University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

8-2010

Gender and Posttraumatic Stress Disorder Screening in the Military: A Measurement Study

Mark Allan Oliver
moliver5@utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

 Part of the Clinical and Medical Social Work Commons, and the Social Work Commons

Recommended Citation

Oliver, Mark Allan, "Gender and Posttraumatic Stress Disorder Screening in the Military: A Measurement Study. " PhD diss., University of Tennessee, 2010.
https://trace.tennessee.edu/utk_graddiss/837

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Mark Allan Oliver entitled "Gender and Posttraumatic Stress Disorder Screening in the Military: A Measurement Study." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Social Work.

William Nugent, Major Professor

We have read this dissertation and recommend its acceptance:

John G. Orme, J. Camille Hall, Robert T. Ladd, Russell Zaretzki

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Mark Allan Oliver entitled "Gender and Posttraumatic Stress Disorder Screening in the Military: A Measurement Study." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Social Work.

William Nugent
Major Professor

We have read this dissertation
and recommend its acceptance:

John G. Orme

J. Camille Hall

Robert T. (Tom) Ladd

Russell Zaretzki

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original Signatures are on file with official student records)

GENDER AND POSTTRAUMATIC STRESS DISORDER SCREENING IN THE MILITARY: A MEASUREMENT STUDY

A Dissertation Presented for
the Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Mark Allan Oliver
August 2010

Copyright © by Mark Allan Oliver
All rights reserved

***Disclaimer**: The views expressed in this dissertation are those of the author, and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U. S. Government.*

DEDICATION

This dissertation is dedicated to my loving wife. It was you who encouraged me to pursue doctoral education when I had my doubts. Your support never wavered through a challenging process. I love you, and once again—you were right!

ACKNOWLEDGEMENTS

Most of all, I thank God for His grace and mercy. I am incredibly grateful for the support of my wife, Jackie, and my kids Caitlin, Haley, and Seth. They patiently put up with me missing dinner to work late and going to the office on weekends—both of which occurred more often than they deserve.

I also owe a great deal to my parents for their love, encouragement, and guidance. My dad, especially, taught me the importance of working hard. While his words were important, it was, and remains, his example that inspires me to try to implement that ethic.

Thanks to Dr. Bill Nugent, my chair, for his advice and counsel throughout this program. You were always available to talk; whether it was about cohort issues, study ideas, or just to answer questions as I wrestled with my research direction. I also want to thank all the members of my dissertation committee for their constructive feedback and support. Dr. John Orme, Dr. Camille Hall, Dr. Tom Ladd, and Dr. Russell Zaretzki, you all helped me think about my topic in new ways and your input has been invaluable.

To Dr. Orme, especially, I am grateful to have served as your teaching assistant. That positive experience opened me up to other teaching opportunities and to a realization of my own love for teaching. Also, thank you for listening and being willing to share your insights and wisdom about the academic profession.

I want to acknowledge Dr. Maria Edelen for her willingness to freely provide her expertise as I worked on completing my analyses. Her responsiveness and detailed, but accessible answers made a difficult subject much simpler to handle.

John Bailey, I've enjoyed our conversations and I want to thank you for always being available to support us students. Thanks for all the hiking advice! It amazed me how quickly problems were remedied just by bringing them to John's attention.

Finally, my sincere thanks go to Lt Col Theresa Lawson, who as my supervisor at Eielson AFB, Alaska, mentored me and supported my efforts to get into the Air Force Institute of Technology (AFIT) doctoral program in social work. There were times when I was not excited about how you pushed, but I can't argue with the results. Thanks again.

ABSTRACT

The Primary Care Posttraumatic Stress Disorder (PC-PTSD) screen (Prins et al., 2003) is used by the Department of Defense to identify military members who are at increased risk of PTSD. This screen has been offered to all returning deployers since 2005. However, validation studies of PC-PTSD scores from military samples have seldom employed a significant number of female subjects and no published studies have examined it for gender bias. Ruling out bias is important because routine under-identification of PTSD risk in any group could result in hindered access to needed assessment and/or care. With the current proportion of military females historically high (Women's Research & Education Institute, 2007), it is imperative that the PC-PTSD be analyzed to ensure measurement equivalence across gender. Using a large sample of male and female veterans returning from deployment, the validity of the PC-PTSD scores was first examined by conducting a differential item functioning (DIF) analysis across male and female subgroups. Then, using a clinical diagnosis as the criterion, both logistic regression and diagnostic likelihood ratio methods were employed to assess for differential predictive validity by gender. Finally, confirmatory factor analysis (CFA) was used to examine convergent and divergent validity in a two-factor model containing both PC-PTSD and depression screen responses. Results revealed no statistically significant gender-related DIF or differential prediction of PTSD by PC-PTSD scores. Good convergent and divergent validity were also observed in the CFA analysis. The results generally supported the continued use of the PC-PTSD with both male and female military veterans returning from deployment. Limitations of the study and recommendations for future research were discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Problem and Significance

With the initiation of the Post-Deployment Health Assessment (PDHA) program during 2004, the Department of Defense (DoD) began routinely screening all military members who were returning from Iraq and Afghanistan for posttraumatic stress disorder (PTSD). This includes members of the US Air Force (USAF; Office of the USAF Surgeon General, 2008). The DoD subsequently augmented this program with the Post-Deployment Health Reassessment (PDHRA) which screens for PTSD three to six months after the return from deployment (Bliese et al., 2008). Both of these initiatives use the same PTSD instrument: the Primary Care PTSD screen (PC-PTSD; Prins et al., 2003). Each program was implemented to maximize early detection of PTSD in order to increase timely access to appropriate care for affected military members (Bliese, Wright, Adler, Thomas, & Hoge, 2007).

