University of Denver

Digital Commons @ DU

Electronic Theses and Dissertations

Graduate Studies

6-1-2010

Development of a Men's Depression Inventory

Andrew Fields
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Clinical Psychology Commons

DEVELOPMENT OF A MEN'S DEPRESSION INVENTORY

_____

A Dissertation

Presented to the College of Education

University of Denver

_____

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

_____

By

Andrew Fields

June 2010

Advisor: Patrick Sherry, Ph.D.

Abstract

Author: Andrew J. Fields
Title: DEVELOPMENT OF A MEN'S DEPRESSION INVENTORY
Advisor: Patrick Sherry, Ph.D.
Degree Date: June 2010

This paper details the development of a scale to more accurately assess depression in men. It first summarizes the literature on depression, depression assessment, and current research on men and masculinity. It is argued that current conceptualizations of both depression and masculinity influence prevalence studies, which consistently find that men experience depression half as often as women. It is argued that an assessment measure that accounts for masculine variants of depression (substance use, anger, withdrawal, and emotional restriction) may identify more frequent depression in men than previously expected. Next, the paper details the development of a men's depression scale using classical test theory, followed by psychometric analysis of the scale using Rasch modeling and structural equation modeling. Implications on use of the scale and issues related to identifying men's depression are discussed.

TABLE OF CONTENTS

TABLES

APPENDICES

CHAPTER ONE

STUDY OVERVIEW

Introduction

Statement of the Problem

Epidemiological studies consistently find twice as many women to be

depressed as men. Some have posited that these finding indicate a true difference

between the genders in the occurrence of depression (Young, et al., 1990), and

this difference has even been explained by men being buffered from depression

via their coping methods (Nolen-Hoeksema, 1987). However, many patterns of

male distress may be overlooked by simply examining diagnostic criteria for

depression. Mental health trends indicate that there is more going on for men than

what is accounted for by traditional diagnostic methods. For example, men have

significantly higher rates of completed suicides (Klerman, 1997). Men are also

overrepresented in cases of substance abuse and dependence (Hanna & Grant,

1997), and severe personality disorders (Golomb, Fava, Abraham, & Rosenbaum,

1995). Researchers have found support for the notion that the gender differences

in depression can also be explained by men's strict adherence to the male gender

role, especially emotional restriction (Shepard, 2002) and aggression (Cohn &

Zeichner, 2006). Substance use, emotional restriction, and aggression may explain

the difference in depression scores found in men compared to women, and

assessment techniques aimed at measuring these symptoms could increase the detection of depression in men.

Purpose of the Study

The current study seeks to extend our knowledge and understanding of depression in men's lives by validating a measure of depression for men. To obtain further evidence of reliability and validity of a men's depression scale, items designed to assess substance use, anger/hostility/aggression, social withdrawal, and emotional restriction, on a sample of male railroad workers will be examined. Scale scores will be compared to the Beck Depression Inventory II (BDI-II) for evidence of convergent validity. In addition to internal consistency reliability measures, Rasch modeling will be used to obtain information about item fit and difficulty. Confirmatory factor analysis will be conducted to confirm the theoretical underlying factor structure of the scale.

Justification for the Study

The high rates of substance use, suicide, and other forms of psychopathology found in men may be indicative of unmet needs for this population. It is important to gain an understanding of the extent to which depresion is not detected in men. It is likely that many men are suffering silently and, given the high rate of substance abuse and suicide completion among men, a measure that enables more accurate detection of difficulties in men may help improve quality of life and even save the lives of many men. A secondary benefit of this study is further understanding of the presentation of depression in men, and

thus a greater understanding of depression as a phenomenon that can inform

diagnosis and treatment of mood disorders in other populations. Finally, there is a

dearth of scales in Counseling Psychology developed using item response theory

(IRT) despite arguments that it is a viable method of scale analysis providing

information not available using classical test theory (CTT, Fox & Jones, 1998).

This study aims to advocate for the utility of IRT as a method of scale

development and analysis.

<div align="center">Research Hypotheses</div>

Hypothesis 1: The scale will demonstrate acceptable internal consistency,

measured by Cronbach's Alpha.

Hypothesis 2: The scale will demonstrate good convergent validity,

evidenced by a high correlation with scores on the BDI-II.

Hypothesis 3: Confirmatory factor analysis will identify four factors:

Substance Use, Anger/Aggression/Hostility, Withdrawal, and Emotional

Restriction and a second-order factor of depression.

Hypothesis 4: Rasch analysis will show that items vary with increasing

amounts of depression in the participant and will cover the range of levels

of depression in the participants.

Hypothesis 5: Discriminant validity will be shown by demonstrating lower

correlations to scores on a measure of PTSD.

Definition of Major Concepts

For the purpose of this study, the following definitions were used.

Depression: A mood disorder characterized by the presence of a Major Depressive Episode, defined as "a period of at least 2 weeks during which there is either depressed mood or the loss of interest or pleasure in nearly all activities" (APA, 2000, p. 349).

Posttraumatic Stress Disorder (PTSD): An anxiety disorder characterized by characteristic symptoms following exposure to an extreme stressor. Symptoms can include avoidance symptoms, reliving/re-experiencing of the trauma, dissociation or detachment, or persistent increased arousal.

Prevalence: The proportion of occurrences of a disorder in a population. The term may refer to point-prevalence, which is the proportion of people suffering from the disorder at a given time point, or lifetime prevalence, which is the proportion of individuals who will suffer from a disorder at some time in their life.

Alexythymia: Literally, no words for emotions. The inability to describe ones emotional experience.

Gender identity: How one perceives themselves in terms of male or female. This is an independent term from biological sex, gender role, and sexual orientation.

Gender role: Societal expectations for acceptable behavior for men and women.

Classical test theory (CTT): A method of scale development and evaluation stating that an individual's true score on a measure is equal to the observed score plus error, or:

$$X_{true} = X_{observed} + Error$$

Item response theory (IRT): A method of scale evaluation that relates a person's ability to item difficulty. A one-parameter model, known as Rasch modeling, will be used for scale analysis. Rasch modeling uses item difficulty as the parameter, and converts scores to a ratio scale to in effect create a "yardstick" where items can be viewed as regularly increasing intervals of difficulty.

Confirmatory Factor Analysis (CFA): A form of scale analysis using structural equation modeling to verify theoretical models purported to be measured by a

scale. Models show relationships among observations and latent variables, and information about variances, covariances, and model fit are used to examine the hypothesized factor structure of the scale.

Summary

This chapter introduced the problem of detecting male depression using traditional depressive symptom measures. It argued that other factors such as anger/hostility/aggression, substance use, withdrawal, and emotional restriction may account for the under-representation of men in epidemiological studies of depression. It was proposed that a scale assessing these constructs can help us more fully understand men's depression and inform assessment and treatment of men. Chapter Two presents an overview of relevant theoretical and empirical research in the areas of depression as a construct, epidemiological research, the psychology of men, gender-role conflict, and current measures used for assessing depression. Chapter Three describes the methodology for the study and outlines the measures, procedures, and statistical analyses. Chapter Four explains the results of this study. And finally, Chapter Five contains a discussion of the results, study limitations, suggestions for future research, and general conclusions.

CHAPTER 2

REVIEW OF THE LITERATURE

Depression as a Construct

The notion of disordered mood has been identified by clinicians, scholars, and philosophers since ancient times (for an in-depth examination of the history of depression see Jackson, 1986). Freud differentiated between what he described as normal and pathological mood states in his paper "Mourning and Melancholia". The distinction made by Freud and other early theorists mirrors the approach taken by modern day diagnosticians, who attempt to differentiate what is considered "normal" mood state versus clinical depression. The Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR; APA, 2000) defines a Major Depressive Episode as "a period of at least 2 weeks during which there is either depressed mood or the loss of interest or pleasure in nearly all activities" (p. 349). In addition, the individual must also experience four or more additional symptoms related to problems sleeping or eating, psychomotor changes, concentration problems, feelings of hopelessness or guilt, or thoughts of death or suicide. Either depressed mood or loss of interest or pleasure is necessary to make the diagnosis.

A Categorical versus Dimensional Construct

While the construct of depression has undergone refinement in subsequent versions of the DSM, the essential components identified by early clinicians remain largely the same. However, the manner in which we conceptualize depression is continually debated. The DSM classification system is based on endorsement of various symptoms or criteria. If a specific number of criteria are met, the person is given the diagnosis. If the specified number of criteria for Major Depressive Disorder is not met, a diagnosis of Depressive Disorder-Not Otherwise Specified (NOS) can be given. This latter diagnostic option was made available to clinicians due to the understanding that an individual who does not meet all diagnostic criteria may still have clinically significant concerns that must be addressed, and highlights an inherent problem with categorical classification of mood disorders.

Shankman and Klein (2002) highlight two key debates in the area of diagnosis of mood disorders. First is a historical question over a century old: Is reactive depression qualitatively different from endogenous (i.e., biologically based) depression, or merely different ends of a severity continuum? This reiterates the Typological Continuity question of whether subclinical depression is a distinct state, or along the same continuum as clinical depression (Flett, Vredenburg, & Krames, 2004). Both of these questions point to a larger debate in psychopathology: Is a categorical model of mental disorders an accurate reflection

of the manner in which they manifest, or would a dimensional model, where symptom severity runs along a continuum, be more appropriate?

Several researchers have approached this issue from a variety of viewpoints. One of the largest efforts to delineate this issue was conducted by the International College of Neuropsychopharmacology (CINP) President's Workshop and presented in a special issue of the Journal of Affective Disorders (Judd, 1997). Two important finding arose from the workshop. First, the research findings support the notion of depression as a "pleomorphic" disorder characterized by various subtypes that fit along a continuum. Second, subclinical levels of depression present as clinically significant problems, and can be viewed as a disease state on the depression continuum.

The current manner in which mood disorders are categorized has also been criticized statistically. Aggen, Neale, and Kendler (2005) criticize the DSM classification system for collapsing symptom clusters into dichotomous variables using a categorical classification. The authors used one- and two-parameter IRT models to evaluate DSM-III-R diagnostic criteria of Major Depression to determine if the criteria form a linear additive measure. They concluded that a dimensional model that views depression symptoms as scaled risk factors better fit the experience of depression than a list of criteria that count toward a categorical threshold. In addition, they found that diagnostic criteria are less sensitive at low-levels of risk, indicating that they have differential efficacy at different severities of depression.

8

While there is research to support the notion of conceptualizing depression dimensionally (Slade & Andrews, 2005), the CINP findings suggest that the answer is likely more complex. Flett, Vredenburg, and Krames (2004) outline the longstanding debate of continuous versus categorical conceptualizations. The authors caution that it is important for research to acknowledge the issue as a complex one and that it is likely that depression contains both continuous and non-continuous aspects. They explore four aspects of the continuity of depression: Phenomenological Continuity (assessing the quantitative differences between individuals with mild, moderate, and severe forms of depression), Typological Continuity (the existence of subtypes of depression differing qualitatively), Etiological Continuity (the extent of subclinical levels of depression and the associated risk for more severe forms of depression), and Psychometric Continuity, described as "the ability of depression measures to assess the full range of depression scores" (p. 398). This latter aspect of continuity has the most relevance for the current study, which seeks to determine the extent to which a standard depression measure captures the full range of depressive symptoms in men. The authors indicate that IRT methods and CFA models focusing on latent factors are useful in determining the performance of a measure in this regard.

Other Issues in Classification

An important limitation of current classification schemes is also highlighted in comorbidity studies. Zimmerman, Chelminski, and McDermut (2002) found that over two-thirds of patients with Major Depressive Disorder had

another comorbid Axis I disorder, and one-third of their sample had two or more comorbid disorders. Similar results were obtained by de Graaf et al. (2002), who found comorbidity to be more likely for mood disorders than anxiety disorders or substance abuse disorders. Early conceptualizations of depression in the first two editions of the DSM included more anxiety features that were made distinct in later editions (Santor, Gregus, & Welch 2006). Interestingly, current conceptualizations of depression and anxiety that separate these symptom clusters also create significant comorbidity of the two disorder spectrums. Comorbidity appears to be the rule, rather than the exception, thus the current classification system does not accurately reflect how individuals are presenting clinically.

Others have criticized the current classification scheme for its lack of rigor in developing nomenclature for variants of depression (Pincus, Davis, & McQueen, 1999). Winokur (1997) argues for a different approach to conceptualizing depression altogether, stating that classification of depression as a disease is problematic due to its multiple etiologies. He proposed that depression be viewed as a syndrome, similar to a fever, which has multiple etiologies and is present in a multitude of conditions. He describes a classification scheme that accounts for family history of depression and alcoholism which would reflect a more accurate conceptualization of depression. The cumulative research on classification of depressive disorders indicates that categorical classification is problematic on many fronts. Dimensional models show promise, although firm consensus on the actual dimensions remains to be established.

Further complicating the identification and treatment of depression is the discrepant approaches to depression based on either a medical model or developmental model of psychopathology. Some have argued that the successful marketing of antidepressant medications has favored treating depression medically while forgoing other forms of treatment (Gussin & Raskin, 2000). Other models have attempted to integrate the biological and environmental influences on depression. The diathesis-stress model first introduced by Zubin and Spring (1977) describes psychopathology as an interaction between genetic vulnerability and environmental stressors, and has been applied to various aspects of depression (for an example, see Kwon & Laurenceau 2002). Thus an accurate measure of depression should have the ability to capture aspects of etiology and environment in order to most accurately describe phenomenologically what is happening beyond pure symptomology.

The classification debates illustrate an important problem for researchers attempting to study depression from any angle. The myriad studies attempting to fine-tune the construct of depression create various naming and classification schemes that are not in accord with one another. Attempting to integrate the research on depression has highlighted for this author just how unclear the concept of depression truly is. To date, the extant literature lacks clarity. The problem is multiple definitions and there is disagreement on the underlying nature of the construct. These are classic disagreements when attempting to understand the nature of the disease. In essence, the manner in which depression is defined

has a large impact on who will be determined to have it, and the literature shows that current definitions of depression are incomplete at best. While the approaches taken to understand and conceptualize depression each have important contributions to psychopathology research, the various finding have yet to be integrated into a coherent framework and, perhaps most importantly, have not been reflected in diagnostic classification systems. Thus, diagnosing depression remains a subjective process of clinical judgment. Attempts to validate clinical impressions with empirical support are likely to be incomplete at best.

Depression and Culture

In order to better understand depression in men, it is important to examine depression within the context of culture. This section will review research and theory regarding the interaction of culture with assessment, diagnosis, and treatment of depression. It will also examine current views on multicultural competence. Finally, issues related to cross-validation of depression measures are addressed. It will be shown that defining culture and cultural interaction is a complex and ever-changing phenomenon. Further, efforts to truly address culture and mental illness are only just beginning to be realized. Methods of measuring and treating depression are evolving as traditional views are being challenged and improved upon.

Studying Culture

Views on the relationship between depression and culture fall into two general camps. The universal view argues that depression is similar across

cultures and thus can be accounted for by one unitary quantitative measure, while the social constructionist view asserts that depression is culture-bound and measures cannot be generalized across cultures (Redmond, Rooney, & Bishop, 2006; Draguns & Tanaka-Matsumi, 2003). These views have significant implications related to how depression is studied across cultures.

The earliest endeavors to create cross-cultural understanding of mental illness is credited to Kleinman (1977), who argued that cultural variations in mood disorder do exist based on that cultures shaping of normal and deviant behavior. He emphasized the need to examine the social implications of illness. Professionals were encouraged to respect indigenous classifications and conceptualizations for disorders. Further emphasis was placed on understanding the limitations of current diagnostic categories, especially in a cross-cultural setting.

Currently the World Health Oraganization (WHO) studies of depression as part of larger epidemiological research on disease and illness. The 1996 publication of *The World Health Report* found depression to be among the top disorders to cause disability (fifth for women, seventh for men; Desjarlais et al., 1996, cited in Lo´pez & Guarnaccia, 2000). Another important finding was the relationship between mental illness and culture. Factors such as hunger, work conditions, and domestic violence were related to levels of depression in women. The authors thus argue that depression is as much a social illness as it is a mental illness. It seems that even a universalist view that depression has commonalities

across cultures will need to account for socially constructed variations in perceptions of the disorder, pathways to treatment, attitudes toward mental health care, and social factors that create/sustain depression.

Methodological Issues in Cross-Cultural Research

One of the earliest large-scale studies of cultural differences in depression was the US-UK Diagnostic Project (Cooper et al., 1972, cited in Draguns & Tanaka-Matsumi, 2003). In this study it was found that a larger percentage of patients in New York were diagnosed with schizophrenia while patients in London were more likely to be diagnosed with depression. The study found that differences disappeared when using standardized diagnostic criteria (ICD-8). An important finding from this study was that clinicians were likely to contribute to cultural differences, not just patients, although the results suggest that use of standardized diagnostic criteria can alleviate such bias. One limitation of the study is the fact that British and U.S. cultures share significant overlap.

Research on the cultural influences on depression has been marred by methodological issues (Redmond, Rooney, & Bishop, 2006). While the US-UK project indicated that validity was improved with the use of standardized diagnostic criteria, Canino, Lewis-Fernandez, and Bravo (1997) state that such criteria drown out cultural nuances and prevent the formation of relevant hypotheses important to culture. They also argue that the criteria are problematic due to being bound by Euro-American ethnocentrism. One difficulty in accurate assessment of psychopathology in cross-cultural situations involves social distance and empathy. That is, the more unfamiliar a person's culture, the more difficult it is for one to

experience it empathically (Draguns & Tanaka-Matsumi, 2003). Indeed our tendency to group and categorize people may prevent us from acknowledging the substantial overlap among various cultures. This leads to pathologizing cultural variation in mental disorders.

Another struggle with studying depression cross-culturally is difficulty finding common terminology. Some cultures do not have a dictionary-equivalent word for depression, and most cultures vary in terms of the connotative meaning of the word (Tanaka-Matsumi & Marsella, 1976). Using U.S. criteria and definition, we do see some broader cross-cultural support for a similar phenomenon we might call depression. Weissman et al. (1996) was one of the first major attempts at multicultural comparison of mood disorders. They found little variation in bipolar mood disorder, which is not surprising due to the disorder's strong biological etiology. However, they found that sleep difficulties and loss of appetite were consistent depressive symptoms found in ten countries. This evidence of course suggests that depression may exist in similar form across cultures.

Chang et al. (2008) examined the validity of using DSM diagnostic criteria with a Korean population. They found that Koreans met diagnostic criteria for depression about one-fourth as often as people in the U.S., suggesting that the diagnostic threshold may differ despite the actual specific criteria being valid with a Korean population. However, there were some variations. Koreans showed four times the amount of work-related impairment than those in the U.S. The type of

symptoms to first appear differed across cultures as well. Depressed mood was first noticed among those in the U.S. (when the course of the disorder was less severe) with psychomotor retardation or agitation and feelings of worthlessness and guilt occurred when depression was more severe. In Koreans, concentration difficulty and low energy symptoms appeared earlier while psychomotor retardation or agitation and feelings of worthlessness and guilt appeared when depression was more severe. It is important to note that depressed mood is often a required symptom for a DSM diagnosis of depression (that or anhedonia). Thus, the cultural differences in the endorsement of depressed mood may be a factor in the observed prevalence differences.

Measuring Depression Cross-Culturally

There is empirical evidence suggesting that universal depression scales can be used to measure depression across cultures (Arrindell, Steptoe, & Wardle, 2003). However, other research cautions that such measures may still miss important cultural nuances and can never avoid ethnocentric interpretation by the assessor (Redmond, Rooney, & Bishop, 2006). Draguns and Tanaka-Matsumi (2003) examined a large body of research pertaining to studying depression across cultures. Several important conclusions were found to guide cross-cultural measurement. First, the authors note that the increase in efforts to standardize measures for cross-cultural use have allowed researchers to test hypotheses about the variation of psychopathology across cultures. The authors state that "Cultural research on psychopathology starts with the development of scales and other

instruments of assessment. It culminates with their application across and within

cultures" (p. 770). Thus cross-cultural validation is considered to be an integral

part of a scale's validation process. The authors also found that cultural variability

was more pronounced when psychopathology was mild, and cultural difference

dissipated as pathology became greater. Symptom clusters such as guilt and

somatization had the greatest variability across cultures. Finally, a major issue

noted in the extant research is that culture of the clinician (researcher, assessor,

etc.) was often left out, overlooked, or deemed as unimportant. This further

highlights the ethnocentric bias persistent in cultural research. The majority of

research on culture thus focuses on the participant's or client's cultural factors,

which is incomplete. True cultural research must look at the discrepancy between

the observer (researcher, clinician, etc.) and the participant or client. Failure to do

so emphasizes the likelihood to pathologize those that are culturally different and

understates the effects of cultural disparity on assessment and diagnosis.

     Hofstede's (1980, 1991, 2001) cultural measure is one of the more widely

used scales to quantify and describe culture in cross cultural studies. It measures

five dimensions of culture (Power Distance; Uncertainty Avoidance;

Individualism/Collectivism; Masculinity/Femininity; and Confucian Dynamism,

cited in Redmond, Rooney, & Bishop, 2006). A variety of research has examined

variability of depression and culture using Hofstede's scale (Arrindell,

Hatzichristou, Wensink et al., 1997; Diener, Diener, & Diener, 1995). However,

the scale has been criticized as being redundant, overly narrow, and used

inappropriately beyond the scope it was originally intended for (Redmond,

Rooney, & Bishop, 2006). Other researchers have critiqued the methodology used

to extract the factors (Bond, 2002) while others have failed to replicate Hofstede's

factor structure and suggest that the factors are suspect (Spector, Cooper, &

Sparks, 2001). While the scale may have continued use in cross-cultural research,

the data suggest that the scale should be used cautiously. At best, Hofstede's

factors appear to be incomplete or too narrow to examine many cultural nuances

in socially mediated facets of depression and psychopathology.

Perspectives on Cross-Cultural Competence

Definitions and perspectives of what it means to possess cultural

competence vary. Sue (1998) describes cultural competence as the possession of

the knowledge and skills of a particular culture to an extent that allows the

delivery of effective services to such a population. Other theories point at the

ability to move between two cultural perspectives or, more broadly, the ability to

recognize the importance of culture and incorporating culture into assessment and

treatment delivery (see Whaley & Davis, 2007). Sue and Torrino (2005) more

recently described cultural competence as follows:

> Cultural competence is the ability to engage in actions or create
> conditions that maximize the optimal development of the client
> and client systems. Multicultural counseling competence is
> achieved by the counselor's acquisition of awareness, knowledge,
> and skills needed to function effectively in a pluralistic democratic
> society (ability to communicate, interact, negotiate, and intervene
> on behalf of clients from diverse backgrounds) and on an
> organizational/societal level, advocating effectively to develop
> new theories, practices, policies, and organizational structures that
> are more responsive to all groups (p. 8).

This perspective on cultural competence includes the acquisition of cultural knowledge as well as a general approach to thinking about, studying, teaching, and developing policy and practice to reflect such thinking.

Whaley and Davis (2007) reviewed the literature related to the increasing need for multicultural competence among clinicians. They cite research to argue that there is an increased need for cultural competency due to the increasing cultural diversity of the U.S. population. In addition, they note issues related to underutilization and overutilization of mental health services. Underutilization refers to ethnic minority groups using dramatically fewer mental health services than Caucasians, which the authors describe as a case of unmet needs. Overutilization refers to ethnic minorities being given diagnoses of more severe disorders or being in greater distress. Other arguments made in the review state that cultural competence research addresses needs put forth by the American Psychological Association (APA) and the American Counseling Association (ACA) code of ethics. Further, issues of external validity are common with regard to generalizing scientific findings to other cultures. Thus, research on cultural competence is called for on grounds of empirical rigor as well. Finally, the authors argue that cultural competence is an essential component of evidence-based therapy. Traditionally, evidence-based therapy research has failed to extend to ethnic minority groups, and the authors see cultural competence as requisite evidence criteria for such therapies.

Hwang and Wood (2007) argue that the guidelines on cultural competence put forth by the American Psychological Association and the U.S. Department of Health do not provide specific guidance for working with culturally diverse clients. They find fault in the assumption that learning about a particular culture will enable one to work effectively with clients from that culture. Rather, they suggest this reinforces views of the client as "the other" rather than examining the cultural discrepancies between the therapist and the client. They emphasize that the therapy relationship provides an opportunity for acculturating a client to therapy as well as affirming the client when he or she corrects the therapist on cultural issues.

In addition to cultural sensitivity being insufficient to work with culturally diverse clients, developing a complex understanding of every culture is practically challenging, if not impossible. It is unlikely that a clinician will possess a high level of expertise on more than a few cultural groups. Chu (2007) proposes the use of a cultural "approach" as a means for working with various cultural groups. The model is essentially a general therapy framework that includes a deliberate effort to "maximize the cultural exchange" (p. 39). The author describes this exchange as ongoing attempts to challenge assumptions and test cultural hypotheses. That is, there is a continual dialogue between the client and clinician regarding culture, cultural assumptions, and cultural interactions in therapy. Thus cultural interaction is not an examination only of the client's cultural background, but instead focuses on the discrepancy between the clinician's culture and the

client's, a sentiment that is in accord with other research findings discussed above.

Chu (2007) argues that this cultural exchange approach can help a clinician watch for what the author calls Type I and Type II cultural errors. Analogous to hypothesis testing errors, the Type I cultural error is the assumption that a clinical issue is cultural when it is not. For example, a client who consistently arrives late for therapy sessions may be viewed as doing so due to cultural attitudes toward time or punctuality. In actuality, this may be an important clinical issue that is unrelated to culture. The Type II cultural error assumes an issues is not cultural when it is, such labeling a client who is acting out cultural beliefs concerning respect toward authority figures as being passive and deferential (Chu, 2007).

Validating Instruments for Multicultural Use

There are a large number of studies that aim to ascertain cross-cultural support for a variety of depression measures, including the DMI-10 Measure of State Depression (Chan, Parker, Tully, & Eisenbruch, 2007), the Beck Depression Inventory II (BDI-II) and Center for Epidemiologic Studies Depression Scale (CES-D; Kojima et al., 2002), and the Hamilton Rating Scale for Depression (HRSD; Fava, Kellner, Munari, & Pavan, 1982). Many of these studies focus on the examination of EFA factor structure, test-retest reliability, and internal consistency reliability of scales translated from English to another language for use with the latter's native population. While consistency is an important facet of

the psychometrics of these translated scales, reliability is necessary but not sufficient to establish validity.

While many cross-cultural validations of depression inventories have merely reported internal consistency reliability and exploratory factor analysis (EFA) results, researchers are beginning to believe that such methods are not sufficient and arguing for use of confirmatory factor analysis (CFA) to test specific hypotheses about the scale structure (Furukawa, 2005; Oei, Hibberd, & O'Brien, 2005). These research endeavors allow for the examination at the scale (or item) level and allows the research to examine the behavior of the factors predicted to be underlying the scale items. While EFA is more widely used in psychometric research, the method is often extended beyond its intended use erroneously.

Conclusions

This section reviewed current thinking and research regarding cultural interactions with depression and mental illness in general. It was argued that definitions of culture are variable and research findings from the WHO and other cross-cultural studies warrant careful examination of the interaction between depression and culture. While studies suggest that a phenomenon called depression likely exists to some similar extent across cultures, caution must be used in understanding cultural nuances. Of utmost importance is examination of culture as an exchange between two people in a clinical setting, rather than merely looking at a particular client's cultural beliefs and values.

This section also reviewed research on cultural competence. It was argued that competence extends beyond cultural sensitivity and includes a comprehensive approach to clinical work, research, teaching, and policy-making. Research on cross-validation of depression measures was examined, and current findings suggest that many translated measures of depression show promise; however the methods used to establish cross-cultural validity are inadequate. It was also argued that the increased use of confirmatory factor analysis (CFA) is warranted to test specific hypotheses regarding a scale's behavior in other cultural settings or with diverse clients.

Depression and Men

It has been argued that conceptualizations of depression have a large impact on who will be found to have the disorder. This section attempts to illustrate how current definitions of depression may misrepresent men. First, research on the male role is presented, followed by an examination of prevalence research with regard to gender. Finally, current depression measures are evaluated and discussed.

Research on the Male Role

Researchers in the mid-1970s began examining the male side of the negative effects of gender roles first posited by feminist researchers. Pleck's (1981) seminal work on the male gender identity introduced the concept of gender role strain and conflict, where strict adherence to the masculine role leads to interpersonal difficulties. It was argued that the male gender role was actually a

pathological social construct. O'Neil, Helms, Gable, David, & Wrightsman (1986) developed the Gender Role Conflict Scale (GRCS) as a measure of gender role conflict, which occurs when "socialized gender roles have negative consequences on the person or others" (O'Neil, Good, & Holmes, 1995, p. 164). The GRCS is composed of four factors examining problematic adherence to male gender roles: Success, Power, and Competition; Restrictive Emotionality; Restrictive Affectionate Behavior between Men; and Conflict between Work and Family Relations.

The work on the male gender role has culminated in the introduction of a "New Psychology of Men" (Levant, 1996) which integrates the need for addressing the problematic male role in a manner that aids men in taking up the new demands to engage in relationships, raise children, assist in housework, and reduce aggression and violence. In the first known work to integrate the Psychology of Men and attempt to offer a comprehensive examination of men's depression, Cochran and Rabinowitz (2000) offer a thorough review of the research on men's depression and offer a context for understanding the unique issues present in identifying, treating, and research men's depression. The next section will summarize their findings and the work of others in the area of male depression.

Prevalence differences explained

The two-to-one prevalence of depression found in women compared to men has been the source of criticism in recent years (Moller-Leimkuhler, 2002).

Many patterns of male distress may be overlooked simply by examining

diagnostic criteria for depression. Men have a four times higher rate of completed

suicides compared to women (Klerman, 1997). Men are also overrepresented in

cases of substance abuse and dependence (Hanna & Grant, 1997), and severe

personality disorders (Golomb, Fava, Abraham, & Rosenbaum, 1995). While

prevalence studies consistently find women to have higher rates of depression

than men, researchers have posited that these differences appear because gender

role expectations mask the presentation of depressed mood and complicate the

accurate assessment and treatment in men (Cochran & Rabinowitz, 2003). Several

sociocultural factors have been identified to explain the disparity in depression

prevalence. It has been posited that external behaviors are over-represented in

men that may symbolize underlying depressed mood (Brownhill, Wilhelm,

Barclay, & Schmied, 2005). This may indicate a disparity between the reporting

of depression in men and the experiencing of depression in men.

Researchers are also finding support for the notion that the gender

differences in depression can be explained by men's strict adherence to the male

gender role, especially emotional restriction. Shepard (2002) posits that male

social expectations create Gender Role Conflict (GRC), and men with high GRC

are more likely to deny or camouflage depressive symptoms. This would present

as "intolerance of depression", with men displaying more somatic and/or

behavioral symptoms and less affective or cognitive symptoms. The author found

that negative attitude is related to the restrictive emotionality (RE) scale on the

Gender-Role Conflict Scale (GRCS) in college men. This lends support to the notion that restricted emotions are a strong predictor of psychological distress in men. Indeed, the presentation of depression in men may not be encapsulated by DSM criteria.

Pollack (1998) agrees that men will be underrepresented in prevalence studies based on DSM criteria since men are socialized to repress vulnerable experiences. He argues that rates of men's depression in research studies are lower than findings of clinicians, and even clinicians under-diagnose depression in men. He states that three factors contribute to the latter: Men's denial of depression due to socialization, men's emotional restriction makes depression more difficult to detect, and clinicians' own unconscious gender stereotypes.

Assessment of men's depression becomes difficult due to the fact that detection of depression often relies on the reporting of cognitive and affective phenomena. Scheibe, Preuschhof, Cristy, and Magby (2003) state that "depressed men…do not appear to preponderate in any of the symptoms relative to depressed women." (p. 231). Assessment methods sensitive to depression in women may not be appropriate for use with men. Winkler, Pijrek, & Kasper (2005) surveyed depressed male and female patients, and found that men were more likely than women to have experienced irritability or to overreact during their last depressive episode. Men also demonstrated lower impulse control and higher substance use than women in this study. Brownhill, Wilhelm, Barclay, and Schmied (2005) hypothesized that a depression scale emphasizing men's symptomatic expression

26

of depression in addition to DSM-IV criteria would yield higher rates of depression in men. Pollack (1998) proposed a new subtype of depression called Major Depressive Disorder-Male Type, including symptoms related to increased withdrawal, anger, denial of pain, substance use, and denial of sadness, among others.

Blair-West, Cantor, Mellsop, & Eyeson-Annan (1999) found that substance abuse and being male are strong predictors of suicide risk. Further, they suggested that the male threshold for depression be lowered, as a diagnosis of depression has limited utility for predicting suicide risk. To further highlight the role of substance abuse in the epidemiological discrepencies, Cochran and Rabinowitz (2000) reviewed literature illustrating the sociocultural impact on gender differences in depression. They cite a study of mental disorders in an Amish community (Egeland & Hostetter, 1983) that found nearly equivalent rates of depression in men and women. This disparity from the general U.S. population was explained by the near-absence of alcohol abuse reported in this population.

The notion of a "male depressive syndrome" has been put forth by researchers studying depression and suicide on the Swedish Island of Gotland (Rutz et al., 1995, 1999). These researchers developed The Gotland Male Depression Scale to aid in assessing this syndrome. Male Depressive Syndrome includes lowered stress tolerance, impulsive behavior, and substance abuse or their equivalents (e.g., workaholism, Moller-Leimkuhler, Bottlender, Straub, & Rutz, 2004). Thus, the symptoms men present with according to the authors are

irritability, anger, hostile or aggressive behavior, and alexithymia, symptoms not assessed in standard depression measures.

Research on the scale has been limited, but supports the notion of a male depressive syndrome. The Gotland Male Depression Scale was used to assess depression in a sample of males seeking help for alcohol use problems (Zierau, Bille, Rutz, & Bech, 2002). The scale was found to have good internal consistency and was positively correlated with an established depression scale. Moller-Leimkuhler et al. (2004) found that these male depression symptoms were not significantly more frequent in men than women. However, exploratory factor analysis showed that, while "typical" depressive symptoms loaded on the first two larger factors for women, the first two factors for men consisted mostly of aggressive, abusive, and antisocial behavior as well as irritability.

The cumulative research on men and depression yield some important conclusions. It is likely that the gender differences in depression prevalence are a consequence of current conceptualizations of depression, rather than true differences in the occurrence of depression. It has been shown above that the manner in which we operationalize the construct of depression may have a large impact on who meets criteria. Symptoms not articulated in the DSM-IV-TR that represent the manifestation of depression in men include substance use, anger, withdrawal, and emotional restriction. The Gotland Men's Depression Scale does assess some of these constructs, but does not directly evaluate emotional restriction, and initial analysis indicates that the factors are not clearly defined.

Consequently, it may be that a measure assessing these symptom clusters above and beyond traditional symptoms will provide a more complete picture of men's depression and would likely explain the gender discrepancies found. Alternatively, these symptoms could represent a new dimension of depression, which, together with traditional symptoms, more completely encompass the range of phenomenological representations of depression. The next section will discuss several depression measures commonly used in research and clinical settings.

Current Measures of Depression

The difficulties with conceptualizing depression have important ramifications for developing assessments. Snaith (1993) examined differences among several popular depression measures, including the Hamilton Rating Scale for Depression (HRSD), the Beck Depression Inventory (BDI) and the Center for Epidemiological Studies Depression Inventory (CES-D). He noted the faulty assumption made by researchers that all scales are based on the same construct, and indicated that the measures differ on the types of symptoms they emphasize (e.g., the BDI contains more cognitive items).

Santor, Gregus, and Welch (2006) sought to elaborate on the work of Snaith (1993) and examine variations in depression inventories from a general perspective. They recognized that the large number of depression scales based on various theoretical frameworks would make it difficult for researchers and clinicians to select appropriate depression measures. They examined the over 280 depression measures available to date. Measures were examined in terms of item

29

characteristics, changes over time, number of scales, and frequency of use. Their findings yield two important conclusions. First, some symptoms included on many depression measures do not coincide well with symptoms thought of as "core" in diagnosis (such as worthlessness). Second, the sheer number of existing scales makes it difficult for new measures to be introduced and accepted by the psychological community, and many of the most commonly used measures were developed over 20 years ago.

Both Snaith (1993) and Santor, Gregus, and Welch (2006) identified the BDI, HRSD, and CES-D as the most commonly used depression scales. The latter authors found the BDI to be the most representative of depression measures in general. The HRSD and CES-D, which are widely used in prevalence and outcome studies, actually differ from general depression conceptualizations. Both studies indicated that the HRSD, often considered the "gold standard" for new scale validation, contains a disproportionately high number of somatic symptoms. The CES-D was found to be even more problematic in that it contains several items that are not unique to depression, such as perceptions of others. The conclusions made by Santor, Gregus, and Welch (2006) have profound ramifications for depression research:

> Two of the primary measures (of depression), namely the HRSD
> and the CES-D, on which much of what we know about basic
> science and treatment outcome studies depends, are not
> representative of how measures of depression have been
> operationalized (p. 151).

Summary

      This chapter reviewed the literature pertaining to conceptualizing depression, and presented an introduction into research on male gender roles. Prevalence differences and current measures of depression were also discussed. The design for testing this hypothesis is discussed in Chapter Three. Chapters Four and Five contain an examination of the results and the general conclusions derived from the study.

CHAPTER 3

METHODOLOGY

The current study seeks to develop a scale for assessing depression more

accurately in men, as well as obtain preliminary psychometric data on the scale.

This chapter will first describe the development and initial evaluation of the

Men's Depression Inventory (MDI). The items for the scale were developed

following tenets of classical test theory (CTT). Development and evaluation of the

initial item pool is described in Fields, Sherry, and Green (2007). While early

versions of the scale contained items assessing traditional depressive symptoms,

for this study, the item pool was revised to create a scale with items more

differentiated from standard depression measures (i.e., no items assessing

traditional depressive symptoms). The scale was designed to specifically assess

four factors: Substance Use, Anger/Aggression/Hostility, Withdrawal, and

Restricted Emotions. Items for the first three factors were newly created for this

study. For the Restricted Emotions factor, items were included from the Gender

Role Conflict Scale's (GRCS) Restrictive Emotionality Subscale (RE), as this

scale has been found to be highly correlated with depression in men (Shepard,

2002). The item pool was submitted for expert review to Patrick Sherry, Ph.D.,

James O'Neil, Ph.D., and Denny Holland, Ph.D. Each of the reviewers has

experience working with depressed men and is interested in furthering the study

of men's depression. Each item was assessed based on wording, relevance to construct, and likelihood to be endorsed if true.

Participants

Participants consisted of male railroad workers, including trainmen and mechanical workers. Rule of thumb estimates for CFA indicate having fifty observations per latent factor, leading to a minimum target sample size of 200. The target sample size of 300 was sought to assure sufficient observations. As shown below this sample size was greatly exceeded. Response rates for voluntary surveys are typically around 50-60%, however previous studies with this population have yielded response rates as high as 95%. A 75% response rate was to be considered acceptable for this study. Individuals who refused the survey were anonymously tallied in order to estimate response rate. Women who took part in the larger study were excluded from the data analysis. Fields and Sherry (2008) have found the proportion individuals reporting moderate to severe depression, according to previously published cutoffs, to be nearly 40%, double the proportion found in the general population, although the full range of depression scores was expected to be represented in this group.

Procedures

The scale was administered as part of a larger study on fatigue, stress, and wellness in transportation workers. Prior to data collection, approval was granted by the Institutional Review Boards (IRB) of the University of Denver for the larger study being conducted. An addendum to the initial IRB proposal was

submitted explaining the purpose and content of the new measures added for this study. Participants were approached by research assistants at the start of their shift and asked to fill out a survey on workplace satisfaction, stress, fatigue, and health and wellness. It was explained that the survey could be taken anonymously, but participants have the option to include their name and email address to be contacted with personalized results.

Due to the high rates of depression previously found in this population, individuals who rated several items in the severe or extreme range or who showed especially high depression scores were briefly interviewed upon completion of the survey. This was done as a safety and health check to insure that the individual was not distressed by the survey questions, not currently experiencing clinically high ranges of depression, and fully aware of the resources available to them. The interview also served to gain further validity for the larger study and to provide referral information for individuals who may be severely depressed and/or suicidal. No individuals were deemed in need of treatment following completion of the survey and the brief interview.

Measures

Men's Depression Inventory. The scale developed for this study is a 29-item self-report rating scale. For each item, respondents are asked to rate the extent to which they agree with the statements based on how they have felt in the past two weeks. Response choices range from 1=Strongly Disagree to 6=Strongly Agree. This was chosen to align with the Restricted Emotions subscale of the

GRCS which uses this rating scale. The scale was constructed to measure four constructs: 1) Anger, aggression, and hostility 2) Substance use 3) Social withdrawal and 4) Restricted emotions. The full scale is presented in Appendix A.

Beck Depression Inventory. The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item self-report rating scale. Each item represents a symptom of depression (e.g., Sleep Difficulty) and asks respondents to select from several options the choice that best describes them. Response choices are labeled from 0 to 3, with higher numbers representing more severe depressive symptoms. Scores of 0-13 indicate minimal depression, 14-19 indicate mild depression, 20-28 indicate moderate depression, and 29-63 indicate severe depression. Higher total scores indicate more severe depressive symptoms.

The BDI is one of the most widely used measures of depression. Beck (1988) presented a review of the psychometric properties of the BDI and its revision, the BDI-II. This review found the scale to have high internal consistency (.86 and .81 in clinical and non-clinical samples, respectively) and concurrent validity with four well-researched measures of depression, including the HRSD and the Depression scale of the Minnesota Multiphasic Personality Inventory (MMPI). In addition to differentiating clinical and non-clinical cases, the BDI was found to discriminate depression from anxiety disorders and identify several subtypes of depression, depending on the sample used. Lasa et al. (2000) found the BDI to have high sensitivity (100%) and specificity (99%), and found no statistically significant differences based on sex or age.

35

PTSD Checklist Civilian Version. The PTSD Checklist Civilian Version

(PCL-C; Blanchard, Jones-Alexander, & Buckley, 1996) is a 17-item rating scale

measure assessing symptoms of PTSD. Response choices range from 1 (Not at

all) to 5 (Extremely), and participants are asked to indicate which response

describes how much they have been bothered by each symptoms (e.g., Repeated,

disturbing dreams of a stressful experience from the past?). A cutoff score of 50

suggests a diagnosis of PTSD. Using this cutoff criterion, Adnrykowski, Cordova,

Studts, and Miller (1998) found the PCL-C to have a sensitivity of 60% and

specificity of 99%. The PCL-C was found to have high internal consistency

reliability (alpha = .97) and convergent validity with other measures of PTSD

(Weathers et al., 1993, cited in Andrykowski et al., 1998),

Data Analysis

Classical Test Theory

Classical test theory (CTT) has traditionally been used to assess

psychometric integrity of new measures. CTT states that an individual's observed

score on a measure is equal to their true score plus error, or:

$$X_{observed} = X_{true} + Error$$

Reliability evidence for the scale will be assessed using internal consistency

reliability (i.e., Cronbach's alpha). Convergent validity evidence will be obtained

by correlating scores on the new scale to scores on the BDI-II. Discriminant

validity evidence will be obtained by examining the correlation of the scale to

scores on the PCL-C.

Rasch Modeling

Item-response theory is a method of relating a person's ability to item difficulty. A one-parameter model, known as Rasch modeling, will be used for scale analysis. Rasch modeling uses item difficulty as the parameter, and converts scores to a ratio scale to in effect create a "yardstick" where items can be viewed as regularly increasing intervals of difficulty. The concepts of item difficulty and person ability seem intuitive on, say, a math test, but they may seem unclear for a construct such as depression. However, the ideas of difficulty and ability remain the same. For both types of scales, person ability describes the level of the latent variable (e.g., math skill or level of depression). Item difficulty corresponds to the amount latent variable needed to endorse the item in the scored direction. Thus, a more difficult depression item would require more depression to endorse in the scored direction (i.e., the direction that indicates depression), just as a more difficult math question would require greater math skill to endorse in the scored direction (i.e., the correct answer).

Although IRT has been identified as a viable alternative to CTT for over fifty years, only in recent years has the value of IRT been widely recognized by researchers in Counseling Psychology, especially for use in smaller-scale assessments of personality and attitudes (Harvey & Hammer, 1999). Fox and Jones (1998) present a discussion on the uses of IRT in Counseling Psychology research. The authors state that IRT enables researchers to test a scale for unidimensionality (an explicit assumption of the IRT model), create an interval

scale of item difficulty, and examine person ability in relation to the latent variable. Thus, IRT can, above and beyond CTT, indicate that a scale is in fact measuring a unitary construct. The interval scale of item difficulty allows one to understand the difference in difficulties among items, and the ability to examine person data is a facet virtually untouched in CTT.

IRT accounts for some of the limitations of CTT. Because the calibrations used in IRT allow the measure to be generalized across samples (due to independence of the items from the sample) the difficulty of the item and person ability are not confounded as they are in CTT. In addition, while CTT provides a standard error estimate for the entire sample, IRT provides a standard error estimate for each person and item (Fox & Jones, 1998). The analyses possible under IRT clearly show utility in scale development. While CTT can serve as more of a "blunt" examination of a scale's psychometric viability, IRT allows detailed examination of individual item functioning, person functioning on the scale, and relation of items to one another.

Harvey and Hammer (1999) reinforce these advantages of using IRT over CTT. The level of analysis is at the item level (as opposed to scale level), allowing for examination of the robustness of each individual item. Further, unlike CTT models, IRT does not need to "assume that the test is equally precise across the full range of possible test scores" (p. 365). This is because the IRT model relies on a continuous function to give "information" (comparable to reliability in CTT). While both IRT and CTT allow for methods to examine the

effects of deleting individual items on the measure, the method used in IRT

(based on the test information function and test standard error) is more specific

and sensitive compared to the global alpha or standard error of measurement used

in CTT. IRT also provides information on person reliability, such as whether

respondents make coherent use of the rating scale and reliability of the rankings

of persons on the specific trait (Hatcher & Gillaspy, 2006).

Another comparative strength of IRT is its ability to provide additional

validity evidence above and beyond what can be done in CTT. While concurrent

validity evidence can be obtained using both IRT and CTT, validity evidence that

is unique to IRT involves the ordering of item difficulty (Fox & Jones, 1998).

Essentially, if item difficulty is arranged consistently with expectations of the

theory, then evidence for validity is present. This relationship can be found by

simply correlating the item logit positions with item rankings provided by

independent experts. A second source of validity evidence unique to IRT has to

do with fit statistics. Fit statistics are provided for both person and item, and

indicate whether responses occur as would be expected. That is, for each item and

each person (to each item), IRT provides information about departures from a

unidimensional construct.

The current study will utilize many of the above analyses using Winsteps

(Linacre, 2007). Since the hypothesized model contains four subscales, Rasch

analysis will examine each subscale separately. Item difficulty and person ability

maps will be generated to examine the item scaling and relative difficulty to the

sample's ability. Items at redundant difficulty levels may lead to deletion of some items (assuming the integrity of the construct is maintained). These maps also show gaps in difficulty levels where items may need to be added to fully tap the construct. Individual item fit statistics will provide information on the performance of each item and further suggest items for deletion. Finally, category probability tables provide information on the functioning of the rating scale. In essence, one can gain information on which response choices provide significant information, and the table indicates the appropriate number of useful response choices. Fit statistics used in the Rasch model are described in Chapter 4, and fit values between .5 and 1.5 will be considered acceptable for this study.

Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a type of structural equation modeling used to test complex measurement models (Kline, 2005). Diagrams are used to display the observed variables, such as scale items, and the latent factors hypothesized to cause these variables. Paths are designated by the researcher to describe how the latent and observed constructs relate to one another theoretically. The theory can then be tested by examining the strength of each path as well as various indices of model fit. Model fit indices are used to determine how well the data fit the hypothesized model.

Figure 1 shows the model to be tested for the current study. The diagram consists of 29 observed variables (presented as rectangles) that relate to the 29 items of the scale. The four latent constructs of substance use, anger/aggression,

40

withdrawal, and emotional restriction are presented as ovals. The second order

latent construct of depression is also presented, and the arrows indicate prediction

pathways among the variables (straight arrows denote causal pathways while

curved double-arrows denote covariance between variables).

There are several indices of model fit used in structural equation models.

The most basic is chi-square, which tests the null hypothesis that the model

perfectly fits the data. Failure to reject the null is desirable here, however, the test

is highly affected by sample size and the null hypothesis of perfect model fit is

unlikely to actually exist. A commonly used fit statistic that accounts for sample

size and does not assume perfect model fit is the Root Mean Square Error of

Approximation (RMSEA). For this study, an RMSEA near .05 will be considered

acceptable. The Comparative Fit Index (CFI) describes the fit improvement of the

model compared to an independence model where all variables are unrelated. A

CFI value greater than .9 is desirable. The Akaike Information Criterion (AIC) is

a goodness-of-fit measure which adjusts model chi-square to penalize for model

complexity. While there are not cutoff values for the AIC, lower scores indicate

better model fit and can be used to compare models. Since the various fit indices

examine different aspects of model fit, all three indices will be reported and the

meanings will be interpreted together.

Summary

This chapter presented the methodology to be used in the study, including

participants, study procedures, and analyses to be conducted. An overview of

41

Classical Test theory and Item-Response Theory explored the types of

information that can be obtained from each. An introduction to Confirmatory

factor analysis was presented, including presentation of the theoretical model and

various types of fit indices. The next chapter describes the results of the study,

including data analysis and relevant findings.

CHAPTER 4

RESULTS

Considerations Made Prior to Data Analysis

Two considerations were addressed prior to data analysis. The sample

contained three female participants from the larger study, whose cases were

removed for the following analyses. In addition, items 9, 10, 12, and 24 are keyed

in the reverse direction. Thus response choice codings were reversed.

Participant Demographics

The sample consisted of 423 male railroad workers. The sample was

73.2% Caucasian, 9.8% Hispanic, 3.2% African American, 2.9% Asian, 1.4%

Native American, 3.2% "other," and 6.3% no response. The average age of the

sample was 44, with ages ranging from 19-66. The men in the study reported an

average of 13.5 years of education (including high school) with a range of 7-18

years. The majority of the men were married (61.9%) with 19.3% reporting never

being married and 7.3% reporting being divorced (11.5% had no response).

Testing of Main Hypotheses

This section will test the four main hypotheses of the study. In accord with

tenets of classical test theory, it was predicted that the scale would have good

internal consistency reliability and convergent validity with the BDI-II. It was

also predicted that the model shown in Figure 1 would demonstrate good model

fit using confirmatory factor analysis. Use of Rasch modeling predicted that the items would be well targeted and show sufficient spread across the range of depression for the sample. Finally, discriminant validity with a measure of PTSD was predicted.

Internal Consistency Reliability of the Scale

Internal consistency reliability was ascertained both for the total scale and for the subscales individually. The total scale showed good internal consistency reliability (Cronbach's Alpha=.882) and no significant improvements in reliability could be made by deleting an item. The Anger subscale had a Cronbach's Alpha of .877. The Withdrawal subscale showed considerably less internal consistency (Crobach's Alpha=.640). The Substance Use subscale had a Cronbach's Alpha of .830 and the Restricted Emotions subscale had a Cronbach's Alpha of .753. The Anger and Withdrawal subscales did not show a significant increase in reliability if items were deleted. However, the Substance Use subscale's reliability would increase to a Cronbach's Alpha of .871 with the removal of item 18 ("I have had an alcohol or substance use problem in the past"). The Restricted Emotions subscale's reliability would increase to a Cronbach's Alpha of .819 with the removal of item 24 ("I find it easy to put my feelings into words").

Convergent Validity

To obtain evidence of convergent validity, the MDI total scores were correlated with respondent's total scores on the BDI-II. A Pearson product-moment correlation of .521 (p<.001) was found, indicating a moderate correlation

between the scales. Correlations between the BDI-II and the four subscales were also obtained. Similar relationships were found for the Anger (r=.523, p<.001), Withdrawal (r=.412, p<.001), and Restricted Emotions (r=.469, p<.001) scales. The Substance Use subscale showed less of a relationship (r=.296, p<.001) compared to the other subscales. The moderate correlations are desirable given the goals of the scale and will be discussed in Chapter 4. However, the correlation was much lower than desirable for the Substance Use subscale. Thus, Hypothesis 1 is partially supported.

Confirmatory Factor Analysis

The proposed model was tested using CFA and is shown in Figure 1. The first step in testing the model is to examine model fit. Because each fit statistic offers somewhat different information, several indices are reported. The comparative fit index (CFI) compares the existing model fit with a null model which assumes the latent variables in the model are uncorrelated, called an independence model. Values range from zero to one, with higher numbers indicating better fit. The root mean square error of approximation (RMSEA) measures discrepancy per degree of freedom, and values at or below .05 typically indicate good model fit. The Akaike Information Criterion (AIC) is a goodness-of-fit measure which adjusts model chi-square to penalize for model complexity. AIC is used to compare models and is not interpreted for a single model, with higher numbers indicating and improvement in fit. Due to the poor fit of the

initially devised model, a revised model was developed by removing four non-significant items from the model.

Table 1 displays the fit statistics for each model iteration. The first model gave a CFI of 0.829, an RMSEA of 0.073, and an AIC of 1422.578. These values do not meet acceptable criterion for model fit, and examination of the model paths revealed four items with insignificant paths (items 9, 10, 12, and 24). For the second iteration of the model, these items were removed from the model, yielding somewhat better model fit. The revised model (Figure 2) gave a CFI of 0.895 an RMSEA of 0.64, and an AIC of 921.865. Unfortunately, these statistics are still lower than acceptable values.

Model 2 was examined to identify changes that would lead to acceptable model fit. A number of theoretically-based changes were identified. Items 8-15 contain wording they may also be endorsed as true by men exhibiting emotional restriction (e.g., "When I'm upset I just want to be left alone"). Thus a revised model (Model 3) with these items crossloading on both the withdrawal factor and the emotional restriction factor was used (see Figure 3). Model 3 was examined and did produce improved model fit. The values were still below acceptable model fit, with CFI= .912, RMSEA= .059, and AIC=846.799.

A second theoretical variation was created for Model 4, where items 8 and 11 were assigned to the emotional restriction subscale alone rather than crossloading them on both the withdrawal and emotional restriction subscales (see Figure 4). Model 3 was examined and did not significantly improve on the fit

compared to Model 3 with CFI= .911, RMSEA=.059, and AIC=847.447. Thus, it was determined that the separate subscales of withdrawal and emotional restriction were not distinct enough to be supported by the model. A fifth model was created that simply collapsed emotional restriction and withdrawal into one subscale (see Figure 5). This model did yield improved fit over previous iterations, with CFI= ..92, RMSEA=.060, and AIC=675.251. The significant drop in the value of the AIC from Model 4 to Model 5 likely indicates that previous models with four subscales were too complex, and Model 5 with three subscales was more parsimonious.

Model 1 did not show adequate fit to the data, thus Hypothesis 2 was not supported. After the removal of four non-significant items, and the addition of crossloadings of the withdrawal items onto the restricted emotions factor, model fit was still not significantly improved. Upon collapsing two of the subscales, model fit was improved. Thus, post-hoc analysis did produce a model that adequately fit the data.

Rasch Modeling Results

The Rasch model assumes that a scale is measuring a unitary construct, thus each subscale was analyzed separately for fit and targeting. Fit refers to how well the data fit the prescriptions of the Rasch model. Targeting refers to the relative ability of the person parameter compared to the difficulty of the item parameter. In other words, were the items of sufficient difficulty to cover the

47

spread of ability (read as amount of the latent trait; in this case, depression) of the participants?

Fit

Fit is expressed using two transformations of chi-square statistics. Infit is an information-weighted sum that accounts for the distance between the person location and item location, and outfit is an unweighted measure. Both values are reported as mean squares (MNSQ) with an expected value of +1, indicating perfect fit, and possible values ranging from zero to positive infinity. A MNSQ of less than +1 indicates less variability than expected, while a MNSQ greater than +1 indicates more variability than expected by the model.

Infit and outfit share the same distribution, but infit leaves the differential effects of weighting in place and thus is less sensitive than outfit to extreme responses. It has been argued that MNSQ fit values between .5 and 1.5 are acceptable, but cutoff values should account some flexibility to allow for researcher judgment (Linacre, 2007).

Subscale Fit and Separation

Rasch analysis was conducted using the original 29 items. While this model did not fit the data well according to the CFA analysis, this was done to obtain detailed information on every item to inform scale revisions. Table 2 shows the fit and separation values for each of the four subscales. Rasch modeling provides information on fit and separation for both persons and items. Person separation reliability is also reported.

48

Person fit. Person fit describes how well the sample behaved in a manner predicted by the model. That is, those having more depression would be expected to have higher scores, while those with less depression would be expected to have lower scores. Person fit for each of the subscales was in expected ranges. The anger subscale MNSQ outfit value was 1.06, and an MNSQ infit of 1.02. The substance use subscale MNSQ outfit value was 1.01, and an MNSQ infit of 1.10. The withdrawal subscale MNSQ outfit value was 1.01, and an MNSQ infit of 1.04. The restricted emotions subscale MNSQ outfit value was 1.04, and an MNSQ infit of 1.08.

Person separation. Person separation shows the spread of persons across the item difficulty levels. Typically separation values greater than +2 are desirable, with values below +1 being unacceptable. Person separation values were marginal for the anger (1.54), withdrawal (1.17), and restricted emotions (1.31) scales. Fit was was too low for the substance use scale (0.81). These findings suggest that persons were too close together along the continuum of item difficulty.

Person separation reliability. Person separation reliability is conceptually similar to Cronbach's alpha, and indicates the extent to which the items order the persons. The anger and restricted emotions subscales showed moderately high person reliability at 0.7 and 0.63, respectively. The withdrawal scale was lower, but still moderately high at 0.58. The substance use subscale showed lower person separation reliability at 0.40.

Item Fit and Separation

Item fit. Item fit describes how well the behavior of the items matched the predictions of the model. Item MNSQ outfit and infit values were within acceptable ranges for all four scales. The anger subscale had an MNSQ outfit of 1.06, with an MNSQ infit of 1.02. The substance use subscale had an MNSQ outfit of 1.01, with an MNSQ infit of 1.02. The withdrawal subscale had an MNSQ outfit of 1.01, with an MNSQ infit of 1.00. Finally, the restricted emotions subscale had an MNSQ outfit of 1.04, with an MNSQ infit of 1.02.

Item separation. Item separation was significantly better than person separation. Separation was strongest for the anger (6.77), withdrawal (7.07), and restricted emotions (8.09) subscales, with the substance use subscale showing lower, but acceptable separation at 2.94. These values indicated that the items covered the range of ability of the persons in the sample.

Individual Item Fit and Logit Position

Table 3 shows the fit and logit position for the anger subscale. Items 2 and 7 show logit difficulties that are higher than the ability of the sample, indicating the items are too difficult. Table 4 shows the fit and logit positions for the withdrawal subscale. Item 14 shows high logit difficulty, indicating that it is too difficult for the sample. Table 5 shows the fit and logit positions for the substance use subscale. Logit difficulty was acceptable for the items in this scale, although the items proved too difficult for the sample. This is discussed in Chapter 5. Table 6 shows the fit and logit postions for the restricted emotions subscale. While item

50

26 has high logit difficulty and item 24 has low logit difficulty, this represented the range of logit ability for the sample. Overall these logit positions are best used to identify how difficult the items are in relation to one another, which can inform scale revisions along with item targeting.

Targeting

One unique strength of Rasch modeling is the ability to examine how well targeted the items and persons are. Targeting refers to the extent to which the item logit difficulty matches with the person ability. This provides information about the extent to which the items cover the breadth of ability in the sample, and also if the range of ability in the sample covers the range of difficulty in the items. This information aids in revising a scale for further use by identifying gaps at various logit positions. In the case of a depression scale, targeting allows a determination of how the wording of items can be adjusted to better assess the response style of a new population.

Targeting is difficult to identify with classical test theory since items and persons are reported using different metrics. Items are measured as means and persons are measured with raw total scores. The Rasch model places person and items using the same metric (logit position) to identify how ability and difficulty relate directly.

Anger scale. Figure 6 shows a person-item map for items 1-7, the anger subscale. "M" denotes the mean response choice location for the sample, which is around choice 1 or 2 for most items. A well-targeted scale would be expected to

51

show an average response choice somewhere between 3 and 4. The figure shows

that most respondents did not strongly endorse this item as predicted by the

model. The items in the scale are much too difficult, with item 2 ("At times I get

so angry I am violent toward others") being the most difficult.

Restricted emotions scale. Figure 7 shows the person-item map for the

restricted emotions subscale. The mean response choice is close to the middle

range for this subscale, indicating moderately good targeting. Many items are still

somewhat difficult for the sample.

Substance use scale. Figure 8 shows the person-item map for the

substance use subscale. The mean response choice is near 1 for this scale,

indicating that the item difficulty was much too high and the scale is poorly

targeted.

Withdrawal scale. Figure 9 shows the person-item map for the withdrawal

subscale. The mean response choice is near the middle range for most items in

this scale, indicating good targeting. In particular, items 9 ("I have people I can

rely on when I am having a hard time"), 11 ("When I'm upset I just want to be

left alone"), and 12 ("I enjoy the support I receive from others") are well-targeted.

Conclusions for Hypothesis 3

Overall, each subscale showed adequate person and item MNSQ fit,

indicating that the items and persons behaved as predicted by the Rasch model.

This was also true for the individual items. In terms of targeting, overall the scale

items are too difficult for the sample. This is particularly the case for the anger

and substance use subscales. The restricted emotions and withdrawal subscales appear well targeted, with the items being only slightly too difficult. Thus, Hypothesis 3 is partially supported.

Discriminant Validity

Evidence for discriminant validity was obtained by correlating MDI scores with the PCL-C. A Pearson correlation of .333 ($p<.001$) was found, indicating a moderately low correlation between the scales. Similar relationships were found for the Anger ($r=.309$, $p<.001$), Withdrawal ($r=.233$, $p<.001$), Substance Use ($r=.291$, $p<.001$), and Restricted Emotions ($r=.299$, $p<.001$) subscales. As expected, the correlations between the MDI and the PCL-C are generally lower than the correlations between the MDI and the BDI-II, thus Hypothesis 4 was supported.

Summary

This chapter detailed the results of the analysis of the Men's Depression Inventory. Considerations made prior to data analysis were addressed, followed by details on the sample demographics. It was found that main Hypotheses 1 and 4 were supported. Hypothesis 2 was rejected for the original model, but supported for a revised model. Hypothesis 3 is partially supported, and a more complete discussion of this is found in Chapter 5. Hypothesis 4 was supported. The next chapter will include a discussion of the results, integration of the findings, and how this affects implications for future revisions of the scale.

CHAPTER 5

DISCUSSION

This study used classical test theory, confirmatory factor analysis, and

Rasch modeling to obtain psychometric data on the Men's Depression Inventory.

The results for each of the four main hypotheses are presented in Chapter 4. This

chapter will discuss the findings as they relate to the four hypotheses. In addition,

it will identify recommendations for amending the scale for future use based on

the findings of this study. Finally, theoretical implications for the study of

depression in men will be discussed.

Hypothesis 1

Hypothesis 1 predicted that the scale would demonstrate acceptable

internal consistency, measured by Cronbach's Alpha. Internal consistency for

each subscale and the whole scale was found to be strong. This suggests that

subscales were measuring a unitary construct, and the notion of four subscales

with items that relate to one another was also supported.

Hypothesis 2

Hypothesis 2 predicted that the scale would have convergent validity

evidenced by a high correlation with scores on the BDI-II. The correlation of the

scale to scores on the BDI-II was moderate. While this was not predicted by

Hypothesis 1, theoretically this finding is appropriate. The scale was designed to

measure a construct of men's depression not tapped by traditional depression measures. It is reasonable that such a scale would share some variance with standard depression measures, but also tap into variance not shared by such a measure. The moderate correlation suggests that the Men's Depression Inventory is measuring a construct similar to that of the BDI-II, but also addressing some aspects that are discreet. Given the intentions of the scale, such a correlation is desirable.

Hypothesis 3

Hypothesis 3 predicted that a confirmatory factor analysis would identify four factors: Substance Use, Anger, Withdrawal, and Emotional Restriction and a second-order factor of depression. The original model shown in Figure 1 did not demonstrate adequate model fit. Elimination of four non-significant items in the scale improved model fit, but fit was still insufficient. Since the items in the withdrawal scale were worded in such a way that they could be endorsed by men experiencing emotional restriction, these items were cross-loaded so they loaded on both the withdrawal factor and the restricted emotions factor. This model (Model 3) shown in Figure 3, demonstrated improved model fit, but fit was still inadequate. Model 4 attempted a variation of Model 3 with no improvement in fit. Model 5 collapsed the withdrawal and emotional restriction subscales into a single factor, which yielded adequate model fit. Thus Hypothesis 3 was not supported, but post-hoc analysis derived a model that fit the data.

Hypothesis 4

Hypothesis 4 predicted that Rasch analysis would show that the items vary with increasing amounts of depression in the participants and will cover the range of levels of depression in the participants. This was tested by examining subscale fit and separation, individual item fit and logit difficulty, and subscale targeting.

Subscale fit and separation. Person and item fit for all four subscales was well within acceptable ranges (see Table 2). These values suggest that the observed item behavior matched that predicted by the Rasch model. In addition, the observed person behavior matched the prediction of the model. Person and item separation values above 2.0 are desirable. The scale showed good item separation, indicating that the sample was large enough to order the items. However, person separation is low, suggesting that there are not enough items to order the people in the sample. The low person separation suggests that the scale would benefit from more items.

Item fit and logit position. As shown in Tables 3-6, the items all show fit values within the range of .5-1.5. Similar to the subscales, the individual items had observed behavior consistent with what was predicted by the Rasch model. The logit position values listed in Tables 3-6 are essentially a measure of relative difficulty of the items. Items with higher logit difficulty require more of the latent trait (e.g., anger, restricted emotions) to endorse. For example, item 2 ("At times I get so angry I am violent toward others") has a higher logit difficulty than item 5 ("I have been more aggressive than usual lately"). Evidence of validity can be obtained by examining logit position to see if the items are ordered as the theory

would predict. In this case, it is reasonable that an individual might have enough anger to endorse item 5, but not enough to endorse item 2.

Targeting. Perhaps one of the more useful features of Rasch modeling is the fact that items are converted into an interval scale, This creates a "yardstick" of sorts with equal intervals (aka: logits). One can examine the logit difficulty of the items side-by-side with the logit ability of the persons. Figures 4-7 illustrate how each item functions to cover (or not cover) the location and spread of person abilities. It was found that the withdrawal and restricted emotions subscales were fairly well targeted to the sample.

The anger and substance use subscales were too difficult for the sample. One explanation for this is the fact that this was not a clinical sample. It is also likely that respondents may have been concerned about the repercussions for endorsing these items. Despite assurance of confidentiality of responses, previous research with this population indicates that suspicion over how the results will be used is prominent. The targeting for these scales indicate either rewording the items to be "easier" or administering the scale to a clinical sample to determine if the thresholds of the subscales were simply not able to be reached in a non-clinical population. Implications for the poor targeting as it relates to theory on men's depression are discussed below.

Hypothesis 5

Hypothesis 5 predicted that the scale would show discriminant validity with a measure of PTSD. The four subscales of the Men's Depression Inventory

showed low correlations with the PCL-C, suggesting that it is tapping a construct

that shares a small amount of variance with PTSD. Theoretically, this makes

sense, as emotional and physical difficulties would be present in each disorder,

but actual symptoms being assessed by the two scales do not have a good deal of

overlap.

<p style="text-align:center;">*Virtues of the Scale and Implications for Modification*</p>

The results of this study identify several useful features of the MDI. It

appears that the scale has good internal consistency, and each of the four

subscales operate as discreet entities. The withdrawal scale seems to tap into

restricted emotions as well, which indicates that

The analysis of the MDI suggests several avenues for modifications to

improve the scale. Interestingly, classical test theory (CTT) methods suggest that

the scale has adequate reliability and validity. The results of the entire analyses

show that the methods of CTT are severely limited in how they inform scale

development and analysis. The results of the CFA and Rasch modeling suggest

detailed improvements to the scale, which will not only yield a psychometrically

superior measure, but will inform our theoretical understanding of the construct of

men's depression in ways that are impossible with traditional analyses.

CFA Implications. The initial model shown in Figure 1 did not yield

adequate model fit, and improvements gained by removing the four non-

significant items were insufficient. However, of import is the fact that these four

items were the only reverse-keyed items in the scale. One possible explanation for

this is that respondents engaged in a repetitive response style, simply endorsing

items in a similar patter, which led to the reverse-keyed items to function poorly.

Given the sample size and in light of other psychometric data, it seems more

likely that the items were much easier to endorse. Indeed tables 3-6 show that

these items have lower logit difficulty. It can be reasoned that men who struggle

to admit their emotional distress on a survey might easily endorse an item such as

item 12 ("I enjoy the support I receive from others"). Conversely, it is also

reasonable that men with reluctance to endorse difficulties would also be reluctant

to *not* endorse an item tapping into healthy interpersonal tendencies. Overall this

suggests that reverse-keyed items should be removed from the scale.

When the model in Figure 3 was tested, which added cross-loadings where

the withdrawal items loaded on to the both withdrawal and restricted emotions

factors, sufficient model fit was obtained. The withdrawal items would

theoretically be endorsed by men experiencing emotional restriction, as in many

ways these behaviors overlap. Since collapsing the factors into one did not yield

sufficient model fit, this suggests that withdrawal continues to be a discreet

construct in men's depression. Thus a revision of the scale should include

withdrawal items worded more precisely, perhaps focusing on overt behaviors

rather than tendencies or preferences, to create more orthogonal factors.

Rasch modeling implications. The results of the study make a strong case

for including Rasch analyses in the psychometric evaluation of a new scale.

Overall the items were found to function together, but the items proved too

difficult for the sample. This was especially true for the anger and substance use subscales. The findings of Aggen, Neale, and Kendler (2005) discussed in Chapter 2 are relevant here, as they found that diagnostic categories become less sensitive at lower levels of severity. This would suggest that many of the scale items should be reworded for clarity and the language "softened" to improve the likelihood that they will be endorsed if true, and also points toward the need to administer the scale to a clinical sample.

However, it is desirable to have items at higher logit difficulties to avoid ceiling effects. Rather than simply reword the existing items, it will be advisable to add additional items to the scale at lower logit difficulties. This ensures that each subscale taps the range of symptomology in respondents. Separation statistics suggested that the sample size was adequate to order the items, but the number of items was insufficient to order the persons. It is recommended that the number of items for each scale be increased as well. Adding a number of items at lower logit difficulties for each subscale will address both of these limitations simultaneously.

Theoretical Implications for Men's Depression

Several findings from the study inform theory on how depression is represented in men. First, the information about targeting indicates that items assessing depression in men may need to have a much lower threshold of sensitivity. This is especially true when assessing substance use and anger. The sensitivity needed for these items, according to the Rasch model, suggests that the

60

expression of anger and substance use may be more subtle, and less overt, than implied the items used in this study. That is, men do not necessarily express depression as angry outbursts and significant substance abuse. These may either be more subtle for depressed men, or depressed men engaging in these behaviors may be less likely to admit them. Again, the simultaneous administration of a scale assessing defensiveness or symptom minimization is indicated. It may also be helpful to compare scores on these subscales to behavioral reports of others to assess the likelihood of underreporting or minimization of these symptoms.

A second important theoretical consideration is the relationship between withdrawal and emotional restriction in depressed men. The final CFA model suggests that men experiencing withdrawal also restrict their emotions, but not necessarily vice versa. It is clear that these items are tapping into discreet factors, as a model that collapses the two factors into one construct had poor model fit compared to the model in Figure 3. It seems more likely that the withdrawal items should be reworded to focus on discreet behaviors rather than feelings, which may overlap conceptually with restricted emotions.

It is important to note that the MDI is not claimed to measure a unique variant of depression. Rather, it aims to identify additional symptoms of depression that may be expressed (or uniquely expressed) in men. It is argued that assessing these symptoms may lead to identification of a depressive disorder in men that traditional measures may miss. As such, a more adequate definition of

depression for men would be a compilation of the traditional symptoms and those assessed in the MDI.

Finally, this study highlights the need for more sophisticated statistical analysis of new measures that go beyond CTT. If one were to draw conclusions about the MDI based on CTT findings (essentially, Hypotheses 1 and 4) it would be reasonable to conclude that the scale demonstrates good internal consistency reliability, convergent validity, and discriminant validity. However, the use of CFA and Rasch modeling highlight problems with the measure that can be addressed to create not only a more useful measure, but a more informed theory of depression in men. Targeting, logit difficulty, person behavior, and model fit are important psychometric aspects of the scale not assessed using CTT analyses.

Limitations of the Study and Future Directions for Research

This study has several limitations. The sample was drawn from a non-clinical population. While these men have been shown to have a high prevalence of depression (see Fields & Sherry, 2007), administration to a clinical sample would provide a broader range of depression and inform targeting of the scale items. In addition, there was no measure of how stringently these men adhered to traditional masculine gender roles. Such a measure would provide information on how responses on the MDI vary with endorsement of a more rigid male gender role. While there was some cultural diversity in the respondents, it was a predominantly Caucasian sample. Future administrations to a culturally diverse sample would inform theory on men's depression, depression and culture, and

62

psychometric properties of the scale when used with culturally diverse respondents.

In addition administering the scale to a more diverse sample, the findings on the MDI suggest several avenues for future research. Based on the results of the CFA and Rasch modeling, the scale will be revised to achieve better targeting. This will include adding additional items to the subscales, particularly at lower logit difficulties, reworking the withdrawal items to focus on overt behaviors, and the removal of reverse-keyed items. Future research studies should include a measure of defensiveness or symptom minimization, and possibly behavioral observations of anger and substance use. Data from a clinical sample is desirable to fine-tune the targeting across the range of depression, and it would be useful to have a greater number of respondents who are more severely depressed.

Conclusion

This study aimed to develop a measure of the unique presentation of depression in traditional men. It described the development of the scale using tenets of CTT, as well as the analysis of the scale using CTT, CFA, and Rasch modeling. It was shown that the scale would be found to have strong psychometric properties if CTT analyses were used alone. The results of CFA indicated a revision of the theoretical model based on the overlap of withdrawal items and restricted emotions items. CFA analysis also highlighted problematic behavior of reverse-keyed items in this scale. The results of the Rasch modeling suggest that the items were not well targeted for the sample, especially for the

anger and substance use subscales. The MDI would benefit from more items, particularly at lower logit difficulties.

It was argued that the use of CTT analyses limits the amount of information that can be gained about a scale, at best, and may lead to overconfidence in the psychometric properties of a new scale, at worst. CFA analyses and Rasch modeling provide information that can be useful in revising and fine-tuning a scale to fill a need in research or practice. Further, these analyses allow researchers to test a variety of hypotheses not accessible through CTT means. The reason for this study was to gain a further understanding of the unique presentation of depression in traditional men. The use of additional statistical procedures allow the answering of important research questions to more completely understand the unique experiences of men.

Figure 1

Original Proposed Model (Model 1)

65
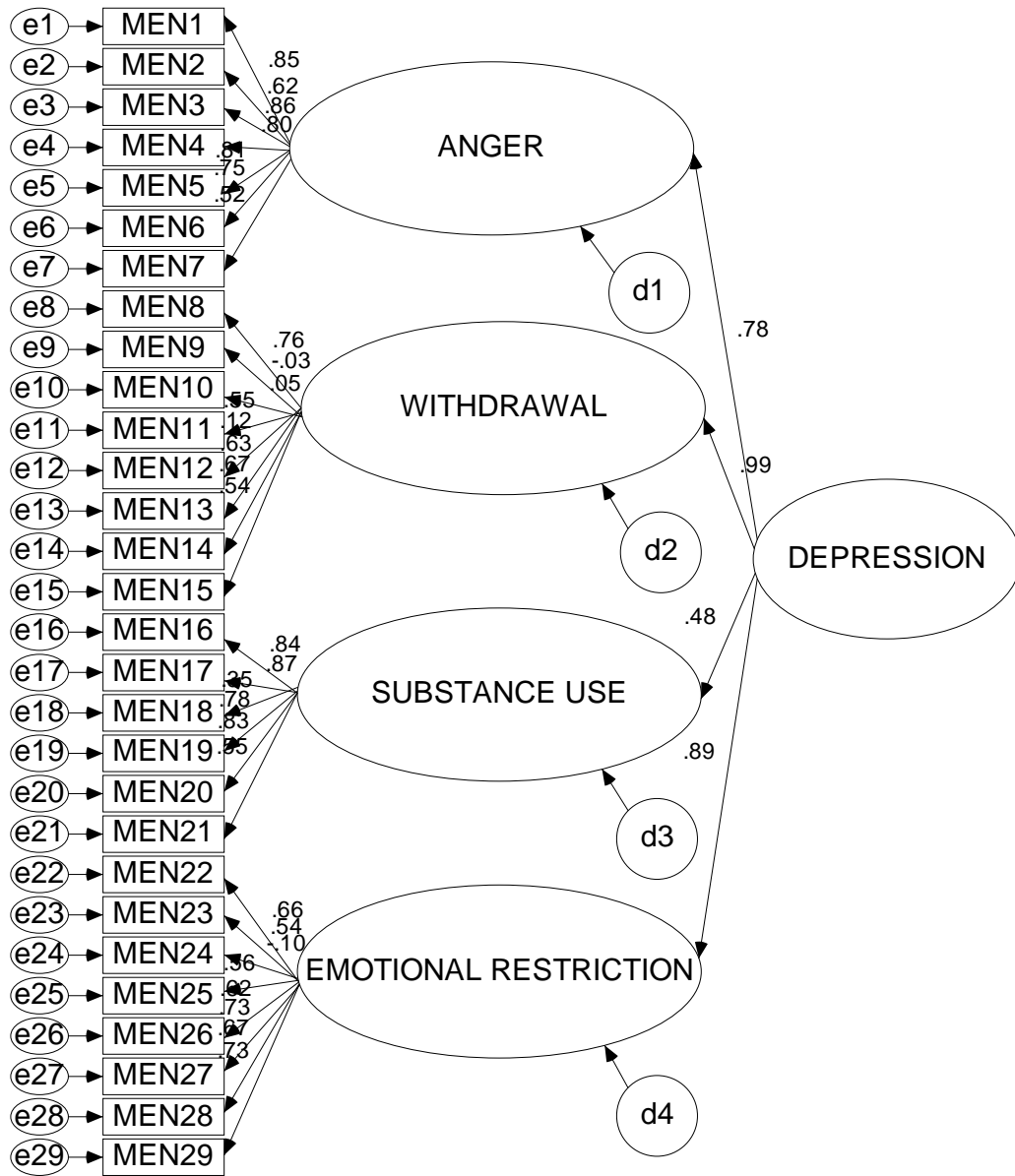
Figure 2

Revised Model – Model 2
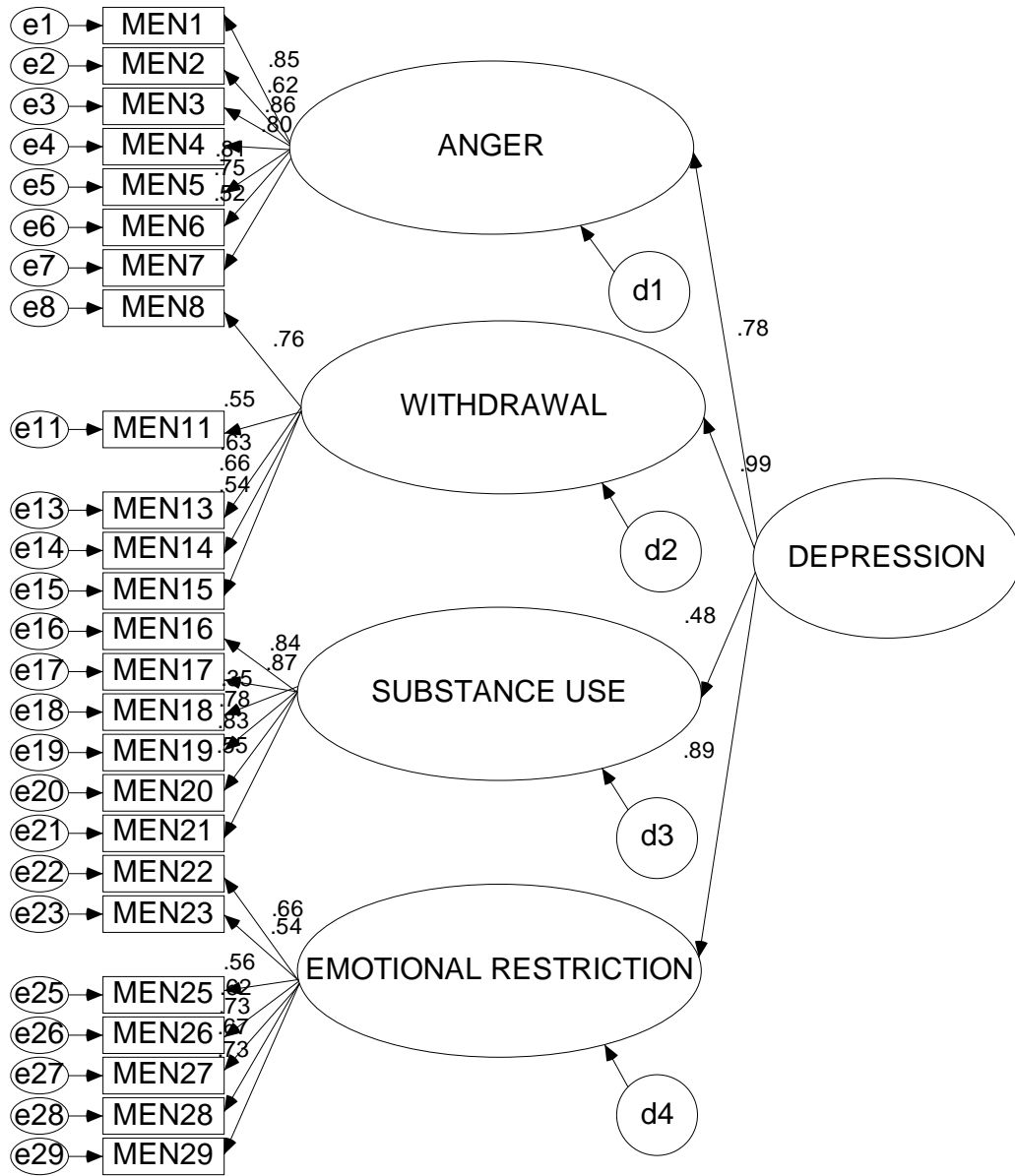
Figure 3

Model 3

Figure 4

Model 4

ANGER

WITHDRAWAL

EMOTIONAL RESTRICTION

SUBSTANCE USE

DEPRESSION

d1

d2

d3

d4

e1 MEN1
e2 MEN2
e3 MEN3
e4 MEN4
e5 MEN5
e6 MEN6
e7 MEN7
e8 MEN8
e11 MEN11
e13 MEN13
e14 MEN14
e15 MEN15
e25 MEN25
e26 MEN26
e22 MEN22
e23 MEN23
e27 MEN27
e28 MEN28
e29 MEN29
e16 MEN16
e17 MEN17
e19 MEN19
e18 MEN18
e20 MEN20
e21 MEN21

.80
.58
.84
.78
.78
.72
.48
.75
.79
.59
.72
.52
.52
.58
.63
.52
.70
.63
.71
.84
.86
.77
.35
.82
.55
.70
.73
.95
.46
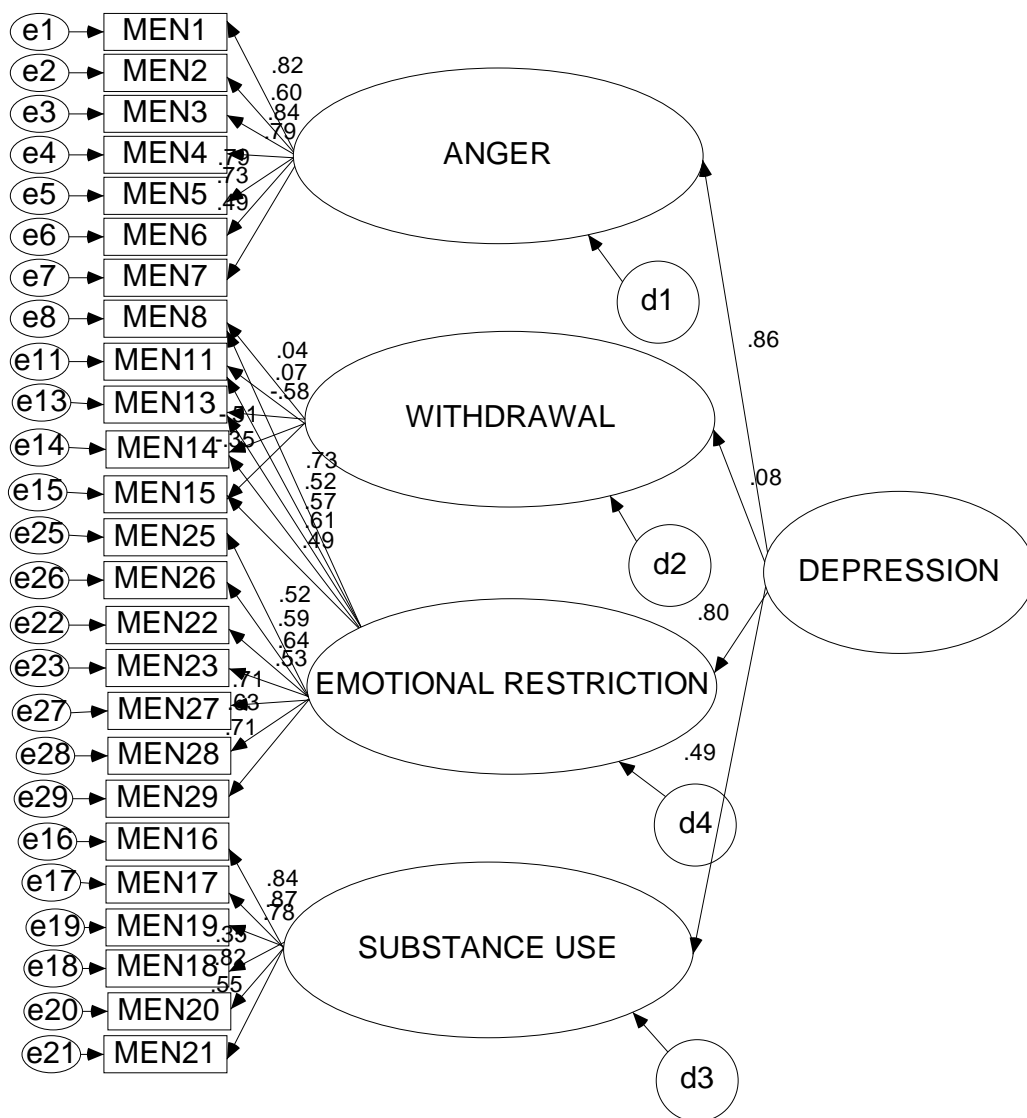
Figure 5

Model 5
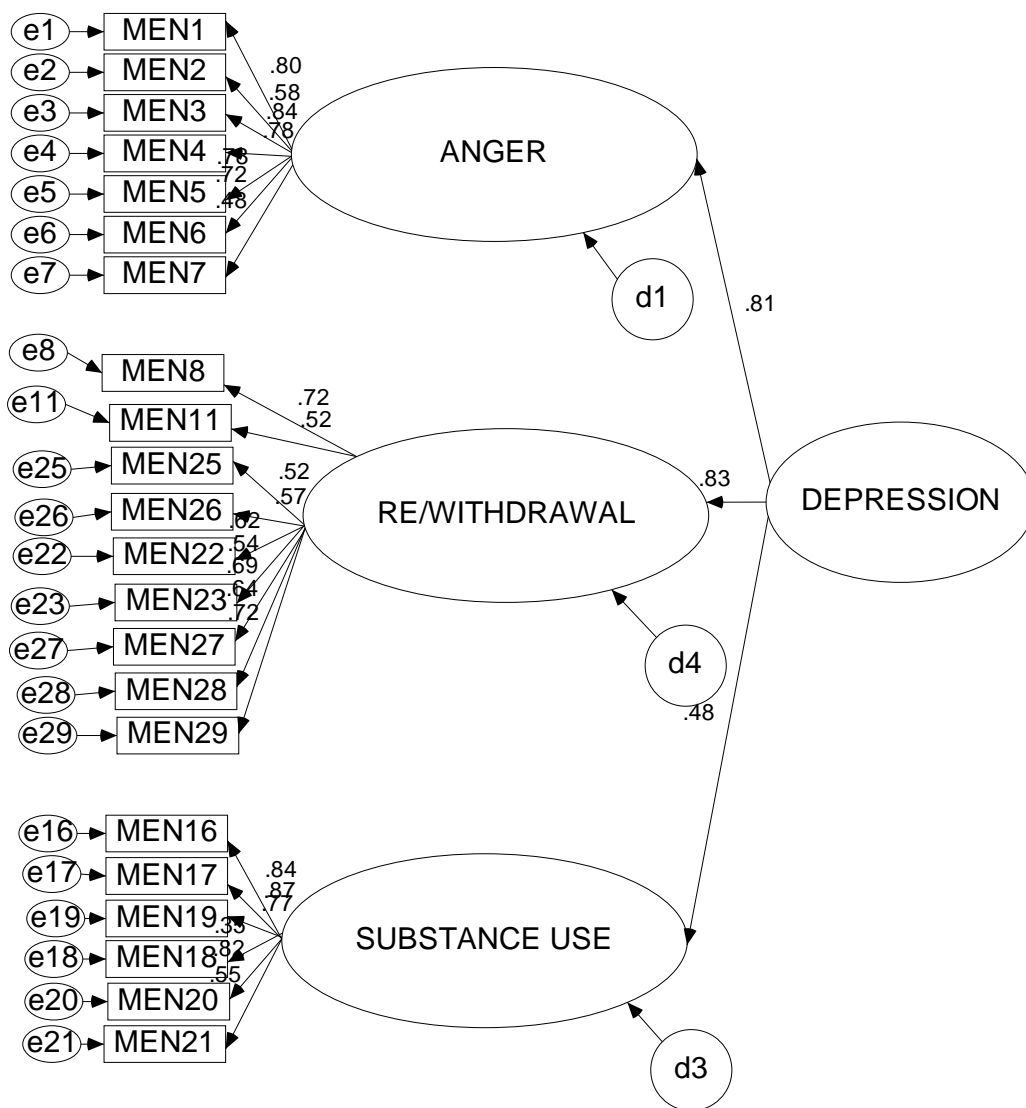
Figure 6

Anger Scale Item Map

```
INPUT: 441 PERSONS  7 ITEMS  MEASURED: 431 PERSONS  7 ITEMS  42 CATS     3.65.0
-----------------------------------------------------------------------------

EXPECTED SCORE: MEAN  (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold)
(ILLUSTRATED BY AN OBSERVED CATEGORY)
-5      -4      -3      -2      -1       0       1       2       3
|------+------+------+------+------+------+------+------|  NUM   ITEM
1               1   :     2 : 3 : 4 :  5   :    6  6    2   MEN2
|                                                     |
1                 1    :    2  : 3 :4 : 5    : 6   6    7   MEN7
|                                                     |
1             1    :    2 :  3:  4 :  5     :   6  6    4   MEN4
1          1    :    2    :   3 : 4 :  5    :   6  6    5   MEN5
|                                                     |
1            1    :    2   :  3 : 4 :  5     :   6  6    1   MEN1
|                                                     |
1         1    :     2  :  3 : 4  :  5     :    6  6    3   MEN3
|                                                     |
1     1    :      2   :  3 : 4  :   5    :    6  6    6   MEN6
|------+------+------+------+------+------+------+------|  NUM   ITEM
-5      -4      -3      -2      -1       0       1       2       3


  1
  0       2    3   3 2 2 1111111111
  3   2   7   17   6 6 0 79819009115846411111 2        PERSONS
     S           M          S           T
```
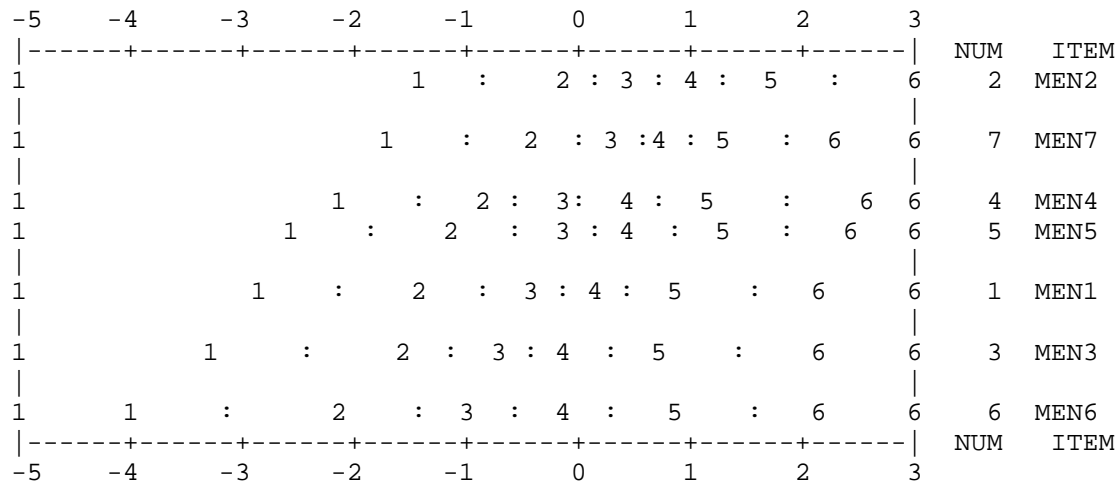
*M=Mean
S=One Std. Dev.
T=Two Std. Dev.

Figure 7

Restricted Emotions Scale Item Map

```
INPUT: 441 PERSONS  8 ITEMS  MEASURED: 429 PERSONS  8 ITEMS  48 CATS      3.65.0
--------------------------------------------------------------------------

EXPECTED SCORE: MEAN  (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold)
(ILLUSTRATED BY AN OBSERVED CATEGORY)
-4     -3      -2      -1       0       1       2       3
|------+------+------+------+------+------+------|  NUM    ITEM
1               1    :     2  : 3 : 4  :    5      :     66    5   MEN26
|                                                  |
1                1   :    2 : 3 : 4 :   5     :     6      6    7   MEN28
1              1     :    2  : 3 :4 :   5    :     6      6    6   MEN27
1            1    :     2  : 3 : 4  :    5     :    6      6    1   MEN22
1            1    :     2  : 3 : 4 :    5      :    6      6    4   MEN25
1            1   :     2   : 3 : 4 :   5    :     6      6    8   MEN29
1            1   :      2 : 3 : 4  :    5     :    6      6    2   MEN23
|                                                  |
|                                                  |
1       1    :     2  : 3  : 4 :   5     :     6            6    3   MEN24
|------+------+------+------+------+------+------|  NUM    ITEM
-4     -3      -2      -1       0       1       2       3

              311112423432211
 3      1    7  6195532953762670208857423         1   PERSONS
           T     S      M      S      T
```

*M=Mean
S=One Std. Dev.
T=Two Std. Dev.

71

Figure 8

Substance Use Scale Item Map

```
INPUT: 441 PERSONS  6 ITEMS  MEASURED: 429 PERSONS  6 ITEMS  36 CATS      3.65.0
-------------------------------------------------------------------------------

EXPECTED SCORE: MEAN  (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold)
(ILLUSTRATED BY AN OBSERVED CATEGORY)
-3         -2         -1          0          1          2          3
|---------+---------+---------+---------+---------+---------|  NUM    ITEM
1                1    :    2  : 3 : 4  :    5       :      6         6     2   MEN17
1              1     :      2  : 3 : 4 :   5        :    6          6     1   MEN16
|                                                                  |
1           1     :        2   : 3 :  4 :     5        :      6          6     5   MEN20
|                                                                  |
1                1     :    2  : 3 :4 :   5      :    6                6     6   MEN21
1      1        :       2    : 3  : 4     :     5          :        6    6     4   MEN19
|                                                                  |
|                                                                  |
1               1    :   2 : 3 :4 : 5      :     6                6     3   MEN18
|---------+---------+---------+---------+---------+---------|  NUM    ITEM
-3         -2          -1          0          1          2          3


2
0         2      2  2 131 1
8    1  12     2  41978 794775922442   1 2              1          PERSONS
          M              S                T
```

*M=Mean
S=One Std. Dev.
T=Two Std. Dev.

Figure 9

Withdrawal Scale Item Map

```
INPUT: 441 PERSONS  8 ITEMS  MEASURED: 431 PERSONS  8 ITEMS  48 CATS      3.65.0
-------------------------------------------------------------------------------

EXPECTED SCORE: MEAN  (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold)
(ILLUSTRATED BY AN OBSERVED CATEGORY)
-3         -2         -1          0          1          2          3
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
1               1    :     2 :  3 : 4 :  5   :    6         6    7   MEN14
|                                                           |
1            1       :     2 :  3 : 4 :  5      :    6      6    6   MEN13
1            1       :    2  : 3 : 4  : 5      :    6       6    1   MEN8
|                                                           |
|                                                           |
1         1     :       2  : 3 : 4 :   5     :    6         6    8   MEN15
|                                                           |
1           1     :      2  : 3 : 4 :  5      :    6        6    3   MEN10
|                                                           |
1     1      :        2 : 3  : 4 : 5       :     6          6    5   MEN12
1        1      :       2 :  3 : 4 : 5      :    6          6    2   MEN9
1        1    :      2  : 3 : 4 :  5       :      6         6    4   MEN11
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
-3         -2         -1          0          1          2          3

           1   11 1111232552321 1
3     5  1  6  0   38 318639302096963121 21                    PERSONS
           T       S        M        S         T
```

*M=Mean
S=One Std. Dev.
T=Two Std. Dev.

Table 1

Summary of Model Fit

	CFI	RMSEA	Chi-Square DF	AIC	CAIC
Model 1	0.827	0.073	1254.7 DF=37	1434.734	1447.905
Model 2	.893	0.065	777.9 DF=272	933.865	943.662
Model 3	.912	.059	681.5 DF=267	847.495	857.921
Model 4	.911	.059	689.4 DF=271	847.447	857.370
Model 5	.92	.060	537.3 DF=206	675.251	682.862

Table 2

Summary of Subscale Person and Item Fit

Subscale	Person Outfit	Person Infit	Person Separation	Person Reliability	Item Outfit	Item Infit	Item Separation	Item Reliability	Residuals
Anger	M=1.06 Std=1.14	M=1.02 Std=.79	1.54	.7	M=1.06 Std=0.33	M=1.02 Std=0.22	6.77	.98	M=.01 Std=1.03
Substance Use	M=1.01 Std=0.79	M=1.10 Std=0.86	0.81	.40	M=1.01 Std=.47	M=1.02 Std=.43	2.94	.90	M=-.02 Std=1.01
Withdrawal	M=1.01 Std=.75	M=1.04 Std=.72	1.17	.58	M=1.01 Std=.19	M=1.00 Std=.17	7.07	.98	M=-.01 Std=1.01
Restricted Emotions	M=1.04 Std=.81	M=1.08 Std=.80	1.31	0.63	M=1.04 Std=0.47	M=1.02 Std=.39	8.09	.98	M=-.03 Std=1.02

M=mean
Std=Standard Deviation

75

Table 3

Anger Subscale Item Logit Position and Fit

Item #	Text	Logit Position	MNSQ Infit	MNSQ Outfit
7	When all else fails, acting aggressively tends to solve my problems.	.44	1.48	1.76
2	At times I get so angry I am violent toward others.	.75	1.09	1.16
6	When things go badly I get angry.	-.67	1.13	1.15
1	I have been getting more angry than usual.	-.25	.89	.94
5	I have been more aggressive than usual lately.	.04	.90	.92
4	Others would say I've had a temper lately.	.14	.86	.66
3	Sometimes I get too angry.	-.46	.82	.85

Table 4

Withdrawal Subscale Item Logit Position and Fit

Item #	Text	Logit Position	MNSQ Infit	MNSQ Outfit
11	When I'm upset I just want to be left alone.	-.32	1.28	1.32
9	I have people I can rely on when I am having a hard time.	-.27	1.23	1.26
10	I find it easy to be around others.	-.13	1.09	1.12
15	I do not feel comfortable having others help me when I'm down.	.04	.97	1.00
12	I enjoy the support I receive from others.	-.26	.96	.93
13	Relying on others is a sign of weakness to me.	.26	.84	.86
8	I find myself pulling away from others.	.24	.85	.84
14	Needing others makes me feel like less of a man.	.46	.80	.78

Table 5

Substance Use Subscale Item Logit Position and Fit

Item #	Text	Logit Position	MNSQ Infit	MNSQ Outfit
18	I have had an alcohol or substance use problem in the past.	-.33	1.86	1.77
21	I tend to avoid situations where I am not able to drink.	-.06	1.29	1.51
19	Drinking alcohol can take the edge off during times of stress.	-.13	.88	.92
20	I tend to drink more when things aren't going well for me.	.06	.75	.73
16	I have been drinking more than usual.	.20	.70	.60
17	Drinking has helped me deal with things more easily.	.27	.62	.53

Table 6

Restricted Emotions Subscale Item Logit Position and Fit

Item #	Text	Logit Position	MNSQ Infit	MNSQ Outfit
24	I find it easy to put my feelings into words.	-.84	2.03	2.26
25	I have difficulty telling others I care about them.	.04	.98	1.05
23	It can be hard to describe how I feel.	-.04	.98	.99
26	Strong emotions are difficult for me to understand.	.42	.92	.86
22	I dislike talking with others about how I feel.	.06	.86	.88
27	Expressing feelings makes me feel open to attack by other people.	.19	.84	.81
28	I have difficulty expressing my emotional needs to my partner	.20	.77	.73
29	I have difficulty expressing my tender feelings.	-.03	.75	.77

Table 7

Summary of Standardized Path Coefficients by Item

Item	Path Coeff for Model		
	1	2	3
1. I have been getting more angry than usual.	.80	.80	.80
2. At times I get so angry I am violent toward others.	.59	.59	.58
3. Sometimes I get too angry	.84	.84	.84
4. Others would say I've had a temper lately.	.78	.78	.78
5. I have been more aggressive than usual lately.	.78	.78	.78
6. When things go badly I get angry.	.72	.72	.72
7. When all else fails, acting aggressively tends to solve my problems.	.48	.46	.48
8. I find myself pulling away from others.	.73	.73	.04
9. I have people I can rely on when I am having a hard time.	-.02	--	--
10. I find it easy to be around others.	.05	--	--
11. When I'm upset I just want to be left alone.	.52	.52	.07
12. I enjoy the support I receive from others.	.11	--	--
13. Relying on others is a sign of weakness to me.	.61	.61	-.58
14. Needing others makes me feel like less of a man.	.65	.64	-.52
15. I do not feel comfortable having others help me when I'm down.	.52	.52	.35
16. I have been drinking more than usual.	.84	.84	.84
17. Drinking has helped me deal with things more easily.	.86	.86	.86
18. I have had an alcohol or substance use problem in the past.	.35	.35	.77
19. Drinking alcohol can take the edge off during times of stress.	.77	.77	.35
20. I tend to drink more when things aren't going well for me.	.82	.82	.82
21. I tend to avoid situations where I am not able to drink.	.95	.85	.55
22. I dislike talking with others about how I feel.	.65	.65	.63
23. It can be hard to describe how I feel.	.52	.52	.62
24. I find it easy to put my feelings into words.	-.10	--	--
25. I have difficulty telling others I care about them.	.54	.54	.51
26. Strong emotions are difficult for me to understand.	.60	.60	.58
27. Expressing feelings makes me feel open to attack by other people.	.71	.71	.70
28. I have difficulty expressing my emotional needs to my partner.	.64	.65	.62
29. I have difficulty expressing my tender feelings.	.71	.71	.70

Using the scale provided, please **circle** the number which corresponds to your response to each item.

Strongly Disagree					**Strongly Agree**
1	**2**	**3**	**4**	**5**	**6**

The following questions refer to the PAST 2 WEEKS, including today.	CIRCLE ONE
1. I have been getting more angry than usual.	1 2 3 4 5 6
2. At times I get so angry I am violent toward others.	1 2 3 4 5 6
3. Sometimes I get too angry	1 2 3 4 5 6
4. Others would say I've had a temper lately.	1 2 3 4 5 6
5. I have been more aggressive than usual lately.	1 2 3 4 5 6
6. When things go badly I get angry.	1 2 3 4 5 6
7. When all else fails, acting aggressively tends to solve my problems.	1 2 3 4 5 6
8. I find myself pulling away from others.	1 2 3 4 5 6
9. I have people I can rely on when I am having a hard time.	1 2 3 4 5 6
10. I find it easy to be around others.	1 2 3 4 5 6
11. When I'm upset I just want to be left alone.	1 2 3 4 5 6
12. I enjoy the support I receive from others.	1 2 3 4 5 6
13. Relying on others is a sign of weakness to me.	1 2 3 4 5 6
14. Needing others makes me feel like less of a man.	1 2 3 4 5 6
15. I do not feel comfortable having others help me when I'm down.	1 2 3 4 5 6
16. I have been drinking more than usual.	1 2 3 4 5 6
17. Drinking has helped me deal with things more easily.	1 2 3 4 5 6
18. I have had an alcohol or substance use problem in the past.	1 2 3 4 5 6
19. Drinking alcohol can take the edge off during times of stress.	1 2 3 4 5 6
20. I tend to drink more when things aren't going well for me.	1 2 3 4 5 6
21. I tend to avoid situations where I am not able to drink.	1 2 3 4 5 6
22. I dislike talking with others about how I feel.	1 2 3 4 5 6
23. It can be hard to describe how I feel.	1 2 3 4 5 6
24. I find it easy to put my feelings into words.	1 2 3 4 5 6
25. I have difficulty telling others I care about them.	1 2 3 4 5 6
26. Strong emotions are difficult for me to understand.	1 2 3 4 5 6
27. Expressing feelings makes me feel open to attack by other people.	1 2 3 4 5 6
28. I have difficulty expressing my emotional needs to my partner.	1 2 3 4 5 6
29. I have difficulty expressing my tender feelings.	1 2 3 4 5 6

Portions of this questionnaire have been developed as a result of a joint effort between the Unions and the University of Denver to assist in developing and understanding employee health and wellness.   The results of this survey will be used to assist in better understanding and possibly developing a comprehensive wellness programs for Railroads. The ultimate goal being to improve work conditions and to make a better and safer work environment.

By completing this questionnaire you indicate your willingness and consent to participate in this project. Your participation is completely voluntary and anonymous and may be discontinued at any time. Individual responses to this questionnaire will be held completely confidential. Responses will be analyzed only by the University of Denver. Final summary reports will present trends, percentages, and written responses to open-ended questions. No information that could identify an employee will be reported.

Please complete the attached questionnaire by circling the number which best indicates your answer. Please complete the ENTIRE questionnaire and turn it in before you leave.

Thank you for your assistance.

Demographic Information

Note: These questions help us determine general characteristics of the people who respond to the questions. Fill out as many as you can. The more you fill out the more complete our results will be. We will NOT be reporting any individual responses. Only group averages will be used.

159. Please indicate your **gender**:   ____ Male   ____ Female

160. **Race**: ____ a) White   ____ b) Asian   ____ c) Black
        ____ d) Am Indian   ____ e) Hispanic ____ f) Other

161. Number of years of Education (*e.g. High school = 12 years*): _____

162.   Marital status: _____

163. How old are you?: _____

164. If you have been injured, whether you reported it or not, how many injuries have you had in the last <u>four</u> years?
        None __ One   __ Two __ Three __ Four   __ Five   __ Six or more

165. Length of time with UPRR: _____ (e.g. 2 yrs, 3 months).

166. Length of time at this location: _____ (e.g. 2 yrs, 3 months).

167. What is your craft? Engineer _____   Conductor _____

168. Please   describe   the   job   you   are   currently   on
        _____

169. Length of time in your present craft/position: _____ (e.g. 2 yrs, 3 months).

170. Is this an assigned job?   ____ Yes _____ No

171. If assigned what type of a schedule do you work?
        ___ 5 days a week ___ 6 days a week ___7 days a week ___ other

172. If assigned, what is your usual start time?   _____

173. What time did you start work today?:
        _____

174. About   what   time   do   you   quit today?:_____

175. Which Pool, or direction did you most recently work?: _____

176. Are you on the extraboard? YES_____   No _____

177. How long does it take for you to commute to work? _____

Name: (Optional): _____

Please give us your **Email** address if you would like a personalized copy of the results:
        _____ @ _____ . _____

**Remember: Only averages and percentages will be reported.
No identifying information will be released!**

138. Sadness	
0	I do not feel sad.
1	I feel sad much of the time.
2	I am sad all the time.
3	I am so sad or unhappy that I can't stand it.
139. Pessimism	
0	I am not discouraged about my future.
1	I feel more discouraged about my future than I used to be..
2	I do not expect things to work out for me.
3	I feel my future is hopeless and will only get worse.
140. Past Failure	
0	I do not feel like a failure.
1	I have failed more than I should have.
2	As I look back, I see a lot of failures.
3	I feel like I am a total failure as a person.
141. Loss of Pleasure	
0	I get as much pleasure as I ever did from the things I enjoy.
1	I don't enjoy things as much as I used to.
2	I get very little pleasure from the things I used to enjoy.
3	I can't get any pleasure from the things I used to enjoy.
142. Guilty Feelings	
0	I don't feel particularly guilty.
1	I feel guilty over many things I have done or should have done.
2	I feel quite guilty most of the time.
3	I feel guilty all of the time.
143. Punishment Feelings	
0	I don't feel I am being punished.
1	I feel I may be punished.
2	I expect to be punished.
3	I feel I am being punished.
144. Self-Dislike	
0	I feel the same about myself as ever.
1	I have lost confidence in myself.
2	I am disappointed in myself.
3	I dislike myself.
145. Self-Criticism	
0	I don't criticize or blame myself more than usual.
1	I am more critical of myself than I used to be.
2	I criticize myself for all of my faults.
3	I blame myself for everything bad that happens.

146. Suicidal Thoughts or Wishes
     0   **I don't have any thoughts of killing myself.**
     1   **I have thoughts of killing myself, but I would not carry them out.**
     2   **I would like to kill myself.**
     3   **I would kill myself if I had the chance.**

147. Crying
     0   **I don't cry anymore than I used to.**
     1   **I cry more than I used to.**
     2   **I cry over every little thing.**
     3   **I feel like crying, but I can't.**

148. Agitation
     0   **I am no more restless or wound up than usual.**
     1   **I feel more restless or wound up than usual.**
     2   **I am so restless or agitated that it's hard to stay still.**
     3   **I am so restless or agitated that I have to keep moving or doing something.**

149. Loss of Interest
     0   **I have not lost interest in other people or activities.**
     1   **I am less interested in other people or things than before.**
     2   **I have lost most of my interest in other people or things.**
     3   **It's hard to get interested in anything.**

150. Indicisiveness
     0   **I make decisions about as well as ever.**
     1   **I find it more difficult to make decisions than usual.**
     2   **I have much greater difficulty in making decisions than I used to.**
     3   **I have trouble making any decisions.**

151. Worthlessness
     0   **I do not feel I am worthless.**
     1   **I don't consider myself as worthwhile and useful as I used to.**
     2   **I feel more worthless as compared to other people.**
     3   **I feel utterly worthless.**

152. Loss of Energy
     0   **I have as much energy as ever.**
     1   **I have less energy than I used to have.**
     2   **I don't have enough energy to do very much.**
     3   **I don't have enough energy to do anything.**

153. Changes in Sleeping Patterns
  - **0** I have not experienced any change in my sleeping pattern.
  - **1a** I sleep somewhat more than usual.
  - **1b** I sleep somewhat less than usual.
  - **2a** I sleep a lot more than usual.
  - **2b** I sleep a lot less than usual.
  - **3a** I sleep most of the day.
  - **3b** I wak up 1-2 hours early and can't get back to sleep.

154. Irritability
  - **0** I am no more irritable than usual.
  - **1** I am more irritable than usual.
  - **2** I am much more irritable than usual.
  - **3** I am irritable all of the time.

155. Changes in Appetite.
  - **0** I have not experienced any change in my appetite.
  - **1a** My appetite is somewhat less than usual.
  - **1b** My appetite is somewhat greater than usual.
  - **2a** My appetite is much less than before..
  - **2b** My appetite is much greater than usual..
  - **3a** I have no appetite at all.
  - **3b** I crave food all of the time.

156. Concentration Difficulty
  - **0** I can concentrate as well as ever.
  - **1** I can't concentrate as well as usual.
  - **2** It's very hard to keep my mind on anything for very long.
  - **3** I find I can't concentrate on anything.

157. Tiredness or Fatigue
  - **0** I am no more tired or fatigued than usual.
  - **1** I get more tired or fatigued more easily than usual.
  - **2** I am too tired or fatigued to do a lot of things I used to.
  - **3** I am too tired or fatigued to do most of the things I used to do.

158. Loss of Interest in Sex
  - **0** I have not notieced any recent change in my interest in sex.
  - **1** I am less interested in sex than I used to be.
  - **2** I am much less interested in sex now.
  - **3** I have lost interest in sex completely.

Appendix E – PTSD Checklist Civilian Version

The following questions refer to the most STRESSFUL incident you can recall. Do you…. (circle number)	Not at all	A little bit	Moder ately	Quite a bit	Extre mely
1. **Have repeated, disturbing memories, thoughts, or images of the stressful extent?**	1	2	3	4	5
2. **Have distressing dreams of this event?**	1	2	3	4	5
3. **Suddenly act or feel as if the stressful event were happening again?**	1	2	3	4	5
4. **Feel very upset when something reminded you of the stressful experience?**	1	2	3	4	5
5. **Have physical reactions (e.g., sweating, trouble breathing, heart pounding) when something reminded you of the stressful event?**	1	2	3	4	5
6. **Avoid thinking or talking about the stressful experience or avoid having feelings related to it?**	1	2	3	4	5
7. **Avoid activities or situations because they remind you of the stressful experience?**	1	2	3	4	5
8. **Have trouble remembering important parts of the stressful experience?**	1	2	3	4	5
9. **Lose interest in activities that you used to enjoy?**	1	2	3	4	5
10. **Feel distant or cut-off from other people?**	1	2	3	4	5
11. **Feel emotionally numb?**	1	2	3	4	5
12. **Feeling as if your future would somehow be cut short?**	1	2	3	4	5
13. **Have trouble falling or staying asleep?**	1	2	3	4	5
14. **Feel irritable or have angry outbursts?**	1	2	3	4	5
15. **Have difficulty concentrating?**	1	2	3	4	5
16. **Become super alert or vigilant?**	1	2	3	4	5
17. **Feel jumpy or easily startled?**	1	2	3	4	5
18. **Were there any fatalities in this incident?**	Yes	No			
19. **Were the persons killed close to you?**	Yes	No			
20. **How many fatalities were there (if none, leave blank).**	1	2	3	4	5

References

Andrykowski, M.A., Cordova, M.J., Studts, J.L., & Miller, T.W. (1998).

Posttraumatic stress disorder after treatment for breast cancer: Prevalence

of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C_ as

a screening instrument. *Journal of Consulting and Clinical Psychology,*

66, 586-590.

Arrindell, W.A., Steptoe, A., & Wardle, J. (2003). Higher levels of state

depression in masculine than in feminine nations. *Behaviour Research and*

Therapy, 41, 809-817.

Arrindell, W.A., Hatzichristou, C., Wensink, J., Rosenberg, E., van Twillert, B.,

Stedema, J., & Meijer, D. (1997). Dimensions of national culture as

predicators of cross-national differences in subjective well-being.

Personality and Individual Differences, 23, 37-53.

Beck, A.T. (1988). Psychometric properties of the Beck Depression Inventory:

Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77-100.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). Manual for the Beck Depression

Inventory-II. San Antonio, TX: Psychological Corporation.

Blair-West, G.W., Cantor, C.H., Mellsop, G.W., & Eyeson-Annan, M.L. (1999).

Lifetime suicide risk in major depression: Sex and age determinants.

Journal of Affective Disorders, 55, 171-178.

Blanchard, E.B., Jones-Alexander, J., Buckley, T.C. (1996). Psychometric

properties of the PTSD Checklist (PCL). *Behavior Research and Therapy,*

34, 669-673.

Brownhill, S., Wilhelm, K., Barclay, L., & Schmied, V. (2005). 'Big build':
Hidden depression in men. *Australian and New Zealand Journal of
Psychiatry, 39,* 921-931.

Bond, M.H. (2001). Reclaiming the individual from Hofsted'es ecological
analysis—a 20-year odyssey: Comment on Oyserman et al. (2002).
Psychological Bulletin, 128, 73-77.

Canino, G., Lewis-Fernandez, R., & Bravo, M. (1997). Methodological
challenges in cross-cultural mental health research. Transcultural
Psychiatry, 34, 163–184Draguns, J.G. & Tanaka-Matsumi, J. (2003).
Assessment of psychopathology across and within cultures: Issues and
findings. *Behaviour Research and Therapy, 41,* 755-776.

Chan, B., Parker, G., Tully, L., & Eisenbruch, M. (2007). Cross-cultural
validation of the DMI-10 Measure of State Depression: The development
of a Chinese language version. *Journal of Nervous and Mental Disease,
195,* 20-25.

Chang, S.M., et al. (2008). Cross-national difference in the prevalence of
depression caused by the diagnostic threshold. *Journal of Affective
Disorders, 106,* 159-167.

Chu, B.C. (2007). Considering culture one client at a time: Maximizing the
cultural exchange. *Pragmatic Case Studies in Psychotherpy, 3,* 34-43.

Cochran, S.V. & Rabinowitz, F. E. (2000). *Men and depression: Clincial and
empirical perspectives.* San Diego, CA: Academic Press.

Cochran, S.V., & Rabinowitz, F.E. (2003). Gender-sensitive recommendations for

assessment and treatment of depression in men. *Professional Psychology:*

Research and Practice, 34, 132-140.

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective

well-being of nations. *Journal of Personality and Social Psychology, 69,*

851-864.

de Graaf, R., Bijil, R.V., Smit, F., Vollebergh, W.A.M., & Spijker, J. (2002). Risk

factors for 12-month comorbidity of mood, anxiety, and substance use

disorders: Finding from the Netherlands Mental Health Survey and

Incidence Study. *American Journal of Psychiatry, 159,* 620-629.

Egeland, J. & Hostetter, A. (1983). Amish Study I: Affective disorders among the

Amish. *American Journal of Psychiatry, 140,* 56-61.

Fava, G.A., Kellner, R., Munari, F., & Pavan. L. (1982). The Hamilton

Depression Rating Scale in normals and depressives: A cross cultural

validation. *Acta Psychiatrica Scandinavica, 66,* 26-32.

Fields, A.J., Sherry, P., & Green, K. (2007, March). *Development and Evaluation*

of a Men's Depression Scale. A poster session presented at the meeting of

the Rocky Mountain Psychological Association.

Fields, A.J., & Sherry, P. (2008, March). *Depression, support, and Safety:*

Prevalence and Implications for Male Railroad workers. A paper

presented at the annual meeting of the APA/NIOSH, Washington , D.C.

Finn, Z., Bille, A., Rutz, W., & Bech, P. (2002). The Gotland Male Depression
    Scale: A validity study in patients with alcohol use disorder. *Nordic*
    Journal of Psychiatry, 56, 265-271.

Flett, G.L., Vredenburg, K., & Krames, L. (2004) The continuity of depression in
    clinical and nonclinical samples. *Psychological Bulletin, 121,* 395-416.

Fox, C.M. & Jones, J.A. (1998). Uses of Rasch modeling in counseling
    psychology research. *Journal of Counseling Psychology, 45,* 30-45.

Furukawa, T.A. et al. (2005). Cross-cultural equivalence in depression
    assessment: Japan-Europe-North American study. *Acta Psychiatrica*
    Scandinavica, 112, 279-285.

Girling, D., Barkley, C., Paykel, E., Gehlhaar, E., Brayne, C., Gill, C., Mthewson,
    D., & Huppert, F. (1995). The prevalence of depression in a cohort of the
    very elderly. *Journal of Affective Disorders, 34,* 319-329.

Hanna, E.., & Grant, B. (1997). Gender differences in DSM-IV alcohol use
    disorders and major depression as distributed in the general population:
    Clinical implications. *Comprehensive Psychiatry, 38,* 202-212.

Harvey, R.J. & Hammer, A.L. (1999). Item respons theory. *The Counseling*
    Psychologist, 27, 353-383.

Hatcher, R.L. & Gillaspy, J.A. (2006). Development and validation of a revised
    short version of the Working Alliance Inventory. *Psychotherapy Research,*
    16, 12-25.

Hofstede, G. (1980). Culture's Consequences: International Differences in Work-
    Related Values. Beverley Hills, CA: Sage.

Hofstede, G. (1991). Cultures and Organizations: Software of the Mind. London:

McGraw-Hill.

Hofstede, G. (2001). Culture's Consequences: Comparing Values, Behaviours,

Institutions, and Organizations Across Nations (2nd edition). Thousand

Oakes, CA: Sage.

Hong, S., Kim, B.S.K., & Wolfe, M. (2005). A psychometric revision of the

European American Values Scale for Asian Americans using the Rasch

model. *Measurement and Evaluation in Counseling and Development, 37,*

194-207.

Hwang, W.C. & Wood, J.J. (2007). Being culturally sensitive is not the same as

being culturally competent. *Pragmatic Case Studies in Psychotherpy,3,*

44-50.

Kleinman AM. 1977. Depression, somatization and the ''New Cross-Cultural

Psychiatry.'' *Social Sciences and Medicine, 11,* 3-10.

Kojima et al. (2002). Cross-cultural validation of the Beck Depression Inventory-

II in Japan. *Psychiatry Research, 110,* 291-299.

Lasa, L., Ayuso-Mateos, J.L., Vazquez-Barquero, J.L., Diez-Manrique, F.J.,

Dowrick, C.F. (2000). The use of the Beck Depression Inventory to screen

for depression in the general population: A preliminary analysis. *Journal*

of Affective Disorders, 57, 261-265.

Levant, R.F. (1996). The new psychology of men. *Professional Psychology:*

Research and Practice, 27, 259-265.

Linacre, J. M. (2007). *A user's guide to winsteps ministep: Rasch-model computer programs*. Chicago: Winsteps.

Loewenthal, K., Goldblatt, V., Gorton, T., Lubitsch, G., Bickness, H., Fellowes, D., & Sowden, A. (1995) Gender and Depression in Anglo-Jewry. *Psychological Medicine, 25,* 1051-1063.

Lo´pez, S.R., & Guarnaccia, P.J.J. (2000). Cultural psychopathology: Uncovering the social world of mental illness. *Annual Review of Psychology, 51,* 571-598.

Moller-Leimkuhler, A.M. (2002). Barriers to help-seeking by men: A review of sociocultural and clinical literature with particular reference to depression. *Journal of Affective Disorders, 71,* 1-9.

Moller-Leimkuhler, AM., Bottlender, R., Straub, A., & Rutz, W. (2004). Is there evidence for a male depressive syndrome in inpatients with major depression? *Journal of Affective Disorders, 80,* 87-93.

Murphy, J., Olivier, D., Monson, R., Sobol, A., & Leighton, A. (1988) Incidence of depression and anxiety: The Stirling County Study. *American Journal of Public Health, 78,* 534-540.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin, 101,* 259-282.

O'Neil, J., Good, G., & Holmes, S. (1995). Fifteen years of theory and research on men's gender role conflict: New paradigms for empirical research. In R. Levant & W. Pollack (Eds.), *A new psychology of men* (pp. 164-206). New York: Basic Books.

O'Neil, J., Helms, B., Gable, R., David, L., Wrightsman, L. (1986). Gender-Role

    Conflict Scale: college men's fear of femininity. *Sex Roles, 14(5/6),* 335-

    350.

Pastor, D.A. & Beretvas, S.N. (2006). Longitudinal Rasch modeling in the context

    of psychotherapy outcomes assessment. *Applied Psychological*

    Measurement, 30, 100-120.

Pleck. J.H. (1981). *The myth of masculinity.* Cambridge, MA: Massachusetts

    Institute of Technology Press.

Pollack, W.S. (1998). Mourning, melancholia, and masculinity: Recognizing and

    treating depression in men. In W. Pollack & R. Levant (Eds.), *New*

    Psychotherapy for men (13-34). New York: John Wiley & Sons, Inc.

Redmond, M., Rooney, R., & Bishop, B. (2006). Unipolar depression across

    cultures: A Delphi analysis of the methodological and conceptual issues

    confronting the cross-cultural study of depression. *Australian e-Journal*

    for the Advnancement of Mental Health, 5, 1-13.

Rutz, w., von Knorring, L., Pihlgren, H., Rihmer, Z., Walinder, J., (1995).

    Prevention of male suicideds: Lessons from Gotland study. *Lancet,* 345-

    524.

Rutz, W. (1999) Improvement of care for people suffering from depression: The

    need for comprehensive education. *International Clinical*

    Psychopharmacology, 14, 27-33.

Santor, D.A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in

    depression. *Measurement, 4,* 135-155.

Shepard, D. (2002). A negative state of mind: patterns of depressive symptoms among men with high gender role conflict. *Psychology of Men and Masculinity, 3(1),* 3-8.

Slade, T. & Andrews G. (2005) Latent structure of depression in a community sample: A taxometric analysis. *Psychological Medicine, 35,* 489-497.

Snaith, P. (1993). What do depression rating scales measure?. *British Journal of Psychiatry, 163,* 293-298.

Spector, P.E., Cooper, C.L., & Sparks, K. (2001).an international study of the psychometric properties of the Hofsted Values Survey Module 1994: A comparison of the individual and country/province level results. *Applied Psychology: An International Review, 50,* 269-281.

Sue, S. (1998). In search of cultural competence in psychotherapy and counseling. *American Psychologist, 53,* 440–448.

Sue, D. W., & Torino, G. C. (2005). Racial–cultural competence: Awareness, knowledge, and skills. In R. T.Carter (Ed.), Handbook of racial–cultural psychology and counseling: *Vol. 2*. *Training and practice* (pp. 3–18). Hoboken, NJ: Wiley.

Tanaka-Matsumi, J., & Marsella, A. J. (1976). Cross-cultural variations in the phenomenological experience of depression: I. Word association studies. Journal of Cross-Cultural Psychology, 7, 379–396.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G., Joyce, P. R., Karam, E. G.,Lee, E. G., Lellouch, J., Lepine, J. P., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickmaratne, P. J.,

Wittchen, H. U., & Yeh, E. K. (1996). Cross-national epidemiology of

major depressive and bipolar disorder. Journal of the American Medical

Association, 276, 293–299.

Whaley, A.L. & Davis, K.E. (2007). Cultural competence and evidence-based

practice in mental health services: A complementary perspective.

American Psychologist, 62, 563-574.

Winkler, D., Pjrek, E., & Kasper, S. (2005). Anger attacks in depression:

Evidence for a male depressive syndrome. *Psychotherapy and*

Psychosomatics, 74, 303-307.

Zimmerman, M., Chelminski, I., & McDermut, W. (2002). Major depressive

disorder and axis I diagnostic comorbidity. *Journal of Clinical Psychiatry,*

63, 187-193.