CROSS-MEASURE EQUIVALENCE AND COMMUNICABILITY IN THE ASSESSMENT OF DEPRESSION: A FINE-GRAINED FOCUS ON FACTOR-BASED SCALES

David Andrés González, B.A.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

August 2012

APPROVED:

Sharon Rae Jenkins, Major Professor
Adriel Boals, Committee Member
Camilo Ruggero, Committee Member
Vicki L. Campbell, Chair of the Department of Psychology
Mark Wardell, Dean of the Toulouse Graduate School

González, David Andrés, Cross-measure equivalence and communicability in the assessment of depression: A fine-grained focus on factor-based scales. Master of Science (Psychology), August 2012, 61 pp., 11 tables, 59 references, 69 titles.

Depression is heterogeneous, however, depression measures conceptualize it as homogeneous. To help fulfill NIMH's strategic plan to focus on components of depression, this study analyzed the psychometrics of factor-based subscales in the BDI-II, CES-D, IDAS, and IDS. CCA was also used to explore redundancy across measures. Using a diverse sample of symptomatic undergraduates, this study found the IDAS to be the best measure, with complete DSM-IV symptom coverage and psychometrically sound subscales. The other measures did not have consistent subscales or coverage of symptoms. Furthermore, CCA revealed low levels of redundancy across measures. These results serve to disabuse the field of a perception that different measures of equivalently measure depression. Conversion tables were provided to empirically compare scores from different measures.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of my advisor, Sharon Rae Jenkins, for going above and beyond the requirements of her position. I would also like to acknowledge the emotional and moral support of my mother, Elouisa Díaz, and father, José de Jesús González Sánchez.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS …. iii

LIST OF TABLES …. vii

Chapter

1. INTRODUCTION …. 1

Depression Heterogeneity and Subtypes

Confronting Homogeneity in Measurement

Importance of Factors to Distinguish Content Domains

Absence of Cross-Measure Equivalence

DSM-IV as Standard for Measures

Evidence of Measure Confusion

Goals of Present Study

Hypotheses

Measure overlap

Descriptive and Psychometric Goals

2. METHOD …. 17

Sample

Measures

Demographics.

Beck Depression Inventory – Second Edition

Center for Epidemiologic Studies – Depression Scale

Inventory for Depression and Anxiety Symptoms

Inventory of Depressive Symptomatology

Patient Health Questionnaire – 9

Procedure

Statistical Analyses

Descriptive and Psychometric Goals

Psychometrics of Subscales

Item-Level Descriptive Comparison of Measures

Conversion Tables

IDS Validation

3. RESULTS …. 29

Descriptives

Subscale-level Psychometric Goals

Internal Consistency Reliabilities

Correlations/Hypothesis Tests

BDI-II Scales

CES-D Scales

IDAS Scales

IDS Scales

Confirmatory Factor Analyses

Canonical Correlation Analyses

Communicative Goals

4. DISCUSSION 37

Factor Psychometrics

Summary and Recommendations

Canonical Correlation Analysis

Conversion Tables

IDS Validation

Limitations

Future Research and Recommendations

REFERENCES …..... 55

LIST OF TABLES

Table 1: Descriptives of Summary Scales …. 45

Table 2: Descriptives for Subscales 45

Table 3: Correlations Among Summary Scales 46

Table 4: Correlations Among Subscales 47

Table 5: Fit of Extant Factor Structures in Current Sample …. 48

Table 6: Canonical Variate Correlations and Redundancy Analysis for BDI-II with Other Scales 49

Table 7: Canonical Variate Correlations and Redundancy Analysis for CES-D with Other Scales 50

Table 8: Canonical Variate Correlations and Redundancy Analysis for IDAS with Other Scales 51

Table 9: Canonical Variate Correlations and Redundancy Analysis for IDS with Other Scales 52

Table 10: Results of Analyses for Predicting Scales Based on Other-Scale Scores …. 53

Table 11: Regression Equations for Predicting Scales Based on Other-Scale Scores …. 54

CHAPTER 1

INTRODUCTION

Major depressive disorder (MDD) is a psychiatric disorder that carries major personal, familial, and economic consequences. According to the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association [APA], 2000), a major depressive disorder is characterized by the presence of a major depressive episode, absence of manic symptoms, and is not better accounted for by any other affective disorders, medical conditions, or substance use. Major depressive episodes are often characterized by depressed mood and anhedonia, and various combinations of seven other symptoms that can lead to distress (APA, 2000). Major depressive disorder (hereafter referred to as depression) can cause undue stress on a family, sometimes resulting in fragmentation of familial bonds or the development of the disorder in other family members, particularly children (Lewinsohn & Essau, 2002). In addition, depression has been found to have a marked negative impact on the economy, with estimates in production losses ranging around $33 billion annually (Greenberg, Kessler, Nells, Finkelstein, & Berndt, 1996).

Although the phenomena of depression has been documented for centuries (Burton, 1621/2001), the recognition of it as a disorder and delineation of specific criteria for it to be diagnosed has not occurred until the 20th century. The criteria needed to diagnose depression and other mental disorders have changed over the years, from Feighner's criteria, to the first *DSM*, to Research Diagnostic Criteria (RDC), and International Classification of Diseases (ICD) criteria (Rogers, 2001). The current standard in the United States for diagnosing mental disorders is the *DSM-IV*.

There is a variety of methods to diagnose depression, including unstructured interviews, structured clinical interviews, and self-report measures. Different methods can vary based on the

diagnostic criteria used, theoretical orientation of the designer (e.g., if they believe that abnormal cognitions are the cause of depression), or time period during which it was developed.

Current diagnostic criteria and measures are not without their critics. Common critiques of the *DSM* depression nosology are that it is too broad (i.e., two people with the same diagnosis may have markedly different presentations), does not take into account the dimensional nature of symptoms (i.e., someone may have significant distress, but only have four symptoms and not be diagnosed with depression), and does not provide meaningful subtypes (Klein, 2008; Watson, 2005). A common critique of measures of depression, specifically self-report measures, is that they may not tap into criteria for depression equally, and do not normally provide factor scores (Watson et al., 2007).

This thesis is focused on the measurement of depression. Specifically, it focuses on understanding differences between different measures of depression, increasing communicability between them, and seeing if the use of factor-based scales is feasible. This paper introduces literature on diagnosis and subtyping of depression, measurement of depression, factors within measures of depression, content validity issues in depression, evidence of how these validity issues have affected clinical trials, and then clearly outline the study goals.

Depression Heterogeneity and Subtypes

Currently, the *DSM-IV* has nine symptoms identified for a major depressive episode (APA, 2000):

1. Depressed mood most of the day, or irritable mood with children and adolescents
2. Decreased interest or pleasure in activities
3. A change in more than 5% of current body weight
4. Insomnia or hypersomnia most of the day
5. Psychomotor agitation or retardation

6. Fatigue or loss of energy
7. Feeling worthless or guilty
8. Diminished ability to think or concentrate
9. Recurrent thoughts of death, suicidal ideation with varying degrees of seriousness

In order for an individual to be diagnosed with depressive episode, they need to have had at least five of the previous symptoms over 2 weeks. As such, individuals can be presenting with different symptoms yet still be diagnosed with the same disorder; that is to say, depression has a heterogeneous presentation. To compensate for this heterogeneity, several researchers have proposed subtypes focused on common symptom patterns or etiology.

Although the *DSM-IV* currently has specifiers for catatonic, melancholic, and atypical subtypes, these are given more attention in psychiatric literature than in psychological literature, were initially derived from clinical experience, and may require future modification (Stewart, McGrath, Quitkin, & Klein, 2007). Two methods have generally been employed to discern subtypes. One method is to measure a range of depressive symptoms and see which ones group together empirically. If certain symptoms tend to cluster together across different samples then this may indicate that there is some common factor causing these symptom groupings, which may lead to focused treatment (Carragher, Adamson, Bunting, & McCann, 2009). Another method espoused by Blatt (2004) is to look at the common experiences that preceded the expression of MDD in individuals and create subtypes based on common preceding events.

Using the first method, researchers have identified a number of subtypes. One study by Carragher et al. (2009) identified four salient subtypes from a nationally representative subsample of 12,180 respondents with depressive symptoms from the NESARC study: severely depressed, psychosomatic, cognitive-emotional, and "non-depressed" subtypes. These symptom groupings were derived from the symptom set listed in *DSM-IV* criteria. Forty percent of

respondents were best categorized by the severely depressed subtype, which endorsed the majority of symptoms. Thirty percent of respondents mainly identified eating and sleeping disturbances along with psychomotor and impaired concentration complaints; these were identified as the psychosomatic subtype. Ten percent of participants mainly endorsed feelings of worthlessness and guilt, concentration difficulties, and suicidal ideation; these were identified by the cognitive-emotional subtype. The last group of "non-depressed" participants endorsed some symptoms, but did not have consistently high ratings on any of the *DSM-IV* A criteria.

The three "actually depressed" subtypes had increased odds-ratios of having a family background with depression, previous life difficulties, a concurrent diagnosis of depression, comorbid anxiety disorder, or personality disorder. The odds-ratios were of different magnitudes for the different subtypes. The odds-ratio for a comorbid Axis II disorder and alcohol dependence was highest with the cognitive-emotional group. The severely depressed group had high odds-ratios for anxiety as well as having a diagnosis of depression. The psychosomatic group was more likely to include those divorced, separated, or widowed and less likely to include Hispanics and those aged 30 - 44. Regarding diagnoses, 92% of those in the severely depressed group had MDD diagnoses, whereas 48% in the psychosomatic, 36% in the cognitive-emotional, and none in the non-depressed group had MDD diagnoses.

In addition to the subtypes identified by Carragher et al. (2009), other subtypes have been identified. Blatt (2004) reviewed several models of depression and, based on an individual's history of experiences and issues, concluded that there are two subtypes of depression: anaclitic and introjective depression. Anaclitic depression consists of a history of interpersonal problems relating to attachment and pleasing others. Introjective depression is characterized by many problems maintaining a positive self-concept, such as locus of control and teleological expectations. These types are not mutually exclusive, and individuals may have a history of both

experiences (Blatt, 2004).

Vanheule, Desmet, Verhaeghe, and Boggaerts (2007) identified an alexithymic subtype of depression. Alexithymia was characterized by somatic symptoms, more suicidal ideation, and less responsiveness to antidepressant medication (Vanheule et al., 2007). The subtype seems to overlap with Carragher's psychosomatic group, although the Carragher study did not identify correlates with psychopharmaceutical treatment. The inclusion of increased suicidal ideation is a bit perplexing and requires further research to determine if in fact these two subtypes are actually distinct, but they seem to point to the same background.

Despite the different approaches to subtyping, it appears that the different methods lead to similar core subtypes with distinct correlates. Although research has demonstrated the existence of clear subtypes, assessment of depression has often focused on a general score and neglected to include subtypes. This has led some researchers to suggest we change the manner in which depression is assessed.

Confronting homogeneity in measurement.

In their strategic plan for mood disorders research, the National Institute of Mental Health (NIMH) identified the assessment of depression as one of the top priorities for research (NIMH, 2003, p. 93):

> Advances in depression research and treatment development are highly dependent on the quality of research procedures to measure, assess, or classify the pathology and its expressed symptomatology. No reliable biological markers or valid behavioral tests exist to define the exact nature of depression and disentangle issues of comorbid pathologies, or co-occurring syndromes or clusters of symptoms; according, diagnostic and classification systems have principally relied upon clinical description and the naming of behavioral signs and symptoms to define the syndrome (e.g., sad mood, sleep difficulties,

diminished interest). The traditional assessment and diagnosis of depression has proved insensitive for the identification of likely responders to existing psychosocial and psychopharmacological treatments...the most widely used instruments in clinical settings have generally failed to provide clear documentation of the symptoms experienced by individuals and instead typically have offered only global indices of depression.

In response to the preponderance of global indices, NIMH urged that our current conceptualization be broken down to subtypes that can be reliably linked to different psychological, social, and genetic factors and outcomes. In breaking down the general construct, and creating reliable and valid measures for these subconstructs, or content domains, one will subsequently improve the effectiveness of clinical trials and interventions by ensuring we are adequately measuring a heterogeneous concept.

Importance of factors to distinguish content domains.

One possible solution to the problem of heterogeneity clearly focused on by the NIMH is to move beyond global indices of depression to specific subscales that may indicate subtypes of depression and the appropriate treatment or probable course. A generally accepted method of creating subscales for measures of a general construct is factor analysis. For depression, factor scales have shown promise for a number of reasons:

- Different factors have different relationship patterns with different variables. For example, one study that divided the Zung Self-Rating Depression Scale into four factors found that somatic and well-being factors predicted survival in individuals with coronary heart disease at a statistically significant level (Barefoot, 2000). The relationship between survival and the somatic factor may be more indicative of general physical complaints and less of a mood disorder.
- Different ethnic or gender groups often have different mean scores for factors. One study

by Endler, Rutherford, and Denisoff (1999) found that Canadian men and women had different mean factor scores, with women scoring higher on cognitive and somatic factors of the BDI. In another study (Yen, Robins, & Lin, 2000), the authors were able to differentiate Chinese inpatients and outpatients, as well as Chinese and American respondents using factor patterns on a Chinese translation of the CES-D.

- Treatment may affect certain factors of depression, but not others. In a meta-analytic review, Faries et al. (2000) reported that subscales derived from the Hamilton Depression Rating Scale (HAMD) consistently outperformed the total score as a measure of change in studies trying to detect differences in anti-depressant medication. The authors reported that studies that used only the global score would have to increase their sample size by one-third in order to detect change that factors could detect.
- Factor scores may respond differently than summary scores as outcome measures for change. Mackinger and Svaldi (2004) found that only the cognitive factor of the BDI moderated the performance of autobiographical memory in sleep apnea patients.

Factors are also important because they allow us to determine how well scores from each measure can communicate with each other. Currently, popular measures of depression offer a global index score which should correlate with other measures' global scores. While there are a number of studies demonstrating high correlations between measures, few studies have explicitly compared factors across measures, or examined how well specific symptom profiles or factor scores on one scale may translate to the score on another scale. This is of particular importance for clinicians and researchers who may want to compare their results with previous research studies. For example, a clinician may want to know if the decrease in depressive symptoms as measured by the Zung after the second week of a treatment is equivalent to the decrease that treatment research studies have reported using the IDS-C.

In summary, factors that distinguish content domains have conceptual and empirical importance for our understanding of depression. However, as NIMH's report suggests, researchers and clinicians consistently marginalize these aspects of depression. Partly because of this marginalization, psychologists are unclear about what factors of depression a measure taps into, and whether different measures of depression equally tap into the same factors.

Absence of cross-measure equivalence.

Although there is a multitude of depression measures, research has demonstrated that these measures do not measure depression in the same fashion. For example, one meta-analysis by Shafer (2006) that compared factors in different measures found less overlap than expected between measures. The author synthesized multiple factor analytic studies on four common measures of depression: the BDI, CES-D, HRSD, and Zung. The author found seven reliable factors across the four measures. Only two factors were shared between all four measures: negative affect, which contained items such as a punitive attitude towards self or sense of sadness, and somatic symptoms, that had items such as weight loss or general somatic complaints. One factor was shared with the CES-D and Zung: positive affect, which contained items related to enjoying activities and having a positive outlook for the future. The four remaining factors were unique to each test: the CES-D had a 2-item interpersonal problems factor with items such as "people don't like me"; the BDI had a performance impairment factor that focused on anhedonia or negative behavioral symptoms; and the HRSD had anxiety and insomnia factors.

In summary, different measures of depression do not seem to tap into the same content domains of depression consistently. When measures do tap into the same domains, they do not give it the same weight. For example, one measure may have two items in the somatic domain, whereas another measure will have seven items for the somatic domain. Although this may not

be surprising given that the measures were developed with different theoretical orientations, what is surprising is that measures of depression are often treated as equivalent. This disregard of content validity can lead to serious problems. In the words of Haynes, Richard, and Kubany (1995):

> Data from an invalid instrument can overrepresent, omit, or underrepresent some facets of the construct and reflect variables outside the construct domain. A content-invalid assessment instrument could erroneously indicate the occurrence or nonoccurrence of clinically significant treatment effects.

Given this varied coverage, it is important that psychologists agree on standard criteria that measures *have* to reflect in item content.

DSM-IV as Standard for Measures

Clearly, measures of depression need more overlap and a standard set of symptoms to draw from. The current standard for measures should be the *DSM-IV* list of depressive symptoms. While measures of depressive symptomatology are not meant to be diagnostic tools, they certainly should be informed by current nosological criteria, especially because they are often used in lieu of structured interviews to indicate possible depression.

One content validity study by Guillion and Rush (1998) measured how well four measures of depression (the IDS-C, IDS-SR, HRSD, and BDI) mapped on to *DSM-IV* criteria for depression. In this study of content validity, the authors reported that the BDI covered eight of the nine criteria (except agitation/retardation; Guillion & Rush, 1998). In addition, about half of the items assessed two criteria: self-blame and hopelessness/suicide. The HRSD also covered eight *DSM* symptoms; in this case missing the concentration symptom. Both measures assess for irritability and somatic complaints, despite the fact that the *DSM* does not list these as symptoms. The IDS-SR tapped into every symptom, including symptoms for *DSM* melancholic and atypical

subtypes and RDC endogenous subtype (Guillion & Rush, 1998). In sum, these measures do not seem to tap into *DSM-IV* criteria equally, with the most content valid being the IDS-SR. It is also important to note that the proposed changes for the *DSM-5* do not affect these nine core symptoms of depression (APA, 2010).

Despite the *DSM-IV-TR*'s status as diagnostic standard, it is not without critics. Most criticism of the *DSM-IV* is aimed at the current inclusion symptom nosology, which marginalizes exclusion and outcome criteria. Succinctly, critics claim that the current diagnostic groupings are based on spurious distinctions as evidenced by high base rates of comorbidity and categories so heterogenous that they are not clinically useful with regard to etiology or treatment indicators (Klein, 2008; Watson, 2005). Critics propose that the *DSM-5* should shift diagnostic criteria from symptoms to etiology or empirical groupings (Andrews, Anderson, Slade, & Sunderland, 2008; Watson, 2005). Based on patterns of comorbidity, Watson (2005) has called for mood and anxiety disorders to be grouped under the diagnostic family of Emotional Disorders. Specifically, major depressive disorder, dysthymic disorder, generalized anxiety disorder and posttraumatic stress disorder would be grouped together as distress disorders.

Alternatively, researchers (Klein, 2008; Watson, 2005) have called for the use of a dimensional classification system. A dimensional system differs from categorical systems in that symptoms are ranked based on severity along a scale as opposed to presence or absence. The benefits of categorizing severity are that symptom severity is often associated with the course and comorbidity of disorders as well as with diagnostic stability over time as opposed to the present/absent changes in diagnosis that may appear over time with dichotomous nosologies (Watson, 2005). To accommodate the fact that discrete categories ease communication, Klein (2008) has proposed that dimensional systems incorporate several thresholds along the continuum to ease communication between professionals. In addition to categorizing symptom

severity along a continuum, Klein (2008) has also proposed that a dimensional scale of chronicity be included in diagnostic criteria for unipolar mood disorders. Such a scale may be a useful criterion because unipolar mood disorders may be chronic or recurrent regardless of severity. For example, an individual may have dysthymia or a single episode of minor depressive disorder, or alternatively, chronic MDD or a single episode of the disorder. Klein (2008) suggests that this two-dimensional model be incorporated within Watson's (2005) hierarchical distress disorder model.

In summary, *DSM* criteria are a useful standard to use for content domains when comparing measures. Content validity studies have found that not all measures of depression equally reflect these core aspects of it. However, critics of the *DSM* focus on how it is inadequate to measure depression independently of anxiety, and focus on depression as a categorical construct. Self-report measures of depression are a perfect compromise inasmuch as they may tap into *DSM* symptoms, as well as anxiety symptoms (Guillion & Rush, 1998), and do so in a dimensional manner.

Evidence of Measure Confusion

Although the literature previously reviewed suggests that giving multiple measures of depression to the same individuals may lead to divergent results, there have been few studies documenting the actual occurrence of such findings. However, a few clinical trials have incidentally documented differential results with different measures. The assumption that measures assess depression equally has led researchers to disregard content validity issues when selecting measures to document change due to an intervention. When assessment instruments are not matched to an intervention, researchers run the risk of not accurately reflecting reality.

One study by Muñoz, Ying, Bernal, and Pérez-Stable (1995) tested the effect of a preventative intervention on a certain sample of people. In their study, they used status as a low-

income minority as an inclusion criterion because it was a good predictor of future depression. They used a current diagnosis of depression as determined by a structured interview as an exclusion criterion because it was solely a preventative intervention. After 6 and 12 month follow ups, the researchers found that there was a statistically significant reduction in depressive symptoms when measured with the BDI, but *not* when measured with the CES-D (Muñoz et al., 1995). How many studies using only one assessment instrument have yielded non-significant results because they did not match that instrument to their intervention?

In a similar situation, Seligman, Schulman, DeRubeis, and Hollon (1999) also tested the effectiveness of a preventative intervention. In their study, they used a sample of college freshman that scored past a cutoff of 9 on the BDI, but did not meet criteria for depression using a structured interview. In their study they found that the intervention had effects when using the BDI, but not when using the HDRS. In both studies, the intervention focused on cognitive interventions. This may account for the significant findings using the BDI, which heavily weights cognitive symptoms. Had the studies used factors, they may have found significant differences on cognitive factors, reducing the risk not documenting specific change.

It has been posited that these differences are due to the fact that the different measures may focus on different symptoms of depression, and may not be the general measures they purport to be (Gillham, Shate, & Freres, 2000). As a result, the authors have recommended the use of multiple measures to see if effects are found with at least one and recommended further research on differences between measures of depression. I would like to extend the previous suggestion, saying that researchers should choose their measures carefully, to make sure the measures both measure adequately what the intervention is attempting to address (i.e., focus on content validity), and cover all diagnostic aspects of depression. Had the researchers focused on factors, or carefully chosen measures, they may have found that the intervention was particularly

helpful for certain aspects of the disorder.

Goals of Present Study

In light of the aforementioned, it becomes clear that the conceptualization and measurement of depression should move toward a focus on factors that measure content domains of depression. The purpose of this study is to increase our understanding of different measures of depression.

In line with NIMH objectives, this study aims to determine the feasibility of using subscale scores on widely used measures of depression. Although some factors across measures overlap, not all do, nor has their relationships been described before in published research. This study will create subscales based on previously reported factor structures for each measure and and explore how they relate to each other.

Given that different measures of depression have different factors, another aim of this study is to increase the understanding of the differences and similarities between measures of depression, as well as facilitating communicability among them. Although correlations between measure summary scores and factor analyses have been used to describe relationships between and structure within measures respectively, this study will use canonical correlation analyses (CCA) to provide a more complete, item-level picture of the relationships between two groups of variables.

Hypotheses.

Based on these goals, several hypotheses are generated:

Measure overlap.

1. The BDI-II Cognitive-Affective subscale will have a stronger, positive correlation with the CES-D's Depressed Affect subscale than with the CES-D's other three subscales.
2. The BDI-II Cognitive-Affective subscale will have a stronger, negative relationship with the

CES-D's Positive Affect subscale than with the CES-D's other three subscales.

3. The BDI-II Cognitive-Affective subscale will have a stronger, positive correlation with the IDS' Cognitive/Mood subscale than with the IDS' other two subscales.

4. The BDI-II Cognitive-Affective subscale will have stronger, positive correlations with the IDAS' Dysphoria, Ill temper, Suicidality scales than with the IDAS' Insomnia, Lassitude, and Appetite change scales.

5. The BDI-II Somatic-Vegetative subscale will have a stronger, positive correlation with the CES-D's Somatic subscale than with the CES-D's other three subscales.

6. The BDI-II Somatic-Vegetative subscale will have stronger, positive correlations with the IDS' ArousalAnxiety and Sleep subscales than with the IDS' Cognitive/Mood subscale.

7. The BDI-II Somatic-Vegetative subscale will have stronger, positive correlations with the IDAS' Insomnia, Lassitude and Appetite change subscales than with the IDAS' other subscales.

8. The CES-D Somatic subscale will have stronger, positive correlations with the IDS' Arousal/Anxiety and Sleep subscales than with the IDS' Cognitive/Mood subscale.

9. The CES-D Somatic subscale will have stronger, positive correlations with the IDAS' Insomnia, Lassitude, and Appetite change subscales than with the IDAS' other subscales.

10. The CES-D Depressed Affect subscale will have a stronger, positive correlation with the IDS' Cognitive/Mood subscale than with the IDS' other two subscales.

11. The CES-D Depressed Affect subscale will have stronger, positive correlations with the IDAS' Dysphoria, Ill temper, and Suicidality scales than with the IDAS' Insomnia, Lassitude, and Appetite change scales.

12. The IDS Cognitive/Mood subscale will have stronger, positive correlations with the IDAS' Dysphoria and Suicidality scales than with the IDAS' other scales.

13. The IDS Arousal/Anxiety subscale will have stronger, positive correlations with the IDAS'

Lassitude and Ill temper scales than with the IDAS' other scales.

14. The IDS Sleep subscale will have stronger, positive correlations with the IDAS' Insomnia and Lassitude scales than with the IDAS' other scales.

Descriptive and psychometric goals.

While it is important to break down the current heterogeneous concept of depression, no study to date has determined which current widely used measures allow us to do this adequately. This study calculated different psychometrics, such as internal consistency, convergent validity, and discriminant validity, for the factors of the BDI-II, CES-D, IDS, and IDAS to see if existing identified factors are psychometrically sound enough to use individually as subscales. If so, then current measures of depression can be used in future studies that focus on subtypes of depression or symptom profiles. Although no psychometrics for individual factors have been published before, most measures of depression have at least adequate reliability and validity statistics for the scale as a whole. Desired reliability statistics are considered to be above .70 for internal consistency and brief test-retest intervals (DeVellis, 2003; Joiner, Walker, Pettit, Perez, & Cukrowicz, 2005). There are no standards for what is "adequate" in convergent, discriminant, and predictive validity (DeVellis, 2003). Generally, stronger coefficients in the expected direction are desired.

Another descriptive goal of this study was to use an item-level analysis (CCA) to describe the relationship between pairs of measures. Previously published studies have either described correlations between the general summary score for each measure, or looked at factor analyses within each measure. Based on these separate techniques, the conclusions have either been: these measures are highly related or, these measures are meaningfully distinct. CCA provides a more complete picture, inasmuch as measures that are more related to each other will have a higher redundancy statistic, or more shared variance.

In the spirit of acknowledging differences between measures and increasing communicability between them, a second psychometric goal will be to create conversion tables for summary scores. Conversion tables will allow clinicians and researchers to accurately compare their results using one measure with previous literature that used another measure. The measures used in this study will be the BDI-II, CES-D, IDS, and IDAS. For example, what score on the IDS corresponds to a score of a 24 on the BDI-II?

A third and final psychometric aim of this study is to validate measures in a new population. The IDS has no published studies validating it with symptomatic college students. Given the rise in suicidality and MDD in student populations, it is important to know if these measures maintain their psychometric characteristics when used with this population.

CHAPTER 2

METHOD

Sample

The sample consisted of 248 undergraduates who had a score above 5 on the Patient Health Questionnaire – Depression scale (PHQ-9; cut-off for mild depression). Of those 248, 30 individuals did not complete all measures, or were careless responders. Careless responding was determined by a series of three validity questions (e.g., have you answered a question about sleep problems in the past few pages) or endorsing only one response type (e.g., all 3s) for entire measures. This screening left a final sample of 218 symptomatic undergraduates.

The sample had an average age of 20.2 (*SD* = 2.7) and was mostly women (74.8%). The sample was also ethnically diverse: 58.1% identified as White, 17.9% identified as Hispanic, 12.6% identified as Black, 11% identified as Asian, and .4% identified as other. With regard to marital status, 74.8% identified as single, 19.9% were in a committed relationship, and 4.5% were married.

Nineteen percent of the sample (*n* = 41) reported receiving a formal diagnosis of a mental disorder. The vast majority of these diagnoses were of mood disorders (*n* = 31). Of the sample, 10.9% (*n* = 24) had received psychotherapy. Furthermore, 12.2% (*n* = 27) reported having a history of mental disorders in their family, with the majority of these being mood disorders (*n* = 22).

Measures

Demographics.

Participants were asked about their age, gender, ethnicity, years of education completed, marital status, major history, location of birth, current educational status, living situation, parent history (e.g., education, income, ethnicity, and employment for SES), current stress, social

network, history of physical and psychological illnesses, and treatment history.

Beck Depression Inventory - Second Edition.

The Beck Depression Inventory (BDI) was originally developed in 1961. It was based on Beck's cognitive theory of depression which posits the main etiology of major depressive disorder to be overly negative construals. In 1979, Beck and colleagues made minor revisions to the measure to improve its psychometric properties. Although the original BDI had good reliability characteristics (Beck et al., 1988) the construct validity was called into question (Nezu, Nezu, Friedman, & Lee, 2008). The next revision of the BDI was in 1996, which removed several items and reworded others to fall in line with *DSM-IV* criteria for depression. The BDI-II consists of 21 questions that an individual can rank on a scale of 0 - 3. Scores on the BDI-II can range from 0 to 63, and are organized by the following cut-offs: 0 -13, minimally depressed; 14 - 19, mildly depressed; 20 - 28, moderately depressed; and 29 - 63, severely depressed (Rush et al., 2008).

The BDI-II has demonstrated very high internal consistency in multiple populations. The measure's manual (Beck, Steer, & Brown, 1996) reported a Cronbach's alpha of .93 for a sample of 120 college students, and a Cronbach's alpha of .92 in a group of outpatient adults. In a sample of undergraduate students, Dozois et al. (1998) found a coefficient alpha of .91. In a sample of nonpatient high school students, Osman et al. (2008) reported a Cronbach's alpha of .92 (CI = .91 - .93). With a sample of inpatient adolescents, they found a similar alpha of .90 (CI = .88 - .92).

The BDI-II has also demonstrated adequate convergent and discriminant validity. Dozois et al. (1998) reported a correlation of .93 between the BDI-II and the original BDI. However, the use of a questionable measure to determine convergent validity may not have been the best choice. Osman et al. (1997) reported a correlation of .63 with the Beck Hopelessness Scale, .57 with the Suicidal Behaviors Questionnaire-Revised, and a correlation of .53 with the State-Trait

Anxiety Inventory-State. They also reported a correlation of -.55 with the Brief Reasons for Living Inventory for Adolescents. While the study indicated appropriate correlations with convergent concepts, it failed to discriminate from the anxiety inventory, which may raise pause. The study by Osman et al. (1997) reported that scores on the BDI-II were able to differentiate between the nonpatient and inpatient adolescents (estimate = .08, odds ratio = 1.08, CI = 1.10 - 1.06).

The BDI-II manual reported a two-factor structure in both their samples through the use of exploratory factor analyses. In Dozois et al.'s study (1998), exploratory factor analyses yielded a two-factor structure that accounted for 46% of the variance. The first factor was a 9-item Cognitive-Affective factor and the second was a 13-item Somatic-Vegetative factor. The authors did not indicate what type of rotation was used. They then proceeded to conduct a confirmatory factor analysis with oblique rotation. The authors reported good fit for their model, which had much better fit than the college-sample factor structure posited by Beck, Steer, and Brown (1996). Unfortunately, the authors used the same data that they used for the exploratory analysis, which could very well account for the high model fit. The authors then performed the same confirmatory oblique rotation on a subsample of their data that had not been use in the exploratory analysis. In this case, they reported adequate fit for their model, and did not compare it with Beck's original model.

In another study, Osman et al. (2008) reported the results of confirmatory factor analyses on four two-factor models using oblique rotation and one general/specific model using orthogonal rotation. In a sample of nonpatient adolescents, they reported a better fit for the general-sub-two factor model.

Center for Epidemiologic Studies - Depression Scale.

The Center for Epidemiologic Studies Depression Scale (CES-D) was published by

Radloff in 1977 and based upon previous measures of depression such as the BDI, MMPI Depression Scale, and Zung Self-Rating Depression Scale. The intent of the scale was to measure depressive symptoms in non clinical populations for epidemiological studies. It consists of 20 items that individuals answer on a Likert-type scale from 0 to 3. Scores can range from 0 to 60, with higher scores indicating higher or more distressful depressive symptoms. Shean & Baldwin (2008) found that cutoff scores of 16 had sensitivity and specificity rates of 86.7 and 76.6 for identifying depressed individuals, whereas a cutoff score of 21 had a sensitivity and specificity rate of 73 and 96.1. The CES-D was validated with African Americans, Caucasians, males, and females (Radloff, 1977). Of note is that the author used subsamples of the main sample that was used to calculate the original psychometrics. By using portions of the same sample, the statistics indicating equality across groups may be inflated. Despite this shortcoming, several studies have used the CES-D in different populations and validated the scale's use in those populations.

In a study with Mexican immigrants, Hiott et al. (2006) found a Cronbach's alpha of .84 for the CES-D. In a sample of freshman and sophomore college students, Shean & Baldwin (2008) found intrascale reliability of .89 for the CES-D. In the initial study by Radloff (1977), the author found that several representative samples of individuals from Washington County, Maryland and Kansas City, Missouri had coefficient alphas of .85, .85, and .84. In a distinct sample of 70 inpatients, Radloff reported a coefficient alpha of .90 (Radloff, 1977).

Radloff also reported a test-retest correlation of .51 and .67 for two and four weeks respectively. Of note is that these correlations were not based off of correlations on a whole sample, but were based off of smaller samples of 139 and 105 individuals who volunteered to send in second copies of their scores. As such, there may be a response bias affecting the scores.

In the same study, Shean and Baldwin (2008) found a Pearson correlation of .86 between

the BDI-II and CES-D. They also evaluated correlations with the Diagnostic Interview Schedule (DIS) results in a subsample of 95 students. They found Spearman rank coefficients of .56 for currently depressed individuals using the DIS and CES-D. Of note is that there was a higher Spearman rank coefficient for individuals who had experienced depression once in their lives according to the DIS (r = .62). This may in part be due to the small number of individuals who were currently depressed (N = 17). In the same study, the CES-D was a stronger predictor of DIS depression than the BDI-II. Radloff reported the CES-D had a correlation of .44 with the HRSD and .54 with the Raskin Rating Scale when assessed at the beginning of treatment. After four weeks of treatment, the correlations increased to .69 and .75 for the respective tests. Radloff also found a low correlation between the Marlowe-Crowne social desireability scale and CES-D of -.18.

Radloff reported four factors for the scale: depressed affect, positive affect, somatic symptoms, and interpersonal problems. The author used principal components analysis with varimax rotation to obtain the four factors which accounted for 48% of the variance. A meta-analysis yielded the same factors with little difference (Shafer, 2006).

Inventory for Depression and Anxiety Symptoms.

The Inventory for Depression and Anxiety Symptoms (IDAS) was developed as a response to the fact that assessment tools often treat depression as a homogeneous construct (Watson et al., 2007). Instead of revamping older, commonly used questionnaires, the authors decided to create a new measure that was multidimensional by design. The IDAS consists of 64 items that individuals rank on a 5-point Likert-type scale. The 64 items are part of two general scales, and 10 specific subscales. The two general scales are: a 20-item General Depression scale, which is similar to more traditional depression scales, and a 10-item Dysphoria scale, which focuses on the cognitive and emotional symptoms of depression. The specific subscales

are: Well-being, Panic, Lassitude, Insomnia, Suicidality, Social Anxiety, Ill Temper, Traumatic Intrusions, Appetite Loss, and Appetite Gain. All the subscales consist of eight or fewer items. There are currently no cut-off scores indicating severity categories.

The original article consisted of three different studies that developed and validated the IDAS with several different populations, including college students and young adults, psychiatric patients, the general community, high school students, and postpartum women (Watson et al., 2007). Across all of the samples, internal consistency for the different scales ranged from .75 - .92. Younger populations, such as the high school students, had a wider range (.77 - .92) than adult psychiatric patients (.84 - .91). The one week test-retest coefficient for the general depression scale was .84 with a patient sample. The same coefficients for the specific symptom scales ranged from .72 to .83.

With regard to validity, the different specific scales had weak to moderate correlations with each other, with the highest correlation being .56 (Watson, et al., 2007). In the same article, all subscales were statistically significantly related to the BDI-II in their combined sample, with the general depression and dysphoria scales showing correlations of .83 and .69 respectively. Discriminant validity was evidenced by most of the depression subscales having significantly stronger correlations with the BDI-II than the BAI. Two of the three anxiety scales (traumatic intrusions, panic) had significantly stronger correlations with the BAI.

Inventory of Depressive Symptomatology.

The Inventory of Depressive Symptomatology (IDS) was developed by Rush et al. (1986, 1996) to measure the extent of depressive symptoms according to *DSM-IV* criteria. This measure differs from most measures of depression in that it did not have a theoretical leaning upon development, and was developed to assess relevant *DSM* symptomatology. There are two versions of the form, the IDS Clinician-rated form (IDS-C) and IDS Self-report form (IDS-SR).

The forms only differ on the items' point of view and intended rater (clinician or patient). Form items were matched to ease comparisons between the two forms. Both forms consist of 30 items that rank the frequency and severity of individual's depressive symptoms. Scores on the IDS range from 0 - 84, with the following cut off score suggested: normal 0 -11; 12 - 23 mildly depressed; 24 - 36 moderately depressed; 37 - 46 moderate to severe depression; and 47 - 84 severely depressed Rush et al. (2008).

Different studies have assessed the internal consistency of the IDS in different settings. For the IDS-SR, in a sample of individuals diagnosed with depression using a structured interview (SCID-III-TR) for a study on cognitive treatment for depression, researchers reported an alpha internal-consistency level of .90 (median; range .72 - .91) for the IDS-SR (Vittengl, Clark, Kraft, & Jarrett, 2005). In another study by Rush et al. (1996), they reported a Cronbach's alpha of .94 for the IDS-SR in a sample of individuals that had been evaluated for a Mood Disorders Program at UT Southwestern. Of note is that the internal-consistency coefficient decreased for a subset of 338 individuals who were currently expressing symptoms for MDD (IDS-SR = .77). In Vittengl et al.'s (2005) study, they found that the IDS-SR had a test-retest zero-order *r* of .14 at the beginning and end of a 20-week cognitive treatment. Such a low correlation is to be expected when measuring a state variable over the course of a long treatment. The same study reported a test-retest *r* of .86 for the IDS-SR with a one-week interval at the end of the treatment.

Vittengl and colleagues (2005) reported convergent validity for the IDS-SR in a nonstandard fashion; the measures were compared against three other measures, and the median zero-order *r* was reported. The IDS-SR had a median convergent validity correlation of .91 when compared with the HRSD-17, BDI, and IDS-C. Rush et al. (2006) reported 89% agreement in scores between the IDS-C and IDS-SR in a sample of patients diagnosed with non-psychotic

MDD who responded to psychopharmacological treatment. The same study reported 91% agreement on MDD remission between the clinician-rated and self-report permutations in a sample of patients.

Rush and colleagues (1996) reported two concurrent validity trials: one where measure scores were measured against the fifth digit severity code in the *DSM-III-R*, and another where they were compared with the HRSD and BDI. The authors reported correlations of .54 for the IDS-SR with the *DSM-III-R* fifth digit. The IDS-SR had correlations of .88 with the HRSD and . 93 with the BDI.

Vittengl and colleagues' study (2005) found a 2-factor structure for the IDS-C and IDS-SR using varimax orthogonal rotation and multiple time points for the same measure. The two factors were "late" and "early," thus scores before treatment tended to agree more than those after treatment. These results were used to conclude that the IDS-SR reflects a single general depression factor. The original Rush et al. (1996) study found a 3-factor structure for the IDS-SR using a principal components analysis with varimax rotation. The IDS-C factors were described as: cognitive/mood, anxiety/arousal, and somatic complaints. The IDS-SR's third factor included appetite regulation and an item for leaden paralysis that characterizes atypical subtypes.

Patient Health Questionnaire - 9.

The Patient Health Questionnaire depression scale (PHQ-9) is a nine-item questionnaire derived from the full PHQ (Kroenke & Spitzer, 2002). The PHQ is a self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) which measures symptoms for five common *DSM-IV* disorders (Nezu et al., 2008). The PHQ-9 has a question for each *DSM* symptom for a depressive episode, which individuals can respond to on a scale of 0 to 3. Total scores can range from 0 to 27, with the following suggested cut-offs: 1 - 4, none; 5 - 9, mild; 10 -

14, moderate; 15 - 19, moderately severe; and 20 - 27, severe (Kroenke & Spitzer, 2002).

With regard to reliability, the PHQ-9 had a Cronbach's alpha coefficient of .86 - .89 in a large sample of hospital patients (Kroenke, Spitzer, Williams, & Lowe, 2010). In the same sample, the test-retest reliability coefficient for 48 hours was .84 (Kroenke, Spitzer, & Williams, 2001). Using a cutoff of 10, the PHQ-9 has demonstrated sensitivity and specificity of .88 in the same sample for any depressive disorder (Kroenke & Spitzer, 2002). With regard to convergent validity, the PHQ-9 had a correlation of .46 and .63 with the SCL-20, before and after treatment, respectively. The PHQ-9 had a greater sensitivity to change than the SCL-20 (d = .71 vs. d = . 91) To date, no factor analysis has been reported in the validation literature.

Procedure

Informed participants completed an on-line battery of measures including demographic information, BDI-II, CES-D, IDAS, IDS-SR, and PHQ-9. To be included in the study, participants PHQ-9 score was at or above a 5. Participants were given course credit as compensation for participation.

Statistical Analyses

To test Hypotheses 1 – 14 and calculate other psychometrics, unweighted factor-based subscales were calculated by summing the scores on individual items that have previously been linked to a factor in the literature, and a correlation matrix including all subscales was created. Correlations were compared using an effect size difference to determine if the difference was small (.10), medium (.30), or large (.50), paralleling Cohen's (1992) use of these descriptors for effect sizes of individual correlations.

Descriptive and psychometric goals.

Psychometrics of subscales.

Psychometric statistics were calculated for each measure and factor-based subscales

derived from previous factor analyses.

a. To evaluate internal-consistency reliability, Cronbach's alpha was computed for the general measure as well as each subscale.

b. As evidence of discriminant validity, correlations between total scores for each scale and the three IDAS anxiety scales were conducted.

c. To test the adequacy of previously defined structures, confirmatory factor analyses were conducted to determine the fit of previous factor structures on a population of symptomatic undergraduates. There are multiple fit statistics, of which four will be used in this study: chi-square test, comparative fit index (CFI), Tucker-Lewis index (TLI), root-mean square-error of approximation (RMSEA), and weighted root mean square residual (WRMR). According to conventions, desirable fit includes a non-significant chi-square, although in actuality, the chi-square test tends to be sensitive to any difference (Kline, 2011). Adequate fit occurs when models have a CFI and TLI of .90 - .95, and RMSEA of $\leq$.06 (Kline, 2011; Osman et al., 2008). WRMR is currently an experimental value, although values close to .95 are suggested (Brown, 2006).

Item-level descriptive comparison of measures.

To fulfill this aim, canonical correlation analysis (CCA) was used. CCA is a multivariate statistical technique that is easily construed as a cross between regression and factor analysis. Essentially, it determines how many latent variables (called variates) are shared across two groups of variables and determines how much variance each variate accounts for in the groups of variables.

One statistic that canonical analysis yields is the redundancy statistic. Essentially, this redundancy statistic reflects the amount of variance accounted for in one group of variables by the latent variables in another group of variables. Although varying measures of depression have

strong correlations by Cohen's (1992) standards, they do not necessarily have overlapping factors. The use of CCA may consolidate these seemingly discrepant findings to further elucidate the relationship between measures of depression and determine whether we should view them as measuring the same construct, or as measuring subtly different constructs. To date, only one study has used canonical analysis to look empirically at the overlap between measures of depression (Suzuki et al., 1995), and only did so with Japanese-language measures designed for the study. If there are large differences in variance accounted for by measures (low redundancy), this would lend further support to NIMH's recommendation to focus on subscale scores instead of general scores.

Conversion tables.

Regression methods were used to provide helpful conversion tables between different measures of depression in symptomatic student populations. Using the obtained regression coefficient and constant, different formulas are provided for the different scales. For example, if a clinician wanted to see if their client's current score on the IDS matches a study that measured depression with the CES-D, they can insert their client's IDS score into the equation provided and obtain the equivalent CES-D score. Equations are provided for every pair in the four measures administered.

IDS validation.

To validate the IDS in a college sample, the same criteria used for subscales were applied to the IDS. Namely:

a. To evaluate internal-consistency reliability, Cronbach's alpha was computed for the general measure as well as each subscale.

b. As evidence of discriminant validity, correlations between total scores for each scale and the three IDAS anxiety scales were used.

c. To test the adequacy of previously defined structures, confirmatory factor analysis was conducted to determine the fit of previous factor structures on a population of symptomatic undergraduates.

CHAPTER 3

RESULTS

Descriptives

Data were screened for outliers, normality, linearity, heteroscedasticity, and multicollinearity. Analyses indicated that no data was out of range and all means matched expected means from the literature for each measure. Means in this sample were generally higher than those reported in normative non-patient samples, but lower than those reported for those seeking treatment in inpatient or outpatient samples. There were some outliers, but these outliers were consistent with other data on these highly distressed individuals (e.g., current diagnoses of bipolar disorder), and were kept without transformation for subsequent analyses. Measures of central tendency, variability, and range are reported in Tables 1 and 2 for the overall scales and respective subscales. All overall scales had means consistent with expectations for a symptomatic sample.

Looking at general summary scores and subscales, no skewness statistic was above 1.3 and no kurtosis statistic was larger than 1.9. According to Kline (2011), skewness values above 3 and kurtosis values above 10 tend to lead to problems with subsequent analyses, so these data are judged to be sufficiently normal. However, visual analysis of probability plots and histograms indicated some deviance from normality (i.e., positive skew) for the Beck Depression Inventory (BDI) Sum, BDI Somatic subscale, Center for Epidemiologic Studies Depression scale (CES-D) Interpersonal subscale, Inventory of Depression and Anxiety Symptoms (IDAS) Insomnia subscale, and IDAS appetite change subscales.

Bivariate scatterplots between all general summary scores and subscales were visually analyzed; all plots appeared to reflect linear relationships, without bivariate outliers. Scatterplots of the variance between all general summary scores and subscales indicated some deviance from

homoscedasticity. However, heteroscedasticity is not a large a concern when using continuous variables, although it can reduce the accuracy of prediction (Tabachnick & Fidell, 2007). A solution for violations of homoscedasticity is to transform variables. However, only the relationships between variables can be interpreted, excluding interpretation of original scores. Given that this study aims to help clinicians and researchers empirically transfer scores from one measure to another it is important for interpretability that scores stay on their original scales; thus using transformations of scores is inconsistent with this study's purpose.

In regard to demographic variables that may confound relationships, ethnicity, age, and gender were not related to any of the main study variables. The relationships among all summary scales were strong and positive. All correlations were statistically significant and "large" by effect size conventions (Cohen, 1992). These results are summarized in Table 3.

Subscale-Level Psychometric Goals

Internal consistency reliabilities.

First, the author calculated internal consistency reliability statistics for each of the subscales from the literature. Results are summarized in Tables 3 and 4. According to conventions, “adequate” Cronbach's alpha would be considered to be around .70, whereas “very good” would be around .80, and “excellent” around .90 (Kline, 2011). For the BDI-II, the internal consistency of the entire scale was in the expected range ($\alpha = .91$). The Cognitive-Affective subscale also had an adequate internal consistency statistic ($\alpha = .90$), but the Somatic-Vegetative subscale did not ($\alpha = .69$). Cronbach's alphas are not generally reported in the literature, so it is difficult to tell if these statistics are consistent with the literature.

The CES-D had adequate internal consistency as an entire scale ($\alpha = .74$), but was lower than generally found in the literature. Two of the subscales had stronger internal consistency reliabilities compared to the entire scale, Negative Affect ($\alpha = .87$) and Positive Affect ($\alpha = .86$).

The Somatic subscale had an adequate Cronbach's alpha statistic (α = .72), but the interpersonal subscale did not (α = .64). Alphas have not generally been calculated in the literature for the CES-D subscales, so it is difficult to tell if this is an artifact of this sample.

The IDAS is a slightly different case. There is a general depression scale and then separate subscales; some items from the subscales are used to create the general depression scale, but additional items are added to create a subscale for specific symptoms. All 11 scales in the IDAS had adequate Cronbach's alphas (αs > .70), with most (8) being above .85. These values are consistent with previous studies using the IDAS.

The IDS had very good internal consistency reliability at the overall scale level (α = .86). The Cognitive/Mood and Anxiety/Arousal subscales also had adequate Cronbach's alpha values (αs = .80 & .79). However, the Sleep subscale did not (α = .38). The low alpha for the Sleep subscale is consistent with the original validation study for the IDS (Rush et al., 1996).

Correlations/hypothesis tests.

To test convergent and discriminant validity, hypotheses were tested for each scale's subscales in relation to subscales from other measures. These hypotheses were based on differences in effect sizes of relationships between subscales and theoretical expectations. The correlations are reported in Table 4.

BDI-II scales.

It was hypothesized that the BDI-II Cognitive-Affective subscale would have stronger correlations with related subscales on other measures than unrelated subscales on the same measure (Hypotheses 1 – 4). All four hypotheses were supported; the strongest correlations were with the CES-D Negative Affect, CES-D Positive Affect, IDS Cognitive/Mood, and IDAS Dysphoria and Suicidality subscales. These correlations were larger than correlations with other subscales with a small to medium sized difference. In addition, the BDI-II Cognitive-Affective

subscale had stronger correlations with these subscales than with the IDAS anxiety scales (i.e., traumatic intrusions, social anxiety, and panic), although the relationship with the anxiety scales was medium in effect size.

Hypotheses 5 – 7 focused on the Somatic-Vegetative subscale of the BDI-II and its correlations with associated subscales on other measures. Only hypothesis 5 was fully supported; the Somatic-Vegetative subscale's correlation with the CES-D Somatic subscale was larger than its correlation with the other CES-D subscales. Hypothesis 6 was not supported; the Somatic-Vegetative subscale had stronger correlations with the IDS Cognitive/Mood subscale than with the Anxiety/Arousal and Sleep subscales. Hypothesis 7 was partially supported, with the Somatic-Vegetative subscale having a stronger correlation with the IDAS insomnia, lassitude, and appetite gain scales than with the IDAS ill-temper and suicidality scales. However, its correlations with the IDAS insomnia, lassitude, and appetite gain scales were negligibly different from its relationship with the IDAS dysphoria scale. The Somatic-Vegetative subscale also had medium-sized correlations with the IDAS anxiety scales, which were negligibly different from its relationships with the IDAS insomnia, lassitude, and appetite gain scales.

CES-D scales.

Hypotheses 8 and 9 focused on the Somatic subscale of the CES-D. Hypothesis 8 was not fully supported; the CES-D Somatic subscale had approximately equal correlations with the IDS Cognitive/Mood and Anxiety/Arousal subscales. Hypothesis 9 was partially supported; the Somatic subscale's relationship with the IDAS Insomnia and Lassitude scales was stronger than its relationships with the IDAS Suicidality scale, but not the IDAS Ill-temper or Dysphoria scales. However, the Somatic subscale's relationship with the IDAS Appetite change scales was not significantly different from the other non-somatic IDAS scales. The Somatic subscale did have a stronger correlation with the IDAS Appetite Gain than Appetite Loss scale, with a small-

sized difference. The CES-D Somatic subscale had approximately equal relationships with the BDI-II Somatic-Vegetative and Cognitive-Affective subscales. Furthermore, the Somatic subscale had medium-sized correlations with the IDAS anxiety scales.

Hypotheses 10 and 11 focused on the Depressed Affect subscale of the CES-D. Both hypotheses were supported, indicating that the Depressed Affect subscale had strong correlations in the expected direction with the IDS Cognitive/Mood subscale and IDAS Dysphoria, Ill-temper, and Suicidality scales. In addition, the CES-D depressed affect subscale had a stronger correlation with the BD-II Cognitive-Affective subscale than with the BDI-II Somatic-Vegetative subscale. The Depressed Affect subscale had stronger correlations with the BDI-II Cognitive Affective, IDAS Dysphoria, and IDS Cognitive/Mood subscales than the IDAS anxiety scales.

No specific hypotheses focused on the CES-D Positive Affect subscale or Interpersonal Problems subscale. However, the Positive Affect and Negative Affect subscales had very strong relationships in opposite directions with the aforementioned scales; the notable exception being having no relationship with the IDAS Lassitude or Insomnia scales. The Positive Affect subscale also had a very strong positive relationship with the IDAS Well-Being scale and small, negative correlations with the IDAS anxiety scales. The Interpersonal Problems subscale had its strongest positive correlations with the Cognitive subscales from the BDI-II and IDS, as well as the IDAS Dysphoria and Suicidality scales. It also had large correlations with the IDAS anxiety scales.

IDAS scales.

No specific hypotheses were created for the IDAS scales because its relationships were tested in previous hypotheses and because the IDAS introduced many new scales that have no appropriate analog in other measures. The IDAS Dysphoria and Suicidality scales had their strongest relationships with the cognitive-mood subscales on other measures (i.e., BDI-II Cognitive-Affective, CES-D Negative Affect, IDS Cognitive/Mood). The IDAS Lassitude and

Insomnia scales had their strongest relationships with the BDI-II Somatic-Vegetative subscale and the CES-D Somatic subscale. The IDAS Appetite change scales had a small negative relationship with each other. However, they had different-sized positive relationships with other subscales. The Appetite Loss scale had its strongest relationships with the CES-D Somatic subscale, and the Appetite Gain scale had its strongest relationship with the IDAS Social Anxiety scale. The IDAS Ill-temper scale had its strongest relationships with the CES-D Somatic and Interpersonal Problems subscales.

IDS scales.

Hypothesis 12 focused on the IDS Cognitive/Mood subscale and was supported; the Cognitive/Mood subscale had stronger relationships with the IDAS Dysphoria and Suicidality scales than with the IDAS's other scales. In addition, the Cognitive/Mood subscale had stronger relationships with the BDI-II Cognitive-Affect subscale and CES-D Negative Affect subscale than with other subscales in their respective scales. Hypothesis 13 was not supported; the IDS Arousal/Anxiety subscale's relationship with the IDAS Dysphoria scale was negligibly different from its relationships with the IDAS Ill-temper and Lassitude scales. However, the Arousal/Anxiety subscale did have its strong relationships with the IDAS anxiety scales, although these were negligibly different than its relationship with the IDAS Dysphoria scale. Hypothesis 14 was partially supported; the IDS Sleep subscale had a large relationship with the IDAS Insomnia scale, but a small relationship with the IDAS Lassitude scale. The IDS Sleep subscale also had stronger relationships with the BDI-II Somatic-Vegetative and CES-D Somatic subscales than with other subscales in their respective measures.

Confirmatory factor analyses.

CFA techniques were used to provide information on whether previously established factor structures had adequate fit in the current sample. Using the polychoric correlation matrix

implicated for ordinal indicators, most of the models did not have good fit with the data as indicated by absolute and relative fit indices. In light of this, two alternative models posited in the literature were also included as comparison, a three-factor BDI-II model and a one-factor IDS model. Results are summarized in Table 5. Specifically, the BDI-II three-factor model had better fit than the BDI-II two-factor model. The IDS three-factor model had better fit than the IDS one-factor model. Of the models tested, the CES-D had the best model fit with previous CES-D literature.

Canonical Correlation Analyses

Another psychometric goal was to use canonical correlation analysis (CCA) to explore the relationship between pairs of measures and shared latent variables. Statistics related to CCA are summarized in Tables 6-9 for each measure. As per Tabachnick and Fidell (2007) only correlations above .30 are interpreted and reported in the tables. First, the number of significant canonical correlations was observed as an indicator of shared latent variables, then redundancy statistics were interpreted as indicators of variance explained by shared latent variables.

First, the BDI-II had four statistically significant canonical correlations with the CES-D, seven with the IDAS, and eight with the IDS. Second, the CES-D shared four canonical correlations with the BDI-II, eight with the IDAS, and five with the IDS. The IDAS shared seven with the BDI-II, eight with the CES-D, and five with the IDS. The IDS shared eight with the BDI-II, five with the CES-D and five with the IDAS. In all cases, the first canonical correlation was reflective of a general depression depression latent variable, although sleep did not tend to group together with this.

With regard to redundancy, all measures of depression had a statistically large amount of redundancy (23% - 41% redundancy). Redundancy statistics are reported in Tables 6 - 9. Of all the measures, the IDS had the lowest redundancy with other measures (23% - 28% redundancy).

This indicates that the IDS, more than other measures, assesses distinct constructs not assessed by other measures.

Communicative Goals

Another set of analyses focused on comparing different measures at a summary score level to allow researchers to empirically compare summary scales. Regression analyses are summarized in Table 10. All relationships were statistically significant. When the CES-D was predicting the BDI-II, a probability plot of residuals indicated that the relationship may not be exactly linear. A quadratic regression model was fitted, and it explained a statistically significant amount of variance beyond the original model. The same occurred for all scales that the IDAS was predicting, as well as when the IDS predicted BDI-II scores. Although the quadratic models explained statistically more variance, the practical change was negligible ($\Delta R^2 = .01$-$.02$). Given the theoretical complexity, both linear and quadratic equations are included in Table 10.

Overall, the IDAS general depression scale explained the most variance in other scales and had the smallest error of estimate. The CES-D explained the least variance in other scales on average, and had the largest standard error of estimate with other scales on average. Table 11 provides regression equations for clinicians and researchers to use to predict an individual score and obtain a 95% confidence interval. This information provides an empirically-informed method of estimating individual scores from other measures of depression instead of assuming that different measures are equivalent.

CHAPTER 4

DISCUSSION

The purpose of this study was to: a) determine whether various depression scale factors with extensive validation in the psychological literature have adequate psychometric characteristics (determined using specific hypotheses) as subscales in a sample of mildly depressed individuals; b) use CCA to better describe shared latent variables in measures and look at the redundancy between measures based on these shared latent variables; c) provide empirical conversion tables for total scores, so that if clinicians and researchers choose to compare results between measures, they do not automatically assume that categorical markers are equal; d) provide initial information to show that the IDS is a valid measure to use in a student population.

The results demonstrated that most of the overall measures did not yield psychometrically reliable and valid subscales, with the notable exception of the IDAS. Further analysis with CCA indicated that there was relatively little shared variance between overall measures based on overlapping latent variables (23% - 41%). These results were surprising given that regression with general summary scores predicts more variance accounted for (31% - 61%). The provided regression tables also demonstrate that although there is a positive relationship between overall measures of depression, specific individuals may score higher on one measure than another because of differences in the latent constructs being measured.

This has implications for the way clinicians and researchers measure depressive symptoms as well as the assumptions of cross-measure equivalence. The field should be cautious when selecting measures of depressive symptoms and carefully weigh the focus of their study with the different measure characteristics highlighted in this study. This careful consideration can yield a fine grained image of the currently heterogeneous construct of depression.

Factor Psychometrics

The results were varied with regard to the psychometric adequacy of factor-based subscales from research-supported factor structures for common measures. Possible solutions for problematic subscales are discussed in the future directions section. The BDI-II's Cognitive-Affective subscale had adequate psychometrics: good internal consistency, and adequate convergent and discriminant validity. The Somatic-Vegetative subscale did not have acceptable psychometrics. There was low internal consistency and although it had statistically significant correlations with expected scales, the discriminant validity was problematic, with evidence of stronger correlations with unexpected scales, such as IDAS Dysphoria or IDS Cognitive/Mood. However, these correlations may be partially explained by the fact that the IDS Cognitive/Mood subscale contains items related to difficulties with appetite changes, weight changes, and energy and the IDAS Dysphoria scale contains two items related to concentration.

For the CES-D, the Somatic subscale demonstrated adequate internal consistency and convergent validity. However, there were some issues with discriminant validity, evidenced by a stronger relationship with the IDAS Dysphoria scale than other theoretically similar scales, perhaps related to the two items in the Dysphoria scale. The Negative Affect and Positive Affect subscales both demonstrated adequate psychometric characteristics: adequate internal consistency coefficients, and convergent and discriminant validity correlations. The Interpersonal scale was problematic to test because there were no adequate analogs for convergent and discriminant validity in other scales given that interpersonal problems are not *DSM-IV* symptoms for depression. In general, it had statistically significant correlations with all scales, the largest being with IDAS Social Anxiety and IDS Arousal/Anxiety. However, it had poor internal consistency, probably related to consisting of two items.

For the IDAS, all the scales demonstrated adequate internal consistency. Convergent and

discriminant validity were difficult to test because many of the scales were specific and did not have analogues in other scales. Generally, scales demonstrated convergent validity by statistically significant correlations with similar subscales. Discriminant validity was difficult to measure, and there were varying results for the appetite change scales. The Appetite Loss scale had adequate discriminant validity with the CES-D subscales, but not with the IDS subscales. The Appetite Gain scale generally had poor discriminant validity, evidencing small to medium correlations with most subscales. It had its strongest relationship with the IDAS Social Anxiety scale.

Two of the IDS subscales, Cognitive/Mood and Arousal/Anxiety, tended to have adequate psychometrics. They had very good internal consistency and convergent validity. There were issues with discriminant validity for the Arousal/Anxiety subscale when looking at appetite change scales in the IDAS and anxiety scales from the IDAS. This reflects the fact that the IDS Arousal/Anxiety subscale does not have items related to appetite or weight. The "anxiety" in the Arousal/Anxiety subscale seems to best tap panic-type symptoms. The Sleep subscale did not have adequate psychometrics: internal consistency was poor, although the convergent and discriminant validity tended to relate in expected directions.

Summary and recommendations.

Based on the above information, it is clearly recommended that researchers use the IDAS. The 64-item scale provides a general depression scale that overlaps with other measures of depression. However, it also has additional theoretically-based scales with adequate psychometrics that measure components of depression. Furthermore, it measures all symptoms required for a *DSM-IV* diagnosis of depression.

If a researcher plans to measure components of depression using older measures such as the BDI-II, CES-D, or IDS, a combined approach is recommended. The CES-D has

psychometrically sound subscales. However, it does not measure all *DSM-IV* symptoms and has a different time length than the *DSM-IV*(one week). The BDI-II and IDS are attuned to *DSM-IV* diagnostic criteria, but do not have sound factors (e.g., the BDI-II Somatic does not contain irritability but contains concentration difficulties, and the IDS Cognitive/Mood contains appetite and weight items). Combining the CES-D with either the IDS or BDI-II provides complete coverage for fine-grained measurement of depression.

Canonical Correlation Analysis

The results of CCA indicate that not all overall measures are created equal; some have more overlap than others. In general, the shared latent variables of overall measures explained much less variance than would be expected based on theory (since these measures are supposed to measure the same construct) and correlations of summary scores (with a typical correlation of .70, explained variance should be around 50%). The redundancy statistics ranged from .23 - .41, meaning that 23% - 41% of the variance was explained by shared latent constructs in measures. The overall CES-D and IDAS General Depression scales had the most shared variance, meaning that they largely tapped similar latent constructs. The IDS tended to have the least shared variance across measures; it had a mean redundancy statistic of .25, meaning that it tapped latent constructs that other scales did not assess. This is not surprising when observing specific items content; there are four items related to sleep and others related to physical pain, general anxiety, and bowel movements.

Another finding was that the first variate across all overall measures tended to include a general depression latent construct that excluded sleep items. After that, several different variates tended to appear that varied in number and content across measures. The variates did not coincide with factors found in the literature. Instead, the variates tended to include items related to single symptoms, such as concentration difficulties, anhedonia, suicidal ideation, or

sleep changes. This is not surprising given our analysis demonstrating that similarly worded subscales (i.e., Somatic-Vegetative and Arousal/Anxiety) did not measure the same symptoms (i.e., appetite and weight changes). Thus the same symptoms from different overall measures were grouping into single variates.

One possible explanation for this is sample characteristics. Although our student sample was symptomatic, the student environment is one that fosters difficulties in sleep due to a number of environmentally-specific conditions, such as irregular sleeping patterns, stress, poor study habits, inadequate time management abilities, and recreational activities. Thus, sleep may not map on to the general depression variate because responses to these items are influenced by factors other than depression.

The implications of these findings are that general summary scores may not be the best use of these measures because they may be measuring subtly different constructs (e.g., the BDI-II having seven items asking about cognitions, or IDS having four items asking about sleep). Although a single general depression variate seems to be present, it does not seem to adequately explain the all variance in the overall measures because of the number of statistically significant variates extracted. These numerous variates tended to focus on diagnostic symptoms and varied in size depending on the number of items each overall measure had for that symptom. Research should continue focusing on the use of factor-based subscales using the information provided in the previous section to provide fine-grained information about the nature of depressive symptoms.

Another reason why the overall measures had less shared variance than expected is that there are slight differences, such as wording of the questions, response formats, and the scale used (e.g., 0-3 or 1-7). Although this should introduce systematic method error in measurement when comparing scores, it should not result in such stark differences in shared variance, and

regardless, this explanation also raises questions about the assumption that different overall measures equivalently assess the same construct.

Conversion Tables

The conversion tables listed are important for a number of reasons. First, instead of assuming that categories within distributions are equivalent across measures, this study shows that the same individuals can be labeled as having different levels of severity depending on the measure they are administered. For example, an individual scoring a 6 on the BDI-II is considered "minimally depressed." However, this same individual is likely to score higher on the CES-D, scoring a 18. Although there are no ordinal conventions for the CES-D, this score is considered to have high levels of sensitivity (>87%) and specificity (>77%) for MDD and is past the suggested MDD cut-score of 16. Using the regression formulas in Table 11, a score of 18 on the CES-D is translated to a score of 11 on the BDI-II, which is within the same category as the original score of 6. As another example, an individual scoring a 6 on the BDI-II would be labeled as minimally depressed, but that same individual would score around a 14 on the IDS, which would label them as mildly depressed.

One may notice that there is not equivalent back-translation of scores (i.e., the BDI-II score of 6 becomes 11 when converting back from the predicted CES-D score). This is because there is greater error in regression for high or low values. The error of prediction decreases asymptotically (and accuracy of score back-translation increases) as the score approaches the mean. When using this conversion table, individuals should exercise more interpretive caution when the scores are different from the means reported in Tables 1 and 2.

Given that each scale appears to assess somewhat different latent and manifest content, these conversion tables are important for intervention research where remission of symptoms is assessed with self-report measures. In some cases individuals may be labeled as likely having

MDD with one measure (e.g., CES-D), when they are labeled as having minimal symptoms using another measure (e.g., BDI-II). This may lead to bias in the results of interventions and influence their implementation and funding of further research. Thus it is of the utmost importance to take into account subtle differences between measures of depression.

IDS Validation

Another contribution of this study was its assessment the psychometric characteristics of the IDS in a student population. To date there are no studies of this nature, which is an important oversight given the different environmental presses in the college environment.. Using a sample of symptomatic undergraduates, the IDS as a measure appears to have good psychometrics. Internal consistency was very good ($\alpha = .86$) at the overall scale level in this symptomatic sample. There was adequate convergent validity with other depression measures (all large effect sizes) and discriminant validity (smaller correlations with IDAS anxiety scales). This study also showed that a 3-factor structure (Cognitive/Mood, Arousal/Anxiety, and Sleep) had better fit compared with the 1-factor structure that has been suggested in the literature.

Limitations

This study has a number of limitations that need to be addressed. One weakness is that it used a collegiate sample. Although all individuals included in this study were at least mildy depressed using criteria on the PHQ-9, there are still environmental circumstances that set college students apart from other groups. One factor that can influence studies of depression scale validity is erratic sleep patterns that can influence the error variance associated with responses to sleep items. There are also lifestyle differences in diet and exercise that can influence responses to items.

Another important weakness is the sample size. Although this sample is considered large enough for most CFA studies, most structural modeling yields results with more confidence

when sample sizes are larger. A larger sample size would yield stronger fit statistics for CFA as well as smaller confidence intervals for the predictions in the conversion tables.

Future Research and Recommendations

With regard to recommendations, researchers should carefully consider the measures that they use. They measure slightly different constructs, which may explain why some intervention studies have found significant effects with some overall measures but not others. Based on the results of this studies, the IDAS is highly recommended for its complete coverage of *DSM-IV* symptoms and various psychometrically sound scales that measure components of depression. If using more traditional measures, a combined approach of the CES-D with the IDS or BDI-II is recommended. Future research may contribute to the literature by suggesting modifications to factor-based subscales that improve their psychometric characteristics. That was beyond the scope of this study, which had a more descriptive focus. Other research can work on extending these results to clinical and community samples to provide regression-based conversion tables for different populations. Informed and detailed approach to assessment of depressive symptoms can improve the state of our science and reduce the incidence of major depression for future generations.

Table 1

Descriptives of Summary Scales

	Mean	Median	*SD*	Range
BDI-II	14.7	13	9.6	0 - 58
CES-D	21.9	20	7.6	3 - 46
IDAS general depression	49.3	48	12.2	27 - 99
IDS	22.0	20	10.8	3 - 70

Note. N = 219. SD = standard deviation. BDI-II = Beck Depression Inventory - 2nd edition.

Table 2

Descriptives for Subscales

	N of items	Mean	Median	SD	Range
BDI-II Cognitive-Affective	15	9.5	8	7.5	0 - 40
BDI-II Somatic-Vegetative	6	5.2	5	2.9	0 - 18
CES-D Somatic	6	6.5	6	3.6	0 - 18
CES-D Negative Affect	7	6.2	5	4.9	0 - 21
CES-D Positive Affect	4	6.8	7	3.2	0-12
CES-D Interpersonal	2	1.4	1	1.5	0 - 6
IDAS Dysphoria	10	24.6	24	7.8	10 - 49
IDAS Lassitude	6	16.3	16	4.5	7 - 30
IDAS Suicidality	6	8.2	6	4.3	6 - 30
IDAS Insomnia	6	14.1	14	6.0	6 - 30
IDAS Appetite Loss	3	6.0	5	3.3	3 - 15
IDAS Appetite Gain	3	6.5	6	3.1	3 - 15
IDAS Ill-temper	5	9.3	8	4.3	5 - 23
IDAS Well-being	8	22.5	23	6.5	8 - 40
IDS Cognitive/Mood	12	9.2	8	5.8	0 - 33
IDS Arousal/Anxiety	11	8.2	7	4.8	0 - 28
IDS Sleep	4	3.7	4	2.3	0 - 9

Note. N = 219. SD = standard deviation. BDI-II = Beck Depression Inventory - 2nd edition. CES-D = Center for

Table 3

Correlations Among Summary Scales

	1	2	3	4
1. BDI-II	[.91]			
2. CES-D	.56	[.74]		
3. IDAS	.74	.68	[.87]	
4. IDS	.78	.60	.71	[.86]

Note. Cronbach's alpha in diagonal. BDI-II = Beck Depression Inventory 2nd edition. CES-D = Center for Epidemiologic Studies Depression

Table 4

Correlations Among Subscales

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. BDI-II Cognitive	[.90]																			
2. BDI-II Somatic	.65	[.69]																		
3. CES-D Somatic	.55	.56	[.72]																	
4. CES-D Negative Affect	.70	.44	.56	[.87]																
5. CES-D Positive Affect	-.68	-.40	-.30	-.58	[.86]															
6. CES-D Interpersonal	.46	.25	.31	.48	-.27	[.64]														
7. IDAS Dysphoria	.71	.54	.67	.68	-.58	.47	[.88]													
8. IDAS Lassitude	.31	.43	.49	.26	**-.10**	.18	.52	[.70]												
9. IDAS Suicidality	.48	.29	.31	.49	-.31	.40	.46	.20	[.92]											
10. IDAS Insomnia	.25	.38	.54	.19	**-.07**	.19	.40	.35	.25	[.86]										
11. IDAS Appetite Loss	.22	.21	.44	.28	-.18	.17	.23	**.12**	.27	.37	[.92]									
12. IDAS Appetite Gain	.24	.34	.22	.23	-.18	.24	.32	.31	.22	.27	-.13	[.83]								
13. IDAS Ill-temper	.39	.20	.42	.39	-.20	.42	.50	.35	.37	.44	.20	.32	[.87]							
14. IDAS Well-being	-.59	-.39	-.24	-.46	.74	-.17	-.47	**-.08**	-.15	**-.02**	**-.11**	**-.10**	**-.09**	[.89]						
15. IDAS Traumatic Intrusions	.43	.32	.47	.50	-.23	.47	.49	.36	.51	.38	.28	.29	.57	-.10	[.82]					
16. IDAS Social Anxiety	.47	.39	.47	.40	-.26	.50	.55	.38	.39	.34	**.11**	.43	.47	-.18	.55	[.86]				
17. IDAS Panic	.39	.33	.48	.43	-.18	.48	.49	.37	.64	.43	.30	.35	.54	**-.07**	.67	.67	[.89]			
14. IDS Cognitive/Mood	.80	.67	.53	.71	-.64	.43	.64	.30	.51	.26	.28	.36	.31	-.58	.43	.50	.43	[.80]		
15. IDS Arousal/Anxiety	.64	.55	.54	.51	-.41	.51	.57	.36	.43	.34	.23	.23	.41	-.15	.42	.47	.54	.67	[.79]	
16. IDS Sleep	.15	.27	.35	**.06**	**-.04**	**.12**	.13	.17	.22	.59	.24	.21	.20	-.19	.30	.34	.35	.23	.34	[.38]

Table 5

Fit of Extant Factor Structures in Current Sample

	N of factors	χ^2	*df*	CFI	TLI	RMSEA	WRMR
BDI-II	2-factor	192.32*	63	.88	.96	.10	1.12
	3-factor	165.31*	63	.90	.97	.09	1.02
CES-D	4-factor	149.79*	58	.93	.97	.09	.97
IDAS	1-factor	679.34*	42	.61	.75	.26	2.64
IDS	1-factor	340.78*	94	.71	.82	.12	1.40
	3-factor	208.94*	90	.85	.92	.08	1.10

Note. * = Statistically significant at p <.01. df = degrees of freedom. CFI = Comparative Fit Index. TLI = Tuck
Lewis Index. RMSEA = Root Mean-Square Error of Approximation. WRMR = Weighted Root Mean Square
Residual. BDI-II = Beck Depression Inventory – Second Edition. CES-D = Center for Epidemiologic Studies –

Table 6

Canonical Variate Correlations and Redundancy Analysis for BDI-II with Other Scales

			CES-D				IDAS						IDS						
Item	Factor (2)	Factor (3)	1st Var.	2nd Var.	3rd Var.	4th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	7th Var.
1 Sadness	C-A	NA	.80	.40			.82						.72					-.30	
2 Pessimism	C-A	NA	.71				.66						.69						
3 Past failure	C-A	NA	.68				.65						.65						
4 Loss of pleasure	C-A	PD	.67				.64			-.48			.66	.32					
5 Guilty feelings	C-A	NA	.54				.52						.64						
6 Punishment feelings	C-A	NA	.57				.57						.57						
7 Self-dislike	C-A	NA	.63				.57	.34					.63						
8 Self-criticalness	C-A	NA	.67				.64						.65						
9 Suicidal thoughts	C-A	NA	.51				.66	-.39		.38			.59	-.49		-.37			
10 Crying	C-A	SE	.50	.42			.51						.44						
11 Agitation	C-A	SE	.52		.52		.41		.40				.56					.39	
12 Loss of interest	C-A	PD	.68				.64		.35				.60	.38					
13 Indecisiveness	C-A	PD	.52				.44		.39		.32		.59						-.40
14 Worthlessness	C-A	NA	.81			-.34	.77						.78						
15 Loss of energy	S-V	PD	.40		.48	.40	.31	.33	.70				.43	.55					
16 Changes in sleep	S-V	SE			.57				.33			.53			.38				
17 Irritability	C-A	PD	.40				.37						.51			.35	-.46	.39	
18 Changes in appetite	S-V	SE	.33				.34				.43		.49		.55				.33
19 Concentration difficulty	S-V	PD	.57	-.55		.44	.47	.65		.33			.59				.33	.45	
20 Fatigue	S-V	PD	.38				.32		.49				.48	.35					
21 Interest in sex	S-V	SE	.42				.33						.47				.37		
Cumulative Explained Variance			.33	.37	.42	.46	.29	.35	.42	.46	.50	.54	.34	.40	.45	.48	.52	.56	.59
Cumulative Redundancy			.25	.27	.28	.30	.22	.24	.28	.30	.31	.32	.28	.31	.34	.35	.38	.39	.40

Table 7

Canonical Variate Correlations and Redundancy Analysis for CES-D with Other Scales

			BDI-II				IDAS								IDS				
	Item	Factor (4)	1st Var.	2nd Var.	3rd Var.	4th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	7th Var.	8th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.
1	Bothered	S	.40				.49			.31					.48				.31
2	Decreased appetite	S						.65	-.59										
3	Shake off blues	NA	.61				.75								.67	-.34			
4	Just as good as others	PA	-.74				-.66								-.67				
5	Concentration	S	.46	-.46		.61	.41			.59			-.47		.50				-.43
6	Depressed	NA	.76				.79				.40				.77				
7	Fatigue	S	.49				.50							.49	.54				
8	Hopeful	PA	-.61				-.60							.42	-.58				
9	Life failure	NA	.77				.69					.35			.67		.35		
10	Fearful	NA	.52				.62								.50		.45		
11	Restless sleep	S	.33		.78			.75	.51						.34	.61	-.51		.31
12	Happy	PA	-.68				-.70			.43					-.70			.36	
13	Talked less	S					.32								.31				
14	Felt lonely	NA	.61				.67				.32				.64	-.30			
15	People unfriendly	IP					.30					.35		-.36		.44			
16	Enjoyed life	PA	-.68				-.74					.38			-.71				
17	Crying	NA	.45	.52			.49								.47			.30	.31
18	Sad	NA	.67	.51			.78								.74				
19	Felt disliked	IP	.58				.62								.62		.37	.30	
20	Could not get “going”	S	.50		.47	.33	.52	.39		.36			.34		.57				
	Cumulative Explained Variance		.31	.37	.43	.48	.34	.40	.45	.50	.54	.58	.61	.66	.33	.38	.44	.48	.53
	Cumulative Redundancy		.23	.26	.28	.30	.28	.32	.34	.37	.38	.39	.40	.41	.26	.29	.31	.33	.34

Table 8

Canonical Variate Correlations and Redundancy Analysis for IDAS with Other Scales

		BDI-II						CES-D								IDS				
	Item	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	7th Var.	8th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.
1	Exhausted			.63					.37						.43			.43	.30	
2	Depressed	.79						.85								.76				
3	Inadequate	.80						.78								.72			-.32	
4	Restless	.32		.41				.31	.54	.35						.34	.38			.37
5	Suicide	.67	-.43		.37			.51					.39			.62		-.59		
6	Poor sleep						.57		.58	.48					.33		.63			
7	Self-blame	.68						.71								.70			-.39	
8	Trouble sleeping						.51		.62	.48							.89			
9	Discouraged	.66	.36					.72								.64				
10	Thought of hurting self	.59	-.38					.46								.53		-.49		
11	Poor appetite					.45			.68	-.53										
12	Decreased hunger					.53			.64	-.60										
13	Enjoyment	.55			-.51			.56			-.34		-.43		-.37	.56				
14	Energy		.41									-.41	-.45			.35		.42		
15	Anhedonia	.51			-.32			.58								.60			.42	
16	Concentrating	.41	.66	.30	.36			.46			.63			-.41		.46		.38		.45
17	Trouble making mind	.42	.35	.30				.49			.42					.45				.38
18	Slow speech	.39		.40				.43								.39				
19	Worry	.53						.66			.30					.65		.38		
20	Effort to get going	.34						.41	.40		.39		-.33	.41		.41		.36		
	Cumulative Explained Variance	.24	.32	.41	.46	.52	.57	.25	.37	.44	.51	.53	.59	.62	.65	.24	.34	.42	.47	.51
	Cumulative Redundancy	.18	.22	.26	.28	.30	.31	.21	.28	.32	.35	.36	.38	.38	.39	.19	.25	.30	.32	.34

Table 9

Canonical Variate Correlations and Redundancy Analysis for IDS with Other Scales

			BDI-II							CES-D					IDAS				
	Item	Factor (3)	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	6th Var.	7th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.	1st Var.	2nd Var.	3rd Var.	4th Var.	5th Var.
1	Falling asleep	S									.53	-.60				.85			
2	Sleep during night	S									.32	-.32				.55			
3	Waking early	S									.34					.33			
4	Sleeping too much	S										.38							
5	Sadness	C-M	.62				-.33	-.47		.74	-.30				.69				
6	Irritability	A-A	.44			.31	-.57	.33		.42					.40				
7	Anxiousness	A-A	.45							.53			.36		.57		.40		
8	Mood responsiveness	C-M	.43						.35	.46					.47				
10	Mood quality	C-M	.56					-.38		.63					.64				
11/12	Appetite change	C-M	.53		.57					.35					.31				
13/14	Weight change	C-M			.32	.30			.30										
15	Concentration	A-A	.55					.41		.53			.30		.46		.47		.44
16	Self-blame	C-M	.65							.37					.63			-.39	
17	View of future	C-M	.73							.72					.73				
18	Suicide	C-M	.64	-.38		-.40				.52					.59				.30
19	Anhedonia	C-M	.55							.57					.56				
20	Energy	C-M	.44	.51						.42					.41		.33	.41	
21	Capacity for pleasure	C-M	.61	.37					.42	.58					.63			.37	
22	Interest in sex	C-M	.47			.45				.30					.32		-.31		
23	Slowed down	A-A	.51							.40					.52				
24	Restless	A-A	.32								.46								.30
25	Bodily pain	A-A											.30						
26	Anxious symptoms	A-A	.32								.47			.37	.29				
27	Phobia	A-A	.39							.30					.37				
28	Bowell movements	A-A	.36								.32								
29	Interpersonal problems	A-A	.60							.62					.52				
30	Physical energy/lead	A-A	.42							.43					.36	.30		.34	
	Cumulative Explained Variance		.21	.25	.29	.32	.35	.39	.42	.20	.26	.30	.34	.37	.19	.26	.30	.34	.37
	Cumulative Redundancy		.17	.20	.22	.24	.25	.27	.28	.16	.19	.21	.22	.23	.15	.19	.22	.24	.25

Table 10

Results of Analyses for Predicting Scales Based on Other-Scale Scores

	Constant	b_1	b_2	β	*p*	Adjusted R^2	SE
BDI-II = x							
CES-D	15.36		.44	.56	<.01	.31	6.29
IDAS	35.45		.94	.74	<.01	.54	8.32
IDS	8.94		.86	.78	<.01	.61	6.71
CES-D = x							
BDI-II	-.84		.71	.56	<.01	.31	8.02
BDI-II Polynomial	9.41	.02	-.22			.33	7.85
IDAS	25.07		1.11	.68	<.01	.46	9.04
IDS	3.16		.86	.60	<.01	.35	8.73
IDAS = x							
BDI-II	-13.78		.58	.74	<.01	.54	6.52
BDI-II Polynomial	-1.08	.01	.06			.56	6.34
CES-D	1.29		.42	.68	<.01	.46	5.56
CES-D Polynomial	12.74	.00	-.05			.47	5.51
IDS	-8.82		.62	.71	<.01	.50	7.70
IDS Polynomial	2.45	.00	.17			.51	7.54
IDS = x							
BDI-II	-.60		.70	.78	<.01	.61	6.01
BDI-II Polynomial	2.04	.00	.46			.61	5.98
CES-D	12.75		.42	.60	<.01	.35	6.08
IDAS	31.69		.80	.71	<.01	.50	8.72

Note. b_1 = unstandardized polynomial regression coefficient. b_2 = unstandardized regression coefficie

β = standardized regression coefficient. SE = standard error of estimate. BDI-II = Beck Depression

Table 11

Regression Equations for Predicting Scales Based o Other-Scale Scores

	Equation	CI
BDI-II = x		
CES-D	15.36 + (.438)x	+/- 12.33
IDAS	35.45 + (.939)x	+/- 16.31
IDS	8.94 + (.875)x	+/- 13.15
CES-D = x		
BDI-II	-.84 + (.71)x	+/- 15.72
IDAS	25.07 + (1.11)x	+/- 17.72
IDS	3.16 + (.86)x	+/- 17.11
IDAS = x		
BDI-II	-13.78 + (.58)x	+/- 12.78
CES-D	1.29 + (.42)x	+/- 10.90
IDS	-8.82 + (.62)x	+/- 15.09
IDS = x		
BDI-II	-.60 + (.70)x	+/- 11.78
CES-D	12.75 + (.42)x	+/- 11.92
IDAS	31.69 + (.80)x	+/- 17.09

Note. CI = 95% confidence interval. BDI-II = Beck Depressio Inventory - 2nd edition. CES-D = Center for Epidemiologic Studies Depression scale. IDAS = Inventory of Depression an

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Arlington, VA: Author.

American Psychiatric Association. (2010). *Proposed revisions: Mood disorders.* DSM-5 Development. Retrieved from http://www.dsm5.org/ProposedRevisions/Pages/MoodDisorders.aspx

Andrews, G., Anderson, T. M., Slade, T., & Sunderland, M. (2008). Classification of anxiety and depressive disorders: Problems and solutions. *Depression and Anxiety, 25*, 274-281.

Barefoot, J. C., Brummett, B. H., Helms, M. J., Mark, D. B., Siegler, I. C., & Williams, R. B. (2000). Depressive symptoms and survival of patients with coronary artery disease. *Psychosomatic Medicine, 62*, 790-795.

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

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.

Blatt, S. J. (2004). *Experiences of depression: Theoretical, clinical, and research perspectives*. Washington, DC: American Psychological Association.

Broderick, J. E., Junghaenel, D. U., & Schwartz, J. E. (2005). Written emotional expression produces health benefits in fibromyalgia patients. *Psychosomatic Medicine, 67*, 326-334.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Burton, R. (1621/2001). *The anatomy of melancholy*. New York: New York Review Books Classics.

Carragher, N., Adamson, G., Bunting, B., & McCann, S. (2009). Subtypes of depression in a nationally representative sample. *Journal of Affective Disorders, 113*, 88-99.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Corruble, E., Legrand, J. M., Zvenigorowski, H., Duret, C., & Guelfi, J. D. (1999). Concordance between self-report and clinician's assessment of depression. *Journal of Psychiatric Research, 33*, 457-465.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10*, 83-89.

Endler, N. S., Rutherford, A., & Denisoff, E. (1999). Beck Depression Inventory: Exploring its dimensionality in a nonclinical population. *Journal of Clinical Psychology, 55*, 1307-1312.

Faries, D., Herrera, J., Rayamajhi, J., DeBrota, D., Demitrack, M., & Potter, W. Z. (2000). The responsiveness of the Hamilton Depression Rating Scale. *Journal of Psychiatric Research, 34*, 3-10.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149 - 1160.

Gillham, J. E., Shatté, A. J., & Freres, D. R. (2000). Preventing depression: A review of cognitive-behavioral and family interventions. *Applied & Preventive Psychology, 9*, 63-88.

Greenberg, P., Kessler, R., Nells, T., Finkelstein, S. & Berndt, E. R. (1996). Depression in the workplace: An economic perspective. In J.P. Feighner & W.F. Boyer (Eds.), *Selective*

serotonin reuptake inhibitors: Advances in basic research and clinical practice. New York: Wiley.

Gullion, C. & Rush, A. J. (1998). Toward a generalizable model of symptoms in major depressive disorder. *Biological Psychiatry, 44*, 959-972.

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

Hiott, A., Grzywacz, J. G., Arcury, T. A., & Quandt, S. A. (2006). Gender differences in anxiety and depression among immigrant Latinos. *Families, Systems, & Health, 24*(2), 137-146.

Joiner, T. E., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Journal of Psychological Assessment, 17*, 267 - 277.

Kessler, R. C., Akiskal, H. S., Ames, M., Birnbaum, H., Greenberg, P., Hirschfeld, R. M., Jin, R., Kerikangas, K. R., Simon, G. E., and Wang, P. S. (2006). Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *American Journal of Psychiatry, 163*, 1561-1568.

Killgore, W. D. S. (1999). Empirically derived factor indices for the Beck Depression Inventory. *Psychological Reports, 84*, 1005-1013.

Klein, D. N. (2008). Classification of depressive disorders in the DSM-V: Proposal for a two-dimension system. *Journal of Abnormal Psychology, 117*, 552-560.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

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

Kroenke, K. Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief

depression severity measure. *Journal of General Internal Medicine, 16*, 606 - 613.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Lowe, B. (2010). The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *Psychiatry and Primary Care, 32*, 345 - 359.

Lewinsohn, P. M., & Essau, C. A. (2002). Depression in adolescents. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression*. New York: Guilford Press.

Mackinger, H. F., & Svaldi, J. J. (2004). Autobiographical memory predicts cognitive but not somatic change in sleep apnea patients vulnerable for affective disorder. *Journal of Affective Disorders, 81*, 17-22.

Muñoz, R. F., Ying, Y., Bernal, G., & Pérez-Stable, E. J. (1995). Prevention of depression with primary care patients: A randomized controlled trial. *American Journal of Community Psychology, 23*, 199-222.

National Institute of Mental Health. (2003). *Breaking ground, breaking through: The strategic plan for mood disorders research.* Washington, DC: National Institutes of Health.

Nezu, A. M., Nezu, C. M., Friedman, J., & Lee, M. (2008). Assessment of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed.). New York: Guilford Press.

Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *Journal of Clinical Psychology, 64*, 83-102.

Osman, A., Downs, W. R., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *Journal of Psychopathology and Behavioral Assessment, 19*, 359-376.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process.

Psychological Science, 8, 162-166.

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.

Rogers, R. (2001). *Handbook of diagnostic and structured interviewing.* New York: Guilford Press.

Rush, A. J., Carmody, T. J., Ibrahim, H. M., Trivedi, M. H., Biggs, M. M., Shores-Wilson, K., et al. (2006). Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services, 57*, 829-837.

Rush, A. J., First, M. B., and Blacker, B. (2008). *Handbook of psychiatric measures* (2nd ed.). Arlington, Va: American Psychiatric Association.

Rush, A. J., Giles, D. E., Schlesser, M. A., & Fulton, C. L. (1986). The Inventory for Depressive Symptomatology (IDS): Preliminary findings. *Psychiatry Research, 18*, 65-87.

Rush, A. J., Gullion, C. M., Basco, M. R., & Jarrett, R. B. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine, 26*, 477-486.

Seligman, M. E. P., Schulman, P., DeRubeis, R. J., & Hollon, S. D. (1999). The prevention of depression and anxiety. *Prevention & Treatment, 2,* no pagination specified.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*, 123-146.

Shean, G., & Baldwin, G. (2008). Sensitivity and specificity of depression questionnaires in a college-age sample. *Journal of Genetic Psychology, 169*, 281-288.

Stewart, J. W., McGrath, P. J., Quitkin, F. M., & Klein, D. F. (2007). Atypical depression: Current status and relevance to melancholia. *Acta Psychiatrica Scandinavica, 115*, 58-71.

Suzuki, T., Yoshino, Y., Tsukamoto, K., & Abe, K. (1995). Concordant factors of depression rating scales and dimensionality of depression. *Biological Psychiatry, 37*, 253-258.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., et al. (2004). The Inventory of Depressive Symptomatology, Clinician rating (IDS-C) and Self-report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician rating (QIDS-C) and Self-report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine, 34*, 73-82.

Vanheule, S., Desmet, M., Verhaeghe, P., & Bogaerts, S. (2007). Alexithymic depression: Evidence for a depression subtype? *Psychotherapy and Psychosomatics, 76*, 315-316.

Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of the Beck Depression Inventory-II: An evaluation. *Assessment, 15*, 177-187.

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2004). Improvement in social-interpersonal functioning after cognitive therapy for recurrent depression. *Psychological Medicine, 34*, 643-658.

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology, 114*, 522-536.

Watson, D., O'Hara, M. W., Simms, L. J., Kotov, R., Chmielewski, M., McDade-Montez, E., et al. (2007). Development and validation of the inventory of depression and anxiety symptoms (IDAS). *Psychological Assessment, 19*, 253-268.

White, H. D. (2009). Scientific communication and literature retrieval. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Yen, S., Robins, C. J., & Lin, N. (2000). A cross-cultural comparison of depressive symptom manifestation: China and the United States. *Journal of Consulting and Clinical Psychology, 68*, 993 - 999.