Recent research has indicated that military service in Iraq and Afghanistan is associated with elevated rates of mental health problems and, specifically, PTSD. Hoge, Auchterlonie, and Milliken (2006) reported that more than 19% of those returning from deployment to Iraq and over 11% of those coming back from Afghanistan self-identified as having some type of mental health difficulty. A large military cohort study (Smith et al., 2008) found that a majority of those who were deployed in the current conflicts were exposed to combat experiences and that as many as 8.7% of those who were exposed later received a new onset PTSD diagnosis or were identified as having PTSD symptoms. By contrast, no more than 3% of non-deployers had a new onset of PTSD or its symptoms.

PTSD has been strongly associated with lifetime comorbidity with serious mental health problems such as major depression, bipolar disorder, and substance abuse disorders in the general population (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Within the military, increased mental health problems have been related to greater attrition from service (Hoge et al., 2006). Veterans with PTSD or trauma symptoms have demonstrated higher rates of self-reported marital dissatisfaction and marital violence (Nelson-Goff, Crow, Reisbig, & Hamilton, 2007; Jordan et al., 1992)

Given these problems and the elevated PTSD risk for veterans, prompt post-deployment screening appears to provide a valuable opportunity to self-identify for helping services for those experiencing combat-related mental health problems (Wright et al., 2007). Studies examining trauma exposure have generally reported that males are more likely than females to report experiencing combat (Flett, Kazantzis, Long, McDonald, and Millar, 2004; Kessler et al., 1995). Recent casualty figures, however, may point to a narrowing of the combat exposure gender gap. Whereas women constituted only about .01% (8 total females) of United States (US) casualties during Vietnam (National Archives and Records Administration, 2007), that figure has increased to 2.4% (121 total females) as of June 6, 2009, in Iraq and Afghanistan combined (Defense Manpower Data Center, 2009a; Defense Manpower Data Center, 2009b). Considering this casualty information along with the historically high proportion (14.3%) of US females currently serving in the active duty military (Women's Research & Education Institute, 2007), it is reasonable to assume that the current counterinsurgency conflicts expose women to combat-type trauma to a greater degree than in the past.

There is also some evidence that PTSD manifests differently in men and women. For instance, despite generally lower rates of exposure to trauma-inducing events (Breslau et al., 1998; Kessler et al., 1995), women have nearly twice the risk of developing PTSD (Kessler et al., 1995; Tolin & Foa, 2006). Several studies have indicated differences in the types of trauma experienced by males and females. Males tend to be more frequently exposed to serious accidents (Kessler et al., 1995; Stein, Walker, Hazen, & Forde, 1997) and, as mentioned above, combat. Females, on the other hand, are disproportionately exposed to sexually assaultive violence (Breslau et al., 1998; Costello, Erkanli, Fairbank, & Angold, 2002). Also, some studies show gender differences in symptom identification on PTSD instruments (Breslau, Chilcoat, Kessler, Peterson, & Lucia, 1999; Fullerton et al., 2001).

When considering apparent gender differences as measured by PTSD instruments, an often implicit assumption is that any measured differences are the result of real differences in the manifestation of the underlying PTSD trait between tested groups. An alternative explanation is that the instrument itself is biased. This would mean that the instrument might, in this instance, systematically show higher or lower scores for one gender despite similar levels of underlying PTSD among males and females.

Despite the widespread employment of the PC-PTSD in screening returning combat veterans, current validation studies for that instrument have thus far not addressed the question of gender bias. In one study (Bliese et al., 2008), only 4% of the respondents were female and another (Bliese, Wright, Adler, & Thomas, 2004) provided no information on the gender breakdown of its active duty military sample. Prins and colleagues (2003) used a more gender-

balanced sample but did not specifically assess for measurement bias. The latter researchers did, however, find markedly lower sensitivity and specificity for women as compared to men at the recommended cutoff level of three out of four "yes" responses.

This result suggests some potential problems associated with the differential performance of the PC-PTSD by gender. Lower sensitivity corresponds to a higher false negative rate (given by 1 – sensitivity), which in this instance indicates that a higher proportion of females at risk of developing PTSD are not identified by the instrument as having that risk. Since those who are identified as having a positive PC-PTSD result are referred for an in-person clinical evaluation (Office of the USAF Surgeon General, 2008) this type of misidentification could lead to fewer women receiving needed assessments. Thus, a higher proportion of women than men at risk for PTSD would not be receiving an early opportunity for clinical care.

Lower specificity, on the other hand, results in a higher false positive rate (1 – specificity). An elevated false positive rate for females on the PC-PTSD means that a greater proportion of women who lack the risk for later PTSD would be erroneously identified by the screen as being at risk. Misidentification of this type would lead to a higher proportion of females than males being unnecessarily referred for more extensive assessment. This could disproportionately pull women off of normal duty, and possibly, place unneeded stress on limited military medical and mental health resources.

Given the evidence for gender PTSD differences and the wide use of the PC-PTSD with both male and female veterans, it is important to ensure that this instrument is measuring PTSD equivalently across genders. If measurement equivalence does not hold, direct comparison of

scores between gender subgroups is rendered impossible since the relationship between the PTSD construct and the observed score differs for men and women (Drasgow & Kanfer, 1985). In the case of the PC-PTSD screen, for example, the same cut score is currently used to signal elevated PTSD risk regardless of gender. If measurement non-equivalence is found within this instrument, a possible implication is that one might need to use different cut scores to interpret elevated PTSD risk when assessing females as opposed to males. Little empirical research has been accomplished in this area with any PTSD instrument. Indeed, in a recent review of PTSD gender differences, Kimerling, Ouimette, and Weitlauf (2007) called for further research to examine whether PTSD assessment instruments exhibit measurement invariance with respect to gender.

Purpose of the Study

The purpose of this study is to examine the validity of PC-PTSD scores for use in identifying probable PTSD with both male and female military members. Initial tests will assess measurement equivalence of the PC-PTSD across genders. This will be accomplished using item response theory (IRT) to test for differential functioning of individual items when comparing male and female subgroups.

Second, this study will examine the utility of the PC-PTSD for predicting a subsequent PTSD diagnosis. This will be accomplished through use of both standard diagnostic accuracy statistics and logistic regression modeling. Concurrent validity will be examined by relating the PC-PTSD cutoff to the presence or absence of a clinical diagnosis of PTSD. Similarly, divergent validity will be assessed by linking the cutoff to a diagnosis of major depression. Interactions for

gender will be tested. Finally, the concurrent and divergent validity of the PC-PTSD will be evaluated using confirmatory factor analysis methods.

CHAPTER II: LITERATURE REVIEW

As mentioned above, current evidence is that women are more likely than men to develop PTSD. A number of potential explanations have been proposed for this disparity. These include the neurophysiological differences between the sexes (Nemeroff et al., 2006; Saxe & Wolfe, 1999; Simmons, 2007), social factors such as gender role expectations or the social milieu during and after the trauma (Kimerling et al., 2007; Nemeroff et al., 2006; Saxe & Wolfe, 1999; Simmons, 2007), methodological gender bias (Kimerling et al., 2007; Simmons, 2007; Tolin & Foa, 2006), and differing male and female exposure rates to particular types of trauma (Kimerling et al., 2007; Nemeroff et al., 2006; Saxe & Wolfe, 1999; Simmons, 2007; Tolin & Foa, 2006).

The proposed study will focus on the measurement aspect of possible methodological gender bias. However, in order to provide context for the importance of this topic, the discussion will begin with a history of the largely male-centered development of the modern diagnosis of PTSD. This will be followed by a review of the epidemiological evidence regarding the prevalence of PTSD among males and females and differences in both exposure to trauma and the associated reactions to that trauma. Next, burgeoning findings on the neurobiology of PTSD will be covered as they relate to sex differences within the diagnosis. The review will then shift focus to gender difference evidence in PTSD symptoms. This section will provide background directly related to the structure of the PC-PTSD: four questions assessing for the presence of PTSD symptoms nested within one overarching query about trauma exposure and the trauma's initial emotional impact. Recommendations will then be presented for further research among

the reviewed areas of knowledge, followed by a review of published validity studies on the PTSD instrument of interest in this study, the PC-PTSD. Finally, specific research questions will be presented.

History of PTSD

The earliest accounts of problems that are now associated with PTSD involved primarily adult males exposed to combat. However, this was not exclusively the case. As research gradually broadened into other groups and types of trauma, this expanded knowledge allowed the experiences of other groups, such as women and children, to contribute to the development of the PTSD diagnosis.

Early Conceptions

According to Rosen (1975), one of the very early descriptions of problematic reactions to a traumatic situation came from the dissertation of a physician named Johannes Hofer in 1678. Hofer coined the term "nostalgia" to label an illness witnessed among despairing soldiers who had been uprooted from their homes or forced to fight in roving, illegitimate armies. Rosen (1975) cites historical accounts which indicate that some men suffering from nostalgia eventually became so indifferent to their circumstances that they lost interest in basic, life-sustaining activities. Following the documentation of nostalgia, however, for nearly two hundred years the scholarly record remained largely silent about any new conceptualizations regarding trauma-related problems. This would change with the appearance of "irritable heart" in the mid-nineteenth century.

Originally reported during the United States Civil War (1861-1865) and recognized in contemporary British soldiers of that era, irritable heart was characterized by the following symptoms: pain in the chest, sweating, heart palpitations, fatigue, shortness of breath, and an inability to increase one's heart rate during physical exertion (Moore & Reger, 2007; Wooley, 2002). With the inclusion of numerous physiological reactions, irritable heart appears to have been more closely related to the current construct of PTSD than was nostalgia. Physicians at the time theorized that the disorder resulted from exposure to combat (Saigh & Bremner, 1999).

In the late 19th century, early psychological theories of trauma began to take shape. Pierre Janet, while studying so-called "hysteria" with Sigmund Freud, developed the belief that severely stressful life events could overwhelm the adaptive capacity of the mind. This could lead to "dissociation" of the traumatic memories. Foreshadowing the current conceptualization of PTSD as a delayed reaction, Janet theorized that these intense memories could continue to cause problems for the sufferer long after the actual traumatic event (Flora, 2002; van der Kolk et al., 2002). At around the same time, Kraepelin described "fright neuroses" resulting from catastrophes such as major fires and railway accidents (Saigh & Bremner, 1999). Interestingly, Freud's original, published understanding of hysteria was similar to Janet's in that intrusive traumatic memories came from intensely stressful situations that had been compartmentalized away from the conscious mind. He posited that hysteria was the consequence of sexual seduction of the child by an adult. However, Freud subsequently revised his view and supplanted this theory with one that saw hysteria as the result of childhood sexual wishes. As

Freud's psychoanalytic view of mental illness grew internationally in influence, Janet's theory about traumatic memory was gradually forgotten (Flora, 2002).

Early to Mid-Twentieth Century

While irritable heart resurfaced during World War I (1914-1918) under the revised term, "soldier's heart," "shell shock" became the more overriding concern. Soldiers suffering from shell shock were typically easily startled, agitated, complained of fatigue, had difficulty concentrating, and exhibited mood swings. Another common presentation was the conversion reaction, characterized by what looked like neurological loss of function in specific parts of the body (Moore & Reger, 2007).

During the intervening years between the world wars, an important paper focused on the plight of sexually abused children. This very early publication by Bender and Blau (1937) described reactions such as hypervigilance, avoidance, nightmares, and reliving the trauma. These sequelae of child sexual abuse are strikingly similar to the general constellation of PTSD as defined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, (DSM-IV; American Psychiatric Association, 1994; Saigh & Bremner, 1999).

In the Second World War (1939-1945), military medical personnel commonly used the term "battle fatigue" to describe a syndrome similar to that denoted by shell shock. Although sharing similar symptoms with shell shock such as loss of function, anxiety, difficulty concentrating, and fatigue, battle fatigue also often involved depression, lack of motivation, and memory loss (Moore & Reger, 2007). Since the United States (US) Army continues to use the term "battle fatigue" in connection with combat stress today, a crucial distinction must be made:

Battle fatigue and PTSD are not identical (Moore & Reger, 2007). By definition, PTSD is a delayed stress reaction while battle fatigue need not be so. However, battle fatigue shares much in common with the modern diagnosis of acute stress disorder (ASD) which has been shown to be strongly associated with later development of PTSD (Kutz & Dekel, 2006).

The progression toward the modern conceptualization of PTSD continued during and after World War II with widespread treatment of veterans suffering from "combat-" or "war neuroses" (Flora, 2002; Saigh & Bremner, 1999). Kardiner (1941) published a work that has been recognized as one of the first characterizations of war neurosis that resembles PTSD (Flora, 2002; van der Kolk et al., 2002). He described a physiological component of hyperarousal and proposed that the individual was behaving as though the threat of the trauma remained present (van der Kolk et al., 2002). Kardiner also identified chronic problems with irritability, being easily startled, reduced interest, and a pattern of dreaming that was typical of war neurosis (Flora, 2002).

First Diagnosis to Modern PTSD

In the post-World War II years, the scale of the numbers of former soldiers suffering from war neuroses coupled with mounting information on the psychological effects of exposure to concentration camps and prisoner of war camps led the American Psychiatric Association (APA) to develop a new category for these problems (Saigh & Bremner, 1999). The DSM-I (1952) introduced "gross stress reaction." Although the APA did not provide specific criteria to be met for such a diagnosis, it acknowledged the severe mental distress that could result in

otherwise normal individuals who had been exposed to stressors such as warfare or serious catastrophe.

The Korean conflict of the 1950's provided further evidence of the PTSD-like symptoms exhibited by many veterans exposed to combat. This time period and the subsequent 1960's were especially significant, however, for the surge in research regarding trauma suffered during civilian accidents and disasters. Interestingly, due to the Cold War much of this work on the psychological and physiological effects of floods, earthquakes, and manmade tragedies was pursued and funded in an effort to better understand the mental consequences of a nuclear attack on a populace (Saigh & Bremner, 1999). In the end, such civilian disaster studies largely confirmed reactions such as avoidance, irritability, nightmares, and reexperiencing that had become familiar to combat trauma researchers and clinicians.

Later in the 1960's and into the 1970's, Vietnam veterans began to return—many suffering from persistent problems with war trauma. However, the diagnosis of problems related to traumatic events remained difficult due to a continuing lack of diagnostic criteria specific to this clinical issue in the new DSM-II (American Psychiatric Association, 1968; Flora, 2002). Instead, the diagnosis currently recognized as PTSD was increasingly unofficially recognized and discussed as a result of the large number of troubled veterans seeking treatment for similar problems at the nation's VA hospitals (Flora, 2002; Saigh & Bremner, 1999).

Around the same time, several researchers finally began to examine women's reactions to traumatic events on a significant scale by turning their attention to the effects of sexual assault. Burgess and Holmstrom (1974) studied women who were rape victims and identified a longer-

term phase of their "rape trauma syndrome" that included fear, avoidance, sexual problems, and aversive thoughts and dreams related to the assault. Other researchers began investigating females' reactions to rape and again reported a symptom pattern similar to that found in other types of trauma (Saigh & Bremner, 1999). Thus, sexual assault was added to a growing list of seemingly disparate traumatic events that each appeared strongly linked to a relatively uniform symptom pattern.

Finally, after the "social epidemic" (Moore & Reger, 2007, p. 165) of troubled Vietnam veterans, PTSD was formally recognized in 1980's DSM-III (American Psychiatric Association). It had become evident to many experts in the burgeoning field of trauma that widely varied events could produce a similar constellation of symptoms across diverse populations. The term PTSD now united the various unique labels previously associated with particular types of trauma. The DSM-III also contained specific diagnostic criteria divided into four categories: extremely stressful event, reexperiencing symptoms, numbing symptoms, and a category containing mostly physiological symptoms. While there have been important changes to these criteria in subsequent DSM editions and revisions, these broad categories of diagnostic criteria remain largely intact (Saigh & Bremner, 1999).

This historical review demonstrates that the impetus for the development of our current PTSD diagnosis was primarily rooted in concern over males suffering symptoms subsequent to combat exposure. While research into the effects of trauma other than combat and experienced by groups besides adult males was initially less voluminous, such knowledge eventually influenced the debate over the PTSD diagnosis in a significant way. In fact, as noted above,

researchers largely recognized that the symptoms represented in PTSD applied to a wide range of traumatic experiences and populations. Indeed, an analysis of data for the DSM-IV field trials (Kilpatrick, Resnick, Saunders, & Best, 1998) revealed no significant differences in the way males and females with PTSD responded to individual PTSD symptoms on the Structured Clinical Interview for DSM-IV (Spitzer, Williams, Gibbon, & First, 1995).

Shortly after PTSD's acceptance as a legitimate diagnosis, work began to establish the prevalence of this new disorder in various populations. The discussion now turns to a review of the epidemiology of PTSD from the perspective of gender.

Gender Differences in PTSD

Epidemiological studies provide the basis for the first two parts of this section: gender differences in PTSD prevalence and exposure rates to potentially traumatizing events. Next, male/female differences in the immediate emotional reactions to trauma exposure will be examined. The section will conclude with coverage of research into differential patterns of PTSD symptom reporting among males and females.

Higher Female PTSD Risk and Prevalence

Several epidemiological studies have reported elevated PTSD rates among women. Notable are those studies using probability sampling methods to attain large, representative samples and assessed with a standardized structured diagnostic interview. Among these, the results from the National Comorbidity Study (Kessler et al., 1995) are frequently cited. This report stated that lifetime PTSD prevalence was 10.4% for women and 5.0% for men. Helzer, Robins, and McEvoy (1987) discovered much lower lifetime prevalence levels in the

Epidemiologic Catchment Area Survey, but again females' PTSD rates (.13%) were more than twice as high as that observed with males (.5%).

Similar findings emerged when researchers assessed for lifetime prevalence only for those exposed to at least one traumatizing event. Kessler and colleagues (1995) analyzed a subset of their sample who had experienced at least one such event and found that trauma-exposed females were significantly more likely to have PTSD than trauma-exposed males (*odds ratio* = 6.13, $p < .05$). Using similar methodology among young adults in the Detroit, MI, vicinity, Breslau et al. (1998) reported the prevalence of PTSD after trauma at 13.0% for females and 6.2% for males.

The pattern of higher female PTSD rates has largely held within international populations, as well. Recent epidemiological work in Mexico (Norris et al., 2003) revealed that females' PTSD lifetime prevalence was also about twice that of males (15% for women, 7% for men). Peters, Issakidis, Slade, and Andrews (2006) assessed the twelve-month prevalence of PTSD among Australian adults. They also discovered a nearly 2:1 ratio of female to male PTSD (4.2% and 2.3%, respectively).

A couple of Canadian studies (Stein, Walker, & Forde, 2000; Stein et al., 1997), using a PTSD self-report instrument instead of structured interviews, found elevated risk for women. Results of Stein et al. (1997) indicated that women were significantly more likely than men to currently have both full PTSD (5.0% vs. 1.7%, respectively) and partial PTSD (5.7% vs. 2.2%). In Stein and colleagues (2000), 8.2% of women versus 1.8% of men were judged by the instrument to have either partial or full PTSD. This difference constituted an odds ratio (*OR*) of

4.79, meaning that the odds of having full or partial PTSD were almost five times higher for women than for men. In both Canadian studies, partial PTSD meant not meeting the full Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association, 1994) criteria, but having at least one symptom in each required symptom category.

Similar results have been observed when the focus shifts to youth. The final report from the National Survey of Adolescents (Kilpatrick and Saunders, 1997) indicated sex differences in prevalence levels for both current PTSD (6.2% for girls, 3.7% for boys) and lifetime PTSD (10.1% and 6.2%, respectively). Also, Breslau, Davis, Andreski, Peterson, and Schultz (1997) found that when they analyzed a subset of their sample who had experienced an initial trauma when they were fifteen or younger, there was a marked disparity between female and male PTSD risk (hazards ratio = 11.4). Costello et al. (1996), on the other hand, failed to find a significant difference in a general population sample of children.

The evidence is somewhat ambiguous, however, with large studies of military samples. In the report on the congressionally-mandated National Vietnam Veterans Readjustment Study, Kulka et al. (1990) indicated that men, not women, had a much higher prevalence of current PTSD. In fact, in this case the male prevalence (15.2%) was nearly twice that of females (8.5%). The National Vietnam Veterans Readjustment Study used a probability sample of 1200 men and women who had served in Vietnam at some time between 1964 and 1975 (women and some minorities were oversampled). PTSD was assessed using the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988).

Turner, Turse, and Dohrenwend (2007) recently reanalyzed the National Vietnam Veterans Readjustment Study data to investigate possible explanations for the prevalence findings in this sample that are so contradictory to the results of most comparable civilian studies. Two relevant findings emerged: First, they found that when they controlled for the pre-war PTSD risk factors minority status, education level, and age, the prevalence difference (while not statistically significant) actually changed directions, with female Vietnam veterans at 4% higher prevalence than male veterans with the lowest combat exposure. Second, when the researchers endeavored to compare males and females with similar degrees of exposure to both medical trauma and combat trauma, no gender-related difference in PTSD prevalence was found.

More recently, Smith and colleagues (2008) analyzed a very large (N = 50,184; 27.6% female) population-based, US military sample from the prospective millennium cohort study. Although the final sample was not random, it included active duty members from all the services, as well as members of the reserves and national guard—none of whom had been deployed to Iraq or Afghanistan prior to joining the larger study. After being sampled for Smith et al. (2008), 23.8% of the participants deployed to Iraq or Afghanistan (26.8% of men, 16.1% of women) during the study period.

The researchers reported that among those subjects with no PTSD symptoms or diagnosis at baseline, 4.7% of females and 3.1% of males experienced a new PTSD diagnosis or were positive for PTSD symptoms within the previous three years as measured by sensitivity-maximizing criteria of the PTSD checklist (PCL; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996). Using criteria from the same instrument that instead maximized specificity

yielded similar results, with PTSD prevalence at 3.8% for women and 2.4% for males. The authors also looked at new onset PTSD (judged using the specific PCL criteria or PTSD diagnosis) by gender for each service. All services but the Marines had a significant OR, indicating that females were significantly more likely to be classified as having new onset PTSD. Of interest for the current study are the results for the Air Force: New onset PTSD was judged in 1.0% of males and 2.0% of females (*OR* = 2.00, 95% *confidence interval*: 1.41 to 2.83).

Indeed, two meta-analytic studies (Brewin, Andrews, & Valentine, 2000b; Tolin & Foa, 2006) on the subject have confirmed women's higher risk for PTSD. Both meta-analyses assessed a mix of veteran and civilian samples. The analysis by Brewin and colleagues (2000b) indicated a relatively small, but significant effect size for gender on PTSD ($r = .13$, $p < .001$). Tolin & Foa's (2006) work revealed a stronger mean effect (*OR* = 1.98, $p < .001$; meaning females had nearly twice the odds of PTSD than men) across all independent samples. In fact, they found that when the gender effect was compared between variables related to methodological differences (e.g., lifetime or current PTSD, interview or questionnaire, epidemiological or convenience sample), none of the methodological factors resulted in a significant change to the effect size indicating women's greater PTSD risk.

It appears that the evidence points to increased risk of PTSD for women when compared to men. Although there is some ambiguity on this among US military and veteran samples, recent work has tended to support the differential seen in the more numerous, epidemiologically-based civilian studies. The discussion will now turn to gender differences in exposure to traumatic events.

Trauma Exposure Differences

As noted above, evidence indicates that females have somewhat lower exposure to potentially traumatic events (PTEs) despite their comparatively higher risk for PTSD. Some of the epidemiological studies mentioned previously have examined this issue. Kessler et al. (1995), Norris et al. (2003), and Stein et al. (1997) all found that a significantly higher percentage of males than females reported experiencing at least one PTE during their lifetime. In the National Comorbidity Study (Kessler et al., 1995) these figures were 60.7% for males and 51.2% for females. Two probability-sampled epidemiological studies using the liberalized DSM-IV criterion for a qualifying event indicated somewhat higher PTE prevalence: Stein and colleagues (1997) reported male lifetime exposure at 81.3% compared to 74.2% for females, and Norris et al. (2003) more recently found comparable figures at 83% and 71% for men and women, respectively.

On average, men also tend to report a higher number of lifetime PTEs than women. Breslau et al. (1998) found that males endorsed an average of 5.3 lifetime traumatic events, compared to 4.3 among females. The averages reported by Norris et al. (2003) were lower (2.6 events for men, 1.7 events for women), but the difference was likewise significant. However, given the unique characteristics of various trauma categories, these overall PTE figures are very gross measures of trauma exposure level. Therefore, the remainder of this section will focus on the research into differential gender exposure to specific types of PTE.

The one large epidemiological study using a representative probability sample of the entire United States (Kessler et al., 1995) reported that males were significantly more likely than

females to identify themselves as experiencing PTE such as combat (6.4% vs. 0.0%, respectively), physical attack (11.1% vs. 6.9%), life-threatening accident (25.0% vs. 13.8%), being personally threatened with a weapon (19.0% vs. 6.8%), and witnessing a serious injury or killing (35.6% vs. 14.5%). On the other hand, females more frequently endorsed rape (9.2% of females vs. 0.7% of males), being sexually molested (12.3% vs. 2.8%), childhood physical abuse (4.8% vs. 3.2%), and parental neglect (3.4% vs. 2.1%).

The Breslau et al. (1998) study of a Detroit-area random sample revealed a similar pattern of differences: Men exhibited heightened risk of exposure to assaultive violence other than rape, such as being physically beaten up (males = 13.1%, females = 9.8%), being either shot or stabbed (males = 8.2%, females = 1.8%), and being threatened with a weapon or mugged (males = 34.0%, females = 16.4%). Women were more at risk for rape (females = 9.4%, males = 1.1%) and other types of sexual assault (females = 9.4%, males = 1.8%). An analysis of the Great Smoky Mountains Study (N = 1420) of randomly selected children and adolescents by Costello and colleagues (2002) revealed that boys were significantly more likely to have a lifetime history of violent physical victimization by a non-family member (0.8% of boys, compared to 0.4% for girls) and causing death or serious harm to others (0.2% of boys, compared to less than 0.1% for girls). Girls exhibited heightened incidence of rape (0.4% of girls, compared to less than 0.01% for boys), sexual abuse (3.4% of girls, compared to 1.8% for boys), and violent physical victimization by a family member (1.2% of girls, compared to 0.5% for boys).

These epidemiological PTE exposure differences mirror findings from outside of the United States, as well. In Mexico (Norris et al., 2003), females were again shown to be more likely than males to identify a history of sexual assault (3.9% vs. 1.1% for women and men, respectively). Males, on the other hand, were disproportionately exposed to combat (3.2% for men vs. 1.1% for women), acts of terror or torture (1.1% vs. 0.3%), being physically assaulted (27.8% vs. 13.5%), experiencing an accident that threatens life (45.1% vs. 21.9%), being threatened with a weapon (28.3% vs. 8.3%), and witnessing someone else being seriously injured or killed (45.6% vs. 26.3%). Stein and colleagues' (1997) Canadian sample showed a similar pattern. Men more often experienced combat, severe MVAs, being personally threatened with a weapon, and witnessing someone else's serious injury or death, but women more frequently endorsed experiencing rape and sexual molestation (all reported as significant, although no specific statistics provided).

A couple of studies have examined trauma exposure with convenience samples containing at least a substantial proportion of individuals affiliated with the military. Bolton, Litz, Britt, Adler, and Roemer (2001) surveyed a highly male (90%) convenience sample of 2947 (mean age = 26) US Army personnel about to deploy to Bosnia-Herzegovina for peacekeeping duty. Once more, females more often reported sexual assault than males (27% and 5%, respectively) and males more often reported witnessing an event involving someone else's serious injury or illness (53% of males, 46% of females). However, the researchers used an unstandardized self-report instrument created specifically for this study which used a smaller

number of broad PTE categories (5) compared to most studies that have investigated this issue (typically, 10+ categories of PTE that are more specific in scope).

Using the Trauma Event Questionnaire (Vrana & Lauterbach, 1994) to assess PTE, Amir and Sol (1999) assessed for trauma in a non-random undergraduate sample within which 17.5% of the females and 23.8% of the males had previously served as Israeli army officers. Women were more exposed to rape and sexual assault (combined: 10% of females vs. 1% of males), while men were more exposed to civilian MVAs (34% of men vs. 27% of women) and military-related trauma connected to the Intifada (42% of men vs. 3% of women), the war in Lebanon (5% of men vs. 1% of women), terror attacks while in the military (23% of men vs. 3% of women), and other military operations (45% of men vs. 3% of women). It was not clear from the article what types of jobs the subjects held while in the military.

To sum up, the weight of the evidence for differential exposure to trauma indicates that women more frequently report sexual trauma such as rape and sexual assault, while men more frequently endorse a history of experiencing combat, accidents, and non-sexual violence. The meta-analysis by Tolin and Foa (2006) confirmed this pattern with the greatest significant differences in adult sexual assault, child sexual abuse, and a category including combat, war, and terrorism experiences. It is tempting to ascribe females' higher incidence of PTSD to these well-supported PTE differences. However, the same meta-analysis (Tolin & Foa, 2006) revealed that the PTSD gender disparity persisted even after controlling for the type of exposure that was experienced. The discussion now turns to neurobiological issues related to gender and PTSD.

Neurophysiology and the PTSD Gender Difference

Researchers' knowledge of the body's stress response has grown dramatically over the last decade. This section will begin with a summary of some major physiological aspects of stress and PTSD. Then, gender differences with regard to these physiological processes will be explored. Finally, the discussion will turn to how such differences may affect one's vulnerability to stress disorders later in life.

Hypothalamic-Pituitary-Adrenal (HPA) Axis

Overview of the HPA axis. Among the most heavily investigated biological mechanisms associated with PTSD and other stress responses is the HPA axis. This system is comprised of the hypothalamus, the pituitary gland, and the adrenal glands. Under normal circumstances these three structures act as a feedback loop to regulate the body's release of cortisol and other glucocorticoids that, in turn, help the body prepare to deal with stress. Among other functions, cortisol facilitates use of the body's energy stores, enhances attention and the encoding of memory, and curbs the sympathetic stress reaction by serving as an "antistress" (Yehuda, 2001, p. 41) hormone (Charney, 2004; Rasmusson & Friedman, 2002; Yehuda, 2009).

The cascade begins when neurotransmitters from other parts of the brain signal the hypothalamus to increase its release of corticotrophin-releasing factor (CRF). When the pituitary gland receives CRF, it is stimulated to secrete adrenocorticotropic hormone (ACTH) into the bloodstream. Circulation then transports ACTH to the adrenal gland where its arrival causes the adrenals to release glucocorticoids, including cortisol, into the blood. Negative feedback is

accomplished in the loop when glucocorticoids binding to glucocorticoid receptors in the hypothalamus and hippocampus exert an inhibitory effect on further CRF secretion.

Ideally, the balance achieved in this system allows the body to adequately prepare for a threat while minimizing the detrimental physiological effects of chronically high levels of cortisol and other glucocorticoids (Charney, 2004; Meyer & Quenzer, 2005). However, the HPA system may not immediately return to normal, baseline levels despite cortisol's negative feedback inhibition. This breakdown, in part, may result when the amygdala perceives the presence of a threat and consequently continues sending signals to stimulate HPA activity. But once the stressor is no longer detected and the amygdala ceases its excitatory signaling, the glucocorticoids can have their full inhibitory effect on the HPA and the hippocampus to rein in the body's stress response (Yehuda, 2001).

PTSD and the HPA axis. PTSD has been associated with alterations in HPA functioning. However, with regard to cortisol, there is disagreement as to the direction of this association (Johnson, Delahanty, & Pinna, 2008). Some studies have reported that individuals with PTSD tend to have lower levels of cortisol (e.g., Bremner, Vermetten, & Kelley, 2007; Gill, Vythilingam, & Page, 2008; Kanter et al., 2001; Yehuda, Boisoneau, Lowy, & Giller, 1995; Yehuda, Boisoneau, Mason, & Giller, 1993; Yehuda, Teicher, Trestman, Levengood, & Siever, 1996). Others (e.g., Lemieux & Coe, 1995; Pitman & Orr, 1990; Rasmusson et al., 2001) instead found that cortisol levels for those with PTSD were higher than for controls. Still others have found no significant association between PTSD and cortisol (Young & Breslau, 2004a, 2004b) or a very small negative relationship despite a large sample size (Boscarino, 1996).

Several possible explanations have been suggested for these disparate findings, including characteristics of the trauma (e.g., severity, frequency, and time elapsed since the trauma), comorbidity, sex, menopausal factors for females (Johnson et al., 2008), and smoking status (Rasmusson et al., 2001). Others have focused on the diverse methods used to measure cortisol. The 24-hour urinary free cortisol test has been criticized for missing brief, periodic changes in cortisol levels that are associated with circadian rhythms (Yehuda, 2009). On the other hand, methodologies utilizing plasma cortisol levels obtained by extracting blood samples every 30-60 minutes during a twenty-four hour period (e.g., Yehuda et al., 1996) are subject to the criticism that being hospitalized or removed from daily life itself could alter the subjects' cortisol levels due to changes in their perceived stress (Rasmusson et al., 2001). Salivary cortisol levels have the advantages of being less invasive and easily performed in a more naturalistic stress setting such as at home (Young & Breslau, 2004b). However, unless the timing of the sampling is adequately controlled they may obscure cortisol differences that tend to only depart from normal levels at certain times of the day (Yehuda, 2009).

Possible mechanisms for the HPA axis role in PTSD. Competing arguments have been made as to what conclusions can be drawn from these seemingly divergent findings on the relationship between PTSD and cortisol levels. Rasmusson, Vythilingam, and Morgan (2003) posit that the case is stronger for an upregulation of HPA activity that results in hypercortisolism. They point out that in many of the studies connecting lower cortisol with PTSD that used urinary cortisol measures (e.g., Yehuda et al., 1993; Yehuda et al., 1995), subjects were confined to either a hospital or home setting which could have resulted in lower stress and thus lower cortisol

levels. Indeed, these authors present a number of reports (e.g., Lemieux & Coe, 1995; Pitman & Orr, 1990; Rasmusson et al., 2001) as evidence that studies using similar cortisol measurement techniques with individuals maintaining more normal, non-sequestered daily routines have typically resulted in higher cortisol for those with PTSD. Further, they point out that such hypercortisolism is consistent with studies that report elevated CRF levels among those with PTSD (e.g., Baker et al., 1999).

Yehuda (2001, 2009) alternatively argues that some of the studies indicating higher cortisol suffered from problems measuring cortisol that call their results into question. She asserts that the majority of studies that used frequent, regular measures of cortisol throughout the day were better at capturing the rhythmic nature of cortisol excretion—and that such investigations tended to show a pattern of lower cortisol in conjunction with PTSD when compared to controls. This researcher reconciles the elevated CRF findings with those revealing lower cortisol levels by citing evidence that glucocorticoid receptors are sensitized in PTSD (Yehuda et al., 1995), which she hypothesizes would result in stronger negative feedback in the HPA system.

Many studies using the dexamethasone suppression test (DST) appear to support this hypothesis (e.g., Yehuda et al, 1993; Yehuda et al., 1995; Griffin, Resick, & Yehuda, 2005). In DST, a small dose of dexamethasone (a synthetically-produced substance with pharmacological action that is very similar to cortisol) is administered in order to test the extent to which endogenous cortisol is affected (Yehuda, 2001). Following DST, studies with various PTSD populations have reported resulting cortisol to be lower than controls—which is consistent with

sensitized glucocorticoid receptors enhancing the negative feedback inhibition of the HPA system.

It appears the bulk of the evidence indicates that PTSD is associated with some form of cortisol dysregulation within the HPA axis. However, until more well-controlled studies with uniform cortisol measurement methods are employed and replicated with similar results, the exact nature of this dysregulation will remain debatable. The discussion now turns to another neurochemical involved in the body's stress response: norepinephrine.

Norepinephrine

Overview of norepinephrine. Norepinephrine (NE) is a catecholamine that has been associated with arousal via the locus coeruleus (LC), memory of stressful events through the amygdala, and cognition through its involvement in the prefrontal cortex (PFC) (Charney, 2004; Southwick, Rasmusson, Barron, & Arnsten, 2005). Specifically, stress activates NE in the LC (Charney, 2004), which contributes to stimulation of the sympathetic nervous system and the enhancement of vigilance (Southwick et al., 1999). In turn, activated NE connections between the LC and the amygdala appear to facilitate enhanced consolidation of memories associated with the acute stress (Charney, 2004). The amygdala also sends signals on to the PFC, among other structures. Under these stressful circumstances, elevated NE interferes with PFC functions. This results in a diminishing of higher cognition (Charney, 2004; Southwick et al., 1999).

PTSD and NE. Clinical studies of combat veterans (Kosten, Mason, Giller, Ostroff, & Harkness, 1987; Yehuda et al., 1998) and victims of child sexual abuse (De Bellis, Baum, Birmaher, & Ryan, 1997; Lemieux & Coe, 1995) have generally shown that subjects with PTSD

have significantly elevated baseline NE when compared to controls. Others have stimulated NE release by administering yohimbine, an antagonist of the α_2 adrenergic receptors which are autoreceptors on noradrenergic neurons. Both Southwick and colleagues (1993) and Bremner et al. (1997a) found that administration of yohimbine resulted in most (70% and 60% in each study, respectively) PTSD subjects experiencing panic attacks and a substantial portion experiencing flashbacks (40% and 30%, respectively). This contrasted with no panic attacks among controls in either study, zero flashbacks in the Bremner et al. (1997a) control group, and only one flashback among controls in the Southwick et al. (1993) study. The latter study (Southwick et al., 1993) also reported a significant post-yohimbine increase in symptoms related to PTSD for those with PTSD compared to controls. These results appear to also lend support to the case for hyper-release of NE among those with PTSD.

Possible mechanisