correctly identified. The revised version was utilized and examined by Depue et al. (1989).

The GBI is a 73-item self-report inventory that assesses behaviors, and nonbehavioral dimensions, associated with depression and hypomania/mania. Nonbehavioral dimensions include: intensity during episodic behaviors (impairment), duration of behaviors at a clinical intensity (minimum of 3 days), rapid behavioral shifts, and frequency of described behavior over time. Respondents rate each item on a 4-point Likert-type scale, ranging from 1 (never or hardly ever) to 4 (very often or almost constantly). A two-dimensional scoring system method was developed, and employed, to include unipolar identification. Each respondent receives a total score for depression and a total score for hypomania/mania and biphasic. Also, the GBI has a low clinical floor to allow a greater range of severities to be identified (Depue et al., 1989).

Depue et al. (1989) examined the validity of the GBI as trait-based in a nonclinical population. A nonclinical population was utilized for several reasons: first, it includes individuals with a personal history of affective disorder but no manifestation of symptoms; second, it includes a full range of severities, from subsyndromal to full syndromal (Depue et al., 1989). If the GBI is able to accurately identify those with histories but no symptoms and those with mild subsyndromal conditions, it might be considered trait-based. The nonclinical sample was obtained from an original randomly acquired group of 1,068 White/Caucasian university students who had been administered the GBI. Subjects were then randomly selected from division lines at 50%, 80%, and 95% of the population of scores on both subscales. These divisions were chosen based on previous work that showed affective disordered cases typically do not score in the lower 80% on either subscale. Also, because the focus was with affective disordered conditions, research had shown those individuals fall in the upper 5% of each subscale score. Lastly, the ultimate goal was to determine cutoff scores and errors usually increase with individuals 5% and

above. Thus, based on these criteria, 205 subjects were chosen for blind interviews, 4 of whom declined, leaving a total of 201 subjects for this study (Depue et al., 1989).

The interviews focused on lifetime diagnostic status. A modified Schedule for Affective Disorders and Schizophrenia—L (SADS-L) was administered prior to the GBI. The interview materials were reviewed by four research diagnosticians and diagnoses were derived by consensus. Diagnostic criteria for mild disorders were more stringent because that allows for greater reliability (Depue et al., 1989).

Results showed that the GBI almost invariably identified conditions that had a chronic-intermittent course (Depue et al., 1989). Among the correctly identified unipolars, 15 had a history of depressive episodes and only one had a mild, episodic condition of low frequency. Among the bipolar diagnoses, the severity of depression was positively correlated with the depression subscale scores. Classification errors were calculated for a larger population using the sample. Three conditions were defined using the cutoff scores: below cutoff meant a nonaffective inventory diagnosis, above the cutoff meant either a unipolar or bipolar condition. The number of sample errors the GBI made and how many would be made in the larger sample were calculated based on the smaller sample of 201 and which of the above three categories they fell into. From these data, standard classificatory indices could be calculated; if this is done the GBI should be considered neutral with regards to short, low-frequency episodic depressions, thus those falling above or below the cutoffs are considered negative or positive (Depue et al., 1989). Results of predictive analyses showed the GBI had high positive predictive power (PPP) (.94 for unipolar and .87 bipolar) and high negative predictive power (NPP) across both affective conditions (.995 unipolar and .93 bipolar). Specificity was high for both conditions (.999

unipolar and .99 bipolar), and sensitivity was passable (.78 unipolar and .76 bipolar), with approximately one fourth of "true" unipolar and bipolar cases being missed (Depue et al., 1989).

There are four points by which this study demonstrated GBI diagnostic efficacy as a first-stage case-identification inventory in a nonclinical population. First, the high PPP and specificities for both affective conditions indicated the GBI was able to correctly identify affective groups with few false positives. Second, while the GBI's sensitivity was only passable for both conditions, this was because the misidentified cases were mild; therefore the GBI's sensitivity for more intense cases was considered good. Third, the scores above the cutoff had little overlap, and thus the two groups were regarded as homogeneous with respect to type of disorder. Fourth, and last, the GBI was able to correctly identify both subsyndromal and completely syndromal affective conditions.

Overall, the GBI performed well. The Depue et al. (1989) suggested three types of data are provided by the GBI: one, a moderately sensitive and highly specific index of nonaffective versus affective status; two, with the affective conditions, reasonably accurate estimates of unipolar-bipolar status were made; and three, by using a dichotic scoring method, a preliminary guideline for the predominance of depression through the course of a disorder was made. Based on rates of diagnoses, the GBI does not appear to overdiagnose affective conditions that are clinically relevant. The lifetime prevalence rates of bipolar conditions reported from the results of this study are comparable to other epidemiologic studies of university populations. Depue et al. (1989) specifically contributed the most to the lifetime prevalence of bipolar conditions by including primarily hypomanic cyclothymic disorder, a milder case of cyclothymia. The ability to identify this condition probably stems from the interest in bipolar traits at all levels of intensity (Depue et al., 1989).

The last note about the GBI is that there are at least three ways a psychiatric inventory having a trait basis can be utilized. First, as a diagnostic index in a huge nonclinical population when interviews are not possible; the GBI's combination of high PPP, NPP, and specificity lend it to giving limited but useful data. Second, the GBI could provide a dimension of frequency of affective disturbance across time in a nonclinical population with a likelihood of a continuum of affective disturbance, such as high-risk offspring. In this situation the GBI could be an index of instability. Third and last of uses for the GBI is: if the purpose is to screen large nonclinical populations to choose a mostly pure affective sample and nonaffective-control group for research purposes, the GBI would be useful. The GBI demonstrated quite low false-positive rates for unipolar and bipolar groups in the nonclinical sample of the current study (Depue et al., 1989). This is consistent with previous studies involving both clinical and nonclinical populations (Depue & Klein, 1988; Depue et al., 1981; Klein et al., 1986).

Over the years there has been an increase in recognition of affective disorders and the fact they often present in a mild, chronic form. Given this increase, there was a newfound need to be able to screen for these mild forms in a clinical setting (Klein, Dickstein, Taylor, & Harding, 1989). The modified General Behavior Inventory (GBI; Depue & Klein, 1988; Depue et al., 1981) was a good candidate because it was created to be able to distinguish between unipolar and bipolar forms of affective disorders, identify a whole range of severity of affective conditions, and to assess symptoms on a trait basis. The modified version only identified bipolar conditions (Depue et al., 1981), and overall it had been shown to have good psychometric properties. At the time of this study there was little evidence of validity in clinical populations, therefore, Klein et al. (1989) aimed to investigate the modified version in a large sample of outpatients with the

concordance between the GBI and blind structured diagnostic interviews being the primary focus.

Subjects were obtained from a community mental health center and a university training clinic. Before their intake the subjects, 266 from the mental health center and 226 from the clinic, completed the modified GBI. Despite some demographic differences in education level, socioeconomic status, and proportion of Whites/Caucasians, the distribution of scores on the GBI was similar across both samples. In order to explore the concordance between the GBI and clinical diagnoses, 167 patients, 77 from the mental health center and 90 from the clinic, were administered structured diagnostic interviews. These patients were disproportionately sampled from high scorers so positive predictive power could be accurately assessed, and to derive cutoffs for case identification. Interestingly, none of the total 492 patients scored both high on the GBI Hypomania-Biphasic subscale and low on the GBI Depression scale (Klein et al., 1989).

The modified GBI (Depue et al., 1987) contains 73 items covering a range of hypomanic and depressive symptoms and biphasic behavior as described in the DSM-III (DSM-III; American Psychiatric Association, 1980). The Depression subscale has 46 items and the Hypomania-Biphasic subscale has 28 items with one item cross-loading and counted on both scales. Each item is rated on a 4-point Likert scale, with only items rated 3 or 4 being counted in the total. Interviews were based on a modified version of the Schedule for Affective Disorders and Schizophrenia (SADS; Spitzer & Endicott, 1978). Interviews were done during the first few treatment sessions by a clinical psychology faculty member and advanced graduate students. Based on these interviews, 15 patients had chronic bipolar conditions, 56 had chronic unipolar conditions, 65 had episodic depressive conditions, and 31 had nonaffective conditions. One hundred and nine nonbipolar depressive diagnoses were followed-up with 6 months after the

interview; 84% provided data. The follow-up assessment consisted of several measures, including another interview, and those who conducted this assessment were blind to the patients' GBI scores.

The results of the diagnostic performance of the GBI were reported for positive predictive power (PPP), negative predictive power (NPP), and overall predictive power for bipolar conditions in both the clinic and mental health center samples. The PPP for bipolar disorders was 85% in the clinic sample and 77% in the mental health center sample. The NPP for bipolar conditions was 99% in the clinic sample and 98% in the mental health center sample. Overall predictive power was 99% in the clinic sample and 97% mental health center sample. The PPP for unipolar conditions was 89% in the clinic sample and 55% in the mental health center sample. The NPP for unipolar conditions was 94% in the clinic sample and 89% in the mental health center sample. Overall predictive power was 93% in the clinic sample and 81% mental health sample. After adjusting for population rates, the GBI correctly classified over 99% of patients with nonaffective disorders as noncases. The predictive validity of the GBI was assessed in nonbipolar depressives over a 6-7 month follow-up. Patients who scored above the GBI Depression cutoff had significantly poorer outcomes on several measures during the follow-up period. The GBI Hypomania-Biphasic subscale was also tested to see if it could predict the development in hypomanic episodes in nonbipolar depressives. A total of 13 patients reported episodes during the follow-up period, and they had significantly higher GBI Hypomania-Biphasic scores than those who did not report such episodes (Klein et al., 1989).

Klein et al. (1989) investigated the validity of the modified GBI as a screening measure in outpatient settings and found the positive predictive power was "adequate-to-good for bipolar conditions" in both samples and "quite good for chronic unipolar disorders" in the clinic sample

(p. 109). Though, the PPP for chronic unipolar conditions in the other sample was quite poor. The negative predictive power was good-to-excellent for both disorders in both samples. Despite its limits, this study provided strong support for the predictive validity of the GBI. However, all the findings of this study must be interpreted with caution because of the criterion measure used. While structured interviews are often used as a gold standard, they are far from infallible, especially with regard to mild, chronic affective disorders (Barrett, 1986). Because of this caution, any concordance may have been limited by errors from the criterion. Lastly, an inventory should be used to supplement, not supplant, careful clinical assessment because many parameters may not be easily assessable via an inventory. Results of Klein et al. (1989) supported the utility of the GBI as a first-stage screening inventory in clinical settings for chronic unipolar and bipolar conditions.

Barr, Markowitx, and Koesis (1992) used the GBI as a screening measure for dysthymic disorder in an outpatient psychiatric sample to validate the GBI as a screening tool for dysthymic disorder given its high negative and positive predictive power. The sample consisted of 43 current and newly admitted patients. In addition to being administered the GBI, blind diagnostic interviews were also conducted and were used as a diagnostic criterion. Previously established cut-off scores were modified for the purposes of this study and these effects were examined.

Overall prevalence of dysthymic disorder was 42%. When the GBI was compared to the interview it had 61% sensitivity and 88% specificity. The threshold for dysthymia was lowered from the previously established cut-off score of 22 to 15, and doing so only caught one additional patient and lowered specificity to 52% (Barr et al., 1992).

According to Barr et al. (1992) the results of this study showed that the GBI was a poor instrument for screening for dysthymic disorder given its low sensitivity. It was believed that

because it was originally developed for college students it was relatively ineffective with other populations. This is illustrated by the fact that it is long, requiring around 20min to complete and clinical populations may not have the reading ability or attention span needed. Additionally, comorbidity of disorders may have rendered the subjects incapable of completing the GBI in a meaningful way (Barr et al., 1992). While sensitivity and specificity are reported in Barr et al. (1992), positive predictive power (PPP) and negative predictive power (NPP) would have been more appropriate; particularly because Barr et al. (1992) references Klein et al. (1989) as supporting the former's results despite the latter reporting PPP and NPP. Barr et al. (1992) decided to calculate the sensitivity and specificity for the data from Klein et al. (1989) so it would be comparable instead of calculating the PPP and NPP to be consistent with Klein et al. (1989). Unfortunately, Barr et al. (1992) does not report the statistics thus PPP and NPP could not be calculated and reported here.

Youngstrom, Murray, Johnson, and Findling (2013) developed a shortened version of the long, 73-item GBI and then validated this new, shorter version. In addition to these goals that would enhance the usability of the GBI, Youngstrom et al. (2013) sought to make a couple improvements: (1) capture the depressive and manic dispositions as distinct components of bipolar disorder; and (2) recognize the continuity between child/adolescent symptoms and adult bipolar disorder (BD). In order to develop the shortened depression and mania scales, secondary analyses used a pool of nine samples (two clinical youth and seven nonclinical adult).

The GBI was established as identifying lifetime diagnoses of BD and syndromal and subsyndromal affective tendencies in clinical and nonclinical populations (Danielson, Youngstrom, Findling, & Calabrese, 2003; Depue et al., 1989). Technically it consists of three categories but two, biphasic and hypomania, are typically combined into one as they both predict

onset of manic episodes (Alloy, Urošević, Abramson, Jager-Hyman, Nusslock, Whitehouse, & Hogan, 2012).

Several steps were employed as a means of reducing the number of items on the GBI, including: exploratory factor analyses on the two data sets, selection of the top 10 items from both EFAs, and Cronbach's alpha as a criterion for high loading in both data sets which was augmented until the scale met the set criterion. Item response theory and correlations were used to compare the new scale with the full-length GBI in order to examine content coverage. For the mania scale, five of the top 10 items were matched from the two EFAs and did not meet the internal reliability criterion so the next ranking items were examined and an additional two were selected. Once these two were added to the first five, internal reliability was met, thus creating a 7-item mania scale. For the depression scale, seven of the top 10 items were matched and met the internal reliability criteria, thus creating a 7-item depression scale. These two new scales were combined and named the 7 Up 7 Down Inventory. The 7 Up 7 Down Inventory had moderate correlations in both the youth and adult samples (rs > .40).

Criterion validity was examined for the 7 Up 7 Down Inventory using ROC analyses. Two comparisons were conducted, mania cases vs. no diagnoses and mania cases vs. other diagnoses. The first comparison had an AUC of .82 for the mania scale and .78 for the depression scale, and the second comparison had an AUC of .59 for the mania scale and .67 for the depression scale with the depression scale performing identically to the full-length version (Youngstrom et al., 2013). Convergent validity was comparable to the full-length version for the 7 Up 7 Down Inventory, and discriminant validity was slightly better for the shorter version.

Youngstrom et al. (2013) aimed to shorten the GBI and make a few improvements. A version using 7 mania items and 7 depression items was created using methods to ensure

application across age groups and to optimize discriminant validity while highlighting the fact that mania and depression are contrasting rather than opposites. The new 7 Up 7 Down Inventory displayed high internal consistency, strong correlations with the original full-length GBI, and good criterion validity, including good discriminant validity. The samples had an age range of 11-86 years old, thus it was validated for that age range (Youngstrom et al., 2013).

Pendergast et al. (2015) evaluated the diagnostic and predictive validity of GBI scores in discriminating bipolar disorder (BD) from unipolar depression (UPD) and ADHD in adolescents and young adults. The purpose of this study arose from the diagnostic challenges that surround differentiating BD from disorders with much shared symptomatology. While the GBI has shown success in assessing BD symptoms and having robust psychometric properties, its utility in distinguishing BD from other disorders with similar symptoms had yet to be examined. The age group of adolescents and emerging/young adults was chosen because it is a clinically important age range in differentiating between persistent ADHD and possibly emerging BD, while discerning UPD.

It was hypothesized that individuals with BD would score significantly higher than nonclinical controls and individuals with either ADHD or UPD on the GBI's hypomanic/biphasic subscale. Also, it was hypothesized that the UPD and BD groups would score higher than the ADHD or nonclinical controls on the depression subscale. An objective of this study was to develop multilevel diagnostic likelihood ratios (DLRs) to facilitate individual decision making.

Pendergast et al. (2015) employed two samples from two different projects conducted at the same university with the same primary investigators. The first consisted on 359 participants aged 14-19 years, the second consisted of 614 participants aged 18-24 years; and they were

oversampled for individuals at risk of bipolar spectrum disorders. A number of measures were used, including: the Schedule for Affective Disorders and Schizophrenia—Lifetime Version, expanded (Exp-SADS-L; Endicott & Spitzer, 1978) and the General Behavior Inventory (GBI; Depue et al., 1981), which consists of two subscales, depression (46 items) and hypomanic/biphasic (28 items). Participants received monetary compensation for their participation (Pendergast et al., 2015).

The samples were analyzed separately and the findings were comparable. After determining there were no significant differences in GBI ratings between the two samples they were pooled to maximize diagnostic likelihood ratios (DLRs) precision and diagnostic efficiency estimates. Based on the Exp-SADS-L diagnoses, participants were grouped into four categories: bipolar spectrum disorders, UPD disorders, current or past ADHD, and nonclinical controls who did not meet criteria for BD, UPD, or ADHD. All categories are hierarchical and allow for comorbidity.

Because of the importance of identifying and discriminating bipolar spectrum disorders from ADHD and UPD, Pendergast et al. (2015) assessed the predictive and diagnostic validity of the GBI. Pendergast et al. (2015) used logistic regression analyses with the following comparisons: BD versus nonclinical controls, mood disorders versus clinical and nonclinical controls, BD versus clinical and nonclinical controls, BD versus any diagnosis, BD versus UPD, and BD versus ADHD. The results reflected the hypotheses that the GBI would provide statistically significant and clinically meaningful discrimination of BD from the other comparison groups. The exception was when an attempt was made to discriminate any mood disorder from all others. In particular, scores from the hypomanic/biphasic subscale significantly contributed to all comparisons. ROC analyses were conducted to examine the value of the two

GBI subscale scores for differentiating individuals within diagnostic groups and showed diagnostic efficiency ranging from fair to good. AUC values for both subscales showed they were significantly better than chance for all comparisons with the hypomanic/biphasic subscale slightly better at differentiating those with BD from those with UPD. In this case both subscales provided incremental information; therefore they should both be interpreted, and sequentially according to the two-step diagnosis process mentioned below. Multilevel DLRs were estimated by dividing the scores on each subscale into sextiles. When several categories emerged as redundant and were thus combined, the end result was the scores divided into three categories: low, moderate, and high. DLRs were also estimated for use in a two-step diagnosis process in which the depression subscale was utilized to determine presence or absence of a mood disorder and the hypomanic/biphasic subscale to determine absence or presence of a more specific range: bipolar spectrum disorders. The DLR values showed that GBI scores can provide helpful information in differentiating BDs from other groups, with very low or high scores changing the odds the most. Overall, the GBI appeared to outperform other tests in the use specified for this study. Additionally, it has several advantages for utility: it is free, it is in the public domain, and it does not require special training to administer or score. Several drawbacks however are that it is long (73-items) and it requires a high school reading level.

While the GBI has been hailed as having the most robust psychometric qualities of any self-report inventory it has one major shortcoming: all its validation has been conducted with predominantly White/Caucasian samples. Because of this shortcoming, Pendergast, Youngstrom, Brown, Jensen, Abramson, and Alloy (2015) examined the structural invariance of GBI scores in White/Caucasian and Black/African-American young adults. A specific issue among Black/African-American individuals and diagnosis is they are overdiagnosed with schizophrenia

or antisocial personality disorder and underdiagnosed with bipolar disorder (BD). However, studies have shown that when reviewing data and the race is unknown Black/African-American participants are less likely to receive the more severe diagnosis. It is presumed that clinicians who use clinical judgment as opposed to other, more structured methods are more likely to be influenced by racial stereotypes and cognitive biases. For Pendergast, Youngstrom et al. (2015), strong evidence for invariance would indicate the GBI could help reduce bias during the assessment process; on the other hand, performance discrepancies may help identify other sources of differential rates of diagnosis that are not heuristics or biases.

Participants in Pendergast, Youngstrom et al. (2015) included 291 Black/African-American and 994 White/Caucasian undergraduate students with a median age of 18 years. The majority were female. There were no significant gender or age differences on any of the scores, total or subscale. Pendergast, Youngstrom et al. (2015) parceled the items because it is considered useful for combining clinical and empirical standards in scale development and validation, and in particular for the GBI it helps assimilate items that ask about changes in energy or mood (biphasic items). Furthermore, the 20 parcels used in previous research (Danielson et al., 2003; Youngstrom et al., 2001; Youngstrom et al., 2013) were also used in this study. Each of the 20 parcels had three or four items with mostly homogeneous content. Differential item functioning (DIF) analyses using item response theory (IRT) were conducted to evaluate racial differences between the items in each parcel relative to the others. The results of these analyses showed minimal DIF, and when the differences were statistically significant, they canceled out at the scale score level.

Multigroup CFAs were utilized to examine structural invariance across race/ethnicity. Invariance was assessed by applying increasingly restrictive constraints across groups. Weak

invariance was met if the slopes of the regression lines between items and factors were equal across groups, strong invariance was met if the intercepts and slopes of regression lines between items and factors were equal across groups, and lastly, strict invariance was met if the slopes, intercepts, and residuals were equal across groups. Some cases were deleted if 30% or more of the GBI items had no response; this reduced the total sample to 285 Black/African-American and 987 White/Caucasian undergraduate students. In concordance with previous research for model fit, eight parcels loaded on the hypomanic/biphasic subscale, fourteen parcels loaded on the depression subscale, and two parcels cross-loaded. For both races the two-factor model fit within the *a priori* limits, however for the Black/African-American participants it was only an adequate fit. All factor loadings were statistically significant except Parcel 2 and the hypomanic/biphasic factor, which was nonsignificant, for both groups.

Because of the discrepancies between Black/African-American and White/Caucasian assessment of BD, rates of diagnosis of BD, access to services, and utilization of services it was crucial to determine measurement invariance as a possible means of reducing assessment bias. The findings of Pendergast, Youngstrom et al. (2015) showed the GBI had an invariant factor structure between Black/African-American and White/Caucasian groups in a nonclinical setting. While the GBI did not have complete measurement invariance it did have weak and strong invariance, thus it can be considered functionally invariant. Overall, the findings indicated that the structure of the dimensions and the relationship between most symptoms and the factors are similar across the two groups. It should be noted that the overall fit for Black/African-Americans was marginal; therefore further research using a mix of approaches is needed to investigate how Black/African-Americans think about behavior and mood problems. It is important for understanding racial and cultural differences given the existing evidence for differences in

attention to symptoms, expression of them, how they are reported, and which symptoms are reported.

A quick summary of the literature review before moving on: the TDI was developed using the Rasch measurement model to avoid the psychometric weaknesses associated with Classical test theory, the TDI has primarily been evaluated using clinical and nonclinical Italian adult samples, the TDI has shown divergent/discriminant validity with the STICSA, there is only one study where the TDI was used in the U.S. with a nonclinical young adult sample, the TDI has good internal consistency and reliability, and lastly, the TDI should be interpreted as unidimensional. The STICSA has better divergent/discriminant validity than the STAI, it has been evaluated using nonclinical and clinical samples, it has good internal consistency and reliability, and the STICSA has support for a four-factor model (state-cognitive, state-somatic, trait-cognitive, and trait-somatic) and a hierarchical model. The GBI was developed to identify unipolar affective disorders then was modified to screen for bipolar affective disorders, the GBI has high PPP and high NPP, it has been evaluated primarily with Caucasian samples, the GBI is considered to have the most robust psychometric properties of any self-report inventory, and the GBI has support for a two-factor model: hypomania-biphasic and depression.

The purpose of the present study was to replicate and extend previous studies by evaluating the convergent and discriminant validity of the TDI, GBI, and STICSA. The primary focus was on the convergent and discriminant validity of the TDI in comparison to the GBI and the STICSA. Again, the hypotheses were as follows: the TDI and GBI depression factor would have high correlations while the TDI and GBI hypomania-biphasic factor and both STICSA subscales (trait and state anxiety) would have lower correlations.

Method

Participants

Participants of this study were one hundred individuals ranging in age from 18-62 (M = 26.75, SD = 11.23). Included in the study were 78 females, 14 males, and 8 participants who identified as other. The mean age and primarily nonclinical sample of the current study is similar to the sample used in Ruan et al. (2016) study of nonclinical young adults in the United States. There was some diversity in race/ethnicity; 78% White/Caucasian, 7% Black/African American, 9% Hispanic/Latino, 4% Asian American, and 2% other. A total of 100 participants completed the TDI and the STICSA, while 95 completed the TDI, STICSA, *and* GBI.

Materials

The presentation of materials and administration of the three scales was conducted via Qualtrics, see Appendix for exact details. The first page was information about the study, an invitation to participate, and the informed consent form. This first page informed the participants about the study and that they had the right to discontinue at any time without penalty. The informed consent form also assured the participants that their responses were anonymous and confidential. They were then provided with the names of the researchers. The second page asked the participants to provide demographic information, which included age, race/ethnicity, gender, sexual orientation, education level ranging from less than high school to above a bachelor's degree or higher, religious affiliation, marital status, and whether they had been formally diagnosed with depression, anxiety, bipolar disorder, or another disorder, which they could specify.

The next page, if the participant did not provide consent, was the debriefing form, which is explained below. If participants consented, the next page following the demographic questions was the first of the three inventories: the TDI, the STICSA, or the GBI.

The last page was the debriefing form, which fully explained the purpose of the study and thanked participants for their participation. This page also had contact information for the principal investigator, faculty sponsor, and mental health resources. The mental health resources were the National Institute of Mental Health, the Anxiety and Depression Association of America, the National Alliance on Mental Health, the Depression and Bipolar Support Alliance, and the National Suicide Prevention Hotline. There was also a note at the bottom informing participants they could contact their local community mental health services or their college campus services if they required attention for mental health difficulties.

Instruments

Teate Depression Inventory (TDI). The TDI is a newer self-report depression inventory that purports to measure depression and identify a range of severities of depression, from mild to severe. The TDI was originally developed in Italy (Balsamo & Saggino, 2013) and was later translated into English for use in the United States (Ruan et al., 2016). The TDI has a total of 21 items, with ratings of 0 = "Never" to 4 = "Always." The TDI has 10 positively phrased items and 11 negatively phrased items, the latter 11 are reverse scored. Higher scores indicated more severe symptomology. Presently, the TDI has evidence of favorable psychometric properties and while it has support for a three-factor bifactor structure with Depressed Mood, Life Satisfaction, and Daily Function subscales, it should be primarily interpreted as unidimensional via the total score (Balsamo et al., 2014; Ruan et al., 2016).

State-Trait Inventory of Cognitive and Somatic Anxiety (STICSA). The STICSA (Ree et al., 2008) is a newer self-report anxiety inventory that measures state and trait anxiety and within those types identifies somatic and cognitive symptomology. The STICSA State and Trait subscales each contain 21 items. The Trait subscale ratings are 1 = "Almost Never" to 4 = "Almost Always." The State subscale ratings are 1 = "Not at all" to 4 = "Very much so." Higher scores indicate more symptomology. Presently, the STICSA has shown favorable psychometric properties and should be interpreted on a state and trait basis (Ree et al., 2008).

General Behavior Inventory (GBI). The GBI is a well-established self-report inventory that identifies a wide range of symptoms of unipolar and bipolar affective conditions. The GBI has 73 items with ratings of 1 = "Never or hardly ever" to 4 = "Very often or almost constantly." Higher scores indicate more severe symptomology. The GBI has two subscales, Depression (46 items) and Hypomania-Biphasic (28 items). The GBI could be interpreted on a subscale basis or as unidimensional, whichever would provide the most relevant information (Depue et al., 1989). The GBI has a long history of proven psychometric properties, including favorable positive predictive power and negative predictive power (Depue et al., 1989; Klein et al., 1989; Barr et al., 1992).

Procedure

Participants were obtained through convenience sampling using posts on social media and the sending of emails which contained a link to the study. All data were collected over a period of eight weeks via Qualtrics, an online survey and data collection software; thus participants were able to complete the study at their convenience. Participation was completely voluntary with no offer of compensation or inventive other than helping to further research on

the TDI, STICSA, and GBI. All procedures were approved by the university's IRB; therefore, the current study met all applicable ethical standards.

The order of these inventories was randomly counterbalanced to control for order effects. The use of the TDI was approved by the publishers (Hogrefe) for research purposes, and the STICSA and the GBI are in the public domain.

Results

Table 1 presents descriptive statistics for selected demographic variables. The sample was primarily female (78%), heterosexual (61%), and White/Caucasian (78%). Regarding a formal diagnosis of mental illness, 10% reported depression, 15% reported anxiety, and 24% reported both.

Table 1
Demographic and Sample Characteristics

Variable	*n*	%
Gender		
Male	14	14
Female	78	78
Nonbinary	5	5
Other	3	3
Sexual Orientation		
Homosexual	4	4
Heterosexual	61	61
Bisexual	23	23
Pansexual	8	08
Queer	4	04
Race/Ethnicity		
White/Caucasian	78	78
Black/African American	7	7
Hispanic/Latino	9	9
Asian American	4	4
Formal Diagnosis		
Depression	10	10
Anxiety	15	15
Depression & Anxiety	24	24
PTSD	4	4
Eating Disorder	3	3
Other/Multiple Diagnoses	10	10
None	34	34

Items that needed to be reverse-scored were first recoded. The TDI Total and Depressed Mood (DM), Life Satisfaction (LS), and Daily Function (DF) subscale scores were calculated. The Cognitive and Somatic subscale scores for the STICSA State and Trait scales were calculated. The Depression and Hypomania-Biphasic subscale scores for the GBI were also calculated. All participants (N=100) completed the TDI and STICSA while 95 participants completed the GBI. The missing data for the GBI is likely a result of the length of the GBI (73

items). Table 2 presents descriptive statistics for the TDI, STICSA, and GBI. Skewness and kurtosis were within normal range for each inventory and their respective subscales.

Table 2
Psychometric Properties of the Major Study Variables

Variable	*M*	*SD*	Range		Skewness	Kurtosis
			Potential	Actual		
Teate Depression Inventory						
Total	42.68	16.57	0-84	6-79	-.15	-.42
Depressed Mood	23.53	9.74	0-44	3-43	-.18	-.72
Life Satisfaction	12.85	5.87	0-28	0-25	-.11	-.68
Daily Function	6.30	2.22	0-12	1-12	-.20	-.12
State-Trait Inventory of Cognitive and Somatic Anxiety						
Trait-Cognitive	25.36	8.08	10-40	10-40	-.32	-.86
Trait-Somatic	19.44	6.34	11-41	11-41	.87	.54
State-Cognitive	22.52	8.55	10-40	10-40	.26	-1.06
State-Somatic	17.17	6.90	11-41	11-37	1.23	1.06
General Behavior Inventory						
Hypomania/Biphasic	22.73	15.91	0-84	0-62	.74	-.37
Depression	57.37	31.36	0-138	2-112	-.04	-1.15

Note. Teate Depression Inventory (TDI) and State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) samples $N = 100$, General Behavior Inventory (GBI) sample $n = 95$ as 5 participants failed to complete the GBI.

Pearson product-moment correlation coefficients were calculated for TDI Total and subscale scores with all other scores on the STICSA and the GBI using SPSS v. 20. The *SimpleStats Test Program* (Watkins, 2007) was used for dependent *t*-tests to compare the convergent and discriminant coefficients for statistically significant differences. All correlations (Table 3) were statistically significant, $p < .001$ (two-tailed).

TDI-GBI comparisons in Table 3 illustrate convergent validity coefficients for the TDI Total, DM, LS, and DF scores with the GBI Depression scale with correlations ranging from .71 to .83. Thus, the TDI shared variance with the GBI Depression subscale ranged from 50.4% to 68.9%. Discriminant validity coefficients for TDI Total, DM, LS, and DF scores with the GBI Hypomania/Biphasic scale with correlations ranging from .41 to .54. Thus, the TDI shared variance with the GBI Hypomania-Biphasic subscale ranged from 16.8% to 29.2%. Convergent validity coefficients were significantly higher than discriminant validity coefficients, $p < .001$. The lower correlations and lower percentages of shared variance between the TDI Total and subscales and GBI Hypomania-Biphasic subscale is evidence of discriminant validity.

TDI-STICSA comparisons in Table 3 illustrate convergent validity coefficients for the TDI Total, DM, LS, and DF scores with the STICSA State-Cognitive and Trait-Cognitive scales with correlations ranging from .56 to .81. The TDI shared variance with the STICSA State- and Trait-Cognitive scales ranged from 31.4% to 65.0%. Discriminant validity coefficients for the TDI Total, DM, LS, and DF scores with the STICSA State-Somatic and Trait-Somatic scales included correlations ranging from .39 to .63. The TDI shared variance with the STICSA State- and Trait-Somatic scales ranged from 15.2% to 39.7%. Convergent validity coefficients were significantly higher than discriminant validity coefficients, $p < .0018$.

Dependent *t*-tests were conducted for differences between correlation coefficients using the following comparisons: TDI Total-GBI Depression vs. TDI Total-GBI Hypomania-Biphasic, $t(92) = 6.89$, $p < .0001$; TDI Total-GBI Depression vs. TDI Total-STICSA State-Cognitive, $t(92) = 2.14$, $p = .0349$; TDI Total-GBI Depression vs. TDI Total-STICSA State-Somatic, $t(92) = 5.97$, $p < .0001$; TDI Total-GBI Depression vs. TDI Total-STICSA Trait-Cognitive, $t(92) = .62$, $p = .5367$ (n. s.); TDI Total-GBI Depression vs. TDI Total-STICSA Trait-Somatic, $t(92) = 4.45$,

$p < .0001$; TDI Total-STISCA State-Cognitive vs. TDI Total-STICSA State-Somatic, $t(97) = 4.13, p < .0001$; TDI Total-STICSA Trait Cognitive vs. TDI Total-STICSA Trait-Somatic, $t(97) = 3.61, p < .0005$; TDI Total-STICSA State-Cognitive vs. TDI Total-STICSA Trait-Cognitive, $t(97) = -2.17, p = .0328$; and TDI Total-STICSA State-Somatic vs. TDI Total-STICSA Trait-Somatic, $t(97) = 3.21, p < .002$. All comparisons were statistically significant except one, which was noted.

Table 3

Pearson Product Moment Correlations and Descriptive Statistics for the Teate Depression Inventory, State-Trait Inventory of Cognitive and Somatic Anxiety, and General Behavior Inventory (N = 100)

			TDI				STICSA				GBI	
	M	*SD*	Total	DM	LS	DF	S-C	S-S	T-C	T-S	D	H/B
Teate Depression Inventory (TDI)												
Total	42.68	16.57	–									
Depressed Mood (DM)	23.53	9.74	.96	–								
Life Satisfaction (LS)	12.85	5.87	.92	.79	–							
Daily Function (DF)	6.30	2.22	.81	.70	.75	–						
State-Trait Inventory of Cognitive and Somatic Anxiety (STICSA)												
State-Cognitive (S-C)	22.52	8.55	.74[C]	.76[C]	.63[C]	.56[C]	–					
State-Somatic (S-S)	17.17	6.09	.47[D]	.46[D]	.39[D]	.44[D]	.56	–				
Trait-Cognitive (T-C)	25.36	8.08	.81[C]	.80[C]	.71[C]	.65[C]	.85	.46	–			
Trait-Somatic (T-S)	19.44	6.34	.63[D]	.62[D]	.53[D]	.56[D]	.62	.80	.65	–		
General Behavior Inventory (GBI)												
Depression	57.37	31.36	.83[C]	.83[C]	.71[C]	.64[C]	.71	.51	.82	.71	–	
Hypomania/Biphasic	22.73	15.91	.53[D]	.54[D]	.41[D]	.50[D]	.56	.63	.65	.68	.72	–

Note. Teate Depression Inventory (TDI) and State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) samples $n = 100$, General Behavior Inventory (GBI) sample $n = 95$ as 5 participants failed to complete the GBI. All correlations statistically significant, $p < .001$.

[C] Convergent validity coefficient

[D] Discriminant validity coefficient.

Discussion

The purpose of the current study was to replicate and extend previous research by examining convergent and discriminant evidence of construct validity for the TDI, a relatively new measure of depression. It was hypothesized that the TDI and the GBI Depression subscale would reflect convergent validity and the TDI and the GBI Hypomania-Biphasic and the STICSA State and Trait Somatic subscales would reflect discriminant validity.

The results of the current study showed a statistically significant correlation between the TDI Total scores and the GBI Depression subscale, which was consistent with previous research showing convergent validity between the TDI and other measures of depression (Balsamo & Saggino, 2014). The results also showed statistically significant but lower correlations between the TDI Total score and the GBI Hypomania-Biphasic subscale and the STICSA State and Trait Somatic subscales, which is also consistent with previous research on discriminant validity of the TDI with other measures of anxiety (Balsamo & Saggino, 2014; Ree et al., 2008; Roberts et al., 2015). It was also hypothesized that there would be somewhat lower correlations between the TDI Total scores and the STICSA State and Trait Cognitive subscales relative to the high correlations between depression factors; however, these correlations were not as low as expected. The correlations between the TDI Total score and the STICSA State and Trait Cognitive subscales were still consistent with other research (Balsamo & Saggino, 2014; Balsamo et al., 2013; Grös et al., 2007).

The relationships between the TDI Total score and the STICSA State and Trait Cognitive subscales are likely due to the overlap that exists with cognitive symptoms of depression and anxiety, such as negative thoughts and difficulty concentrating. In contrast, there is less overlap between symptoms of depression and somatic symptoms of anxiety; depression is typically

associated with decreased arousal while anxiety is typically associated with increased arousal. This contrast is reflected by the relationship between the TDI Total scores and the STICSA State and Trait Somatic subscales (Balsamo et al., 2013; Grös et al., 2007).

Some limitations of the current study included its small sample size, no offer of an incentive for participation, which likely contributed to the small sample size, and 5 cases with missing GBI data. As stated before, the missing GBI data was probably due to the length of the GBI as it includes 73 items and the item content is quite long, so they take considerably more time to read. The small sample size contributed to the lack of generalizability of the results. The sample was only one hundred participants, and there was not much diversity in the sample, therefore it would be inappropriate to broadly generalize these results to other populations.

Future research should investigate the TDI, STICSA, and GBI with regard to invariance of measurement across a number of important demographics such as sex/gender, race/ethnicity, and age to confirm they measure similarly across these categories. Future research comparing race/ethnicity is necessary because of the lack of research with ethnically diverse samples. Most, if not all, of the research on these three scales has been conducted primarily with White/Caucasian samples, with the exception of Pendergast, Youngstrom et al. (2015), whose study investigated GBI measurement differences with Black/African-Americans compared to White/Caucasians. Additional research should examine how the TDI, STICSA, and GBI measure individuals of varying ages as most of the existing research has been conducted with young adult samples. Finally, given the increasing visibility of the LGBTQA (Lesbian, Gay, Bisexual, Transsexual, Queer, and Asexual) community, and the known mental health problems associated with non-heterosexual and/or non-binary individuals, it is important for there to be measures that are valid and reliable in assessing individuals from the LGBTQA community.

To conclude, the TDI is a strong measure of depression in comparison to the GBI Depression subscale. The TDI showed discriminant validity when compared to the STICSA, a scale that measures anxiety, and the GBI Hypomania-Biphasic subscale. Results showed a convergent relationship between the TDI and STICSA Cognitive factors, which was discussed. Lastly, further research should focus on the invariance of measurement of the TDI, STICSA, and GBI across a variety of demographics.

References

Alloy, L. B., Urošević, S., Abramson, L. Y., Jager-Hyman, S., Nusslock, R.,...Hogan, M. (2012) Progression along the bipolar spectrum: a longitudinal study of predictors of conversion from bipolar spectrum conditions to bipolar I and II disorders. *Journal of Abnormal Psychology*, 121(1), 16-27. doi: 10.1037/a0023973

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR®*. Arlington, VA: American Psychiatric Publishing; 2000.

Angst, J. (1995) The epidemiology of depressive disorders. *Europ Neuropsychopharmacol*, 5 Suppl: 95–98.

Balsamo, M. & Saggino, A. (2013). Teate Depression Inventory. Firenze, IT: Hogrefe.

Balsamo, M. & Saggino, A. (2014) Determining a diagnostic cut-off on the teate depression inventory. *Neuropsychiatric Disease and Treatment*, 2014(10), 987-995. doi: 10.2147/ndt.s55706

Balsamo, M., Carlucci, L., Sergi, M. R., Murdock, K. K., & Saggino, A. (2015) The mediating role of early maladaptive schemas in the relation between co-rumination and depression in young adults. *PLoS ONE*, 10(10), 1-14. doi: 10.1371/journal.pone.0140177

Balsamo, M., Giampaglia, G. & Saggino, A. (2014) Development of a new measure of depression using Rasch analysis. *Neuropsychiatric Disease and Treatment*, 2014(10), 153-165. doi: 10.2147/ndt.s53425

Balsamo, M., Imperatori, C., Sergi, M. R., Belvederi Murri, M., Continisio, M., Tamburello, A.,...Saggino, A. (2013) Cognitive vulnerabilities and depression in young adults: an ROC curves analysis. *Depression Research and Treatment* 2013, 1-8. doi: 10.1155/2013/407602

Balsamo, M., Imperatori, C., Sergi, M. R., et al. Cognitive vulnerabilities and depression in young adults: an ROC curves analysis. *Depress Res Treat*. 2013;2013:407602.

Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M.,...Saggino, A. (2015) Measurement of external shame: an inside view. *Journal of Personality Assessment*, 97(1), 81-89. doi: 10.1080/00223891.2014.947650

Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., & Saggino, A. (2013) The state-trait anxiety inventory: shadows and lights on its construct validity. *J Psychopathol Behav Assess* 2013(35), 475-486. doi: 10.1007/s10862-013-9354-5

Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., Saggino, A. The State-Trait Anxiety Inventory: shadows and lights on its construct validity. *J Psychopathol Behav Assess*., 2013(35): 475–486.

Baron, R. M., and Kenny, D. A. (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of - Personality and Social Psychology*, 51(6), 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Barr, L. C., Markowitz, J. C., & Koesis, J. H. (1992) The General Behavior Inventory as an Outpatient Screen for Dysthymic Disorder. *Journal of Psychopathology and Behavioral Assessment*, 14(4), 337-342. doi: 0882-2689/92/1200-0337$06.50/0

Barrett, J. (1986). Case identification for category validation: The challenge of disorder-specific assessment. *Comprehensive Psychiatry*, 27, 81-100. doi: 10.1016/0010-440X(86)90018-0

Beck, A. T. and Steer, R. A. (1988) *Beck Hopelessness Scale manual*. San Antonio, TX: The Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: The Psychological Corporation.

Beck, A.T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961) An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.

Berto, P., Ilario, D. D., Ruffo, P., Di Virgilio, R., & Rizzo, F. (2000) Depression: cost of illness studies in the international literature, a review. *The Journal of Mental Health Policy and Economics*, 3, 3-10. doi: 10.1002/1099-176X(200003)3:13.0.CO;2-H

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., ...Nelson, B. (1996) Childhood and adolescent depression: a review of the past 10 years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(11), 1427-1439 doi: 10.1097/00004583-199611000-00011

Cattell, R. B. (1966). Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 288–329). Chicago: Rand McNally.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, 100, 316–336. doi: 10.1037/0021843X.100.3.316

Danielson, C. K., Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003) Discriminative validity of the general behavior inventory using youth report. *Journal of Abnormal Child Psychology*, 31(1), 29-39. doi: 0091-0627/03/0200-0029/0

Depue, R. A. & Klein, D. N. (1988) Identification of unipolar and bipolar affective conditions in nonclinical and clinical populations by the General Behavior Inventory. *Relatives at Risk for Mental Disorder*, 179-202.

Depue, R. A., Krauss, S., & Spoont, M. R. (1987). A two-dimensional threshold model of seasonal bipolar affective disorder. In D. Magnusson & A. Ohman (Eds.), *Psychopathology: An interactional perspective* (pp. 95-123). New York: Academic Press.

Depue, R. A., Krauss, S., Spoont, M. E., & Arbsi, P. (1989) General Behavior Inventory Identification of Unipolar and Bipolar Affective Conditions in a Nonclinical University Population. *Journal of Abnormal Psychology*, 98(2), 117-126. doi: 0021-843X/89/S00.75

Depue, R. A., Slater, J. F., Wolfstetter-Kausch, H., Klein, D., Goplerud, E., and Farr, D. (1981) A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: a conceptual framework and five validation studies. *Journal of Abnormal Psychology*, 90(5), 381-437. doi: 10.1037/0021-843X.90.5.381

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, 35, 837– 844. doi: 10.1001/archpsyc.1978.01770310043002

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric Properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory. *Psychological Assessment*, 19(4), 369-381. doi: 10.1037/1040-3590.19.4.369

Grös, D. F., Simms, L. J., & Antony, M. M. (2010). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads. *Behavior Therapy*, 41, 277–284. doi: 10.1016/j.beth.2009.07.001

IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Klein, D. N., Depue, R. A., & Slater, J. F. (1986) Inventory identification of cyclothymia IX: validation in offspring of bipolar I patients. *Arch Gen Psychiatry*, 43(5), 441-445. doi: 10.1001/archpsyc.1986.01800050043005

Klein, D. N., Dickstein, S., Taylor, E. B., & Harding, K. (1989). Identifying Chronic Affective Disorders in Outpatients: Validation of the General Behavior Inventory. *Journal of Consulting and Clinical Psychology*, 57(1), 106-111. doi: 0022-006X/89/J00.75

Lehrer, P. M., & Woolfolk, R. L. (1982). Self-report assessment of anxiety: Somatic, cognitive, and behavioral modalities. *Behavioral Assessment*, 4, 167–177.

Lopez Molina, M. A., Jansen, K., Drews, C., Pinheiro, R. S., & Souza, L. (2014) Major depressive disorder symptoms in male and female young adults. *Psychology, Health & Medicine*, 19(2), 136-145. doi: 10.1080/13548506.2013.793369

Lovibond, P. F., & Lovibond, S. H. (1995). Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia.

Meng, X.-l., Rosenthal, R., and Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175. doi: 10.1037/0033-2909.111.1.172

Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for the BIDR version 6*. Unpublished manuscript, Department of Psychology, University of British Columbia, Vancouver, Canada.

Pendergast, L. L., Youngstrom, E. A., Brown, C., Jensen, D.,...Alloy, L. B. (2015) Structural Invariance of the General Behavior Inventory (GBI) Scores in Black and White Young Adults. *Psychological Assessment*, 27(1), 21-30. doi: 10.1037/pas0000020

Pendergast, L. L., Youngstrom, E. A., Merkitch, K. G., Moore, K. A., Black, C. L.,...Alloy, L. B. (2015) Differentiating Bipolar Disorder From Unipolar Depression and ADHD: The

Utility of the General Behavior Inventory. *Psychological Assessment*, 26(1), 195-206. doi: 10.1037/a0035138

Qualtrics Research Suite (Version 2.1) [Computer software]. (2009). Provo, UT: *Qualtrics Labs, Inc.* Retrieved from http://www.qualtrics .com

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401. doi: 10.1177/014662167700100306

Ree, M. J., French, D., MacLeod, C., & Locke, V. (2008). Distinguishing Cognitive and Somatic Dimensions of State and Trait Anxiety: Development and Validation of the Sate-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behavioral and Cognitive Psychology*, 2008(36), 313-332. doi: 10.1017/S1352465808004232

Ree, M. J., MacLeod, C., French, D., & Locke, V. (2000, November). *The State–Trait Inventory for Cognitive and Somatic Anxiety: Development and validation.* Poster session presented at the annual meeting of the Association for the Advancement of Behavior Therapy, New Orleans, LA.

Roberts, K. E., Hart, T. A., & Eastwood, J. D. (2015). Factor Structure and Validity of the State-Trait Inventory for Cognitive and Somatic Anxiety. Psychological Assessment, 28(2), 134-146. doi: 10.1037/pas0000155

Robins, L., Regier, D. (Eds.). (1991). *Psychiatric Disorders in America: the Epidemiologic Catchment Area Study* (pp. 328-366). Free Press: New York.

Rose, A. J. (2002). Co–rumination in the friendships of girls and boys. *ChildDev* (73) 1830–1843. PMID: 12487497

Ruan, L., Pendergast, L., Dixon, T., Liao, P. C., Jones, P., & von der Embse, N. (2016) Measuring depression in college students: structural validity of the teate depression inventory (TDI). Temple University, 1-31.

Schwartz, G. E., Davidson, R. J., & Goleman, D. J. (1978). Patterning of cognitive and somatic processes in the self-regulation of anxiety: Effects of meditation versus exercise. *Psychosomatic Medicine*, 40, 321–328. doi: 10.1097/00006842-197806000-00004

Sobel, M. E. (1982) Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13 (1982), 290-312. doi: 10.2307/270723

Spielberger, C. D. (1966). Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3–19). New York: Academic Press.

Spielberger, C. D. (1972). Conceptual and methodological issues in anxiety research. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (pp. 481–492). New York: Academic Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y).* Palo Alto, CA: Consulting Psychologists.

Spitzer, R. L., & Endicott, J. (1978). *Schedule for the Affective Disorders and Schizophrenia* (3rd ed.). New York: New York State Psychiatric Institute, Biometrics Research.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293. doi:10.1126/science.3287615

Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R.,...Cardenas, S. J. (2007). Robust dimensions of anxiety sensitivity: Development and initial validation

of the Anxiety Sensitivity Index-3. *Psychological Assessment*, 19, 176–188. doi: 10.1037/1040-3590.19.2.176

Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *Handbook of research methods in psychology, Vol. 1: Foundations, planning, measures, and psychometrics* (pp. 723-744). Washington, DC: American Psychological Association.

Watkins, M. W. (2007). SimpleStat Tests [Computer software]. Phoenix, AZ: Ed & Psych Associates.

Watson, D., & Clark, L. A. (1991). *The Mood and Anxiety Symptom Questionnaire*. Unpublished manuscript, Department of Psychology, University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Young, J. E. (2003). Young Schema Questionnaire—Long Form 3 (YSQ-L3). New York, NY: Schema Therapy Institute.

Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. R. (2001) Discriminative validity of parent report of hypomanic and depressive symptoms on the general behavior inventory. *Psychological Assessment*, 13(2), 267-276. doi: 10.1037//1040-3590,13.2.267

Youngstrom, E. A., Murray, G., Johnson, S. L., & Findling, R. L. (2013) The 7 Up 7 Down Inventory: A 14-Item Measure of Manic and Depressive Tendencies carved From the General Behavior Inventory. *Psychological Assessment*, 25(4), 1377-1383. doi: 10.1037/a0033975

Appendix

Consent Form

Thank you for your interest in this study! I am an undergraduate psychology departmental honors student conducting research to compare how three different questionnaires measure symptoms or characteristics of mood and worry. In the mental health field, it is essential that valid and reliable tools are used to provide the best services to those in need. Your participation and responses to these questions based on your experiences are helpful.

My first task is to gather more information about how several newer questionnaires work in measuring individual's reports of fear, worry, and various moods.

Participation in the study is anonymous and will be extremely beneficial to building a better understanding of how well these newer questionnaires work. All information will be confidential, but some of the items or questions could make some individuals feel uncomfortable. In the event that you feel concerned about mental health, contact information for national mental health organizations is provided at the end of the survey. Although there are no direct benefits to the participants, individuals may gain insight about mental health through completing the survey and help contribute valuable information to the mental health field.

Completing the surveys usually takes between 20 and 45 minutes. Participation in the study is voluntary; you are free to discontinue at any time without penalty.

If you have any questions, please contact the primary investigator, Eleanor Crouse at ecrouse@eiu.edu, or the faculty sponsor, Dr. Gary Canivez at glcanivez@eiu.edu.

If you have any questions or concerns about the treatment of human participants in this study, you may call or write:

Institutional Review Board, Eastern Illinois University
600 Lincoln Ave., Charleston, IL 61920
Telephone: (217) 581-8576
E-mail: eiuirb@www.eiu.edu

You will be given the opportunity to discuss any questions about your rights as a research subject with a member of the IRB. The IRB is an independent committee composed of members

of the University community, as well as lay members of the community not connected with EIU. The IRB has reviewed and approved this study.

Debriefing Form

If you are looking for more information regarding mental health, please contact a national organization.

National Institute of Mental Health

Website: https://www.nimh.nih.gov/index.shtml

Health and Information: https://www.nimh.nih.gov/health/index.shtml

Telephone: 1-866-615-6464 (toll-free)

Monday through Friday

8:30 a.m. to 5:00 p.m. ET

Email: nimhinfo@nih.gov

Anxiety and Depression Association of America

Understanding Anxiety: https://www.adaa.org/understanding-anxiety

Finding Help: https://www.adaa.org/finding-help

Contact Information: https://www.adaa.org/contact-adaa

Telephone: 240-485-1001

Email: information@adaa.org

National Alliance on Mental Health

Website: http://www.nami.org/

Finding Support: http://www.nami.org/Find-Support

Helpline: 800-950-6264

Depression and Bipolar Support Alliance

Website: http://www.dbsalliance.org/site/PageServer?pagename=home

Education: http://www.dbsalliance.org/site/PageServer?pagename=education_landing

Toll-free Phone: (800) 826-3632

National Suicide Prevention Line

1-800-273-8255

24 hours, 7 days a week

If you are looking for mental health services, please contact mental health counselors in your community or college campus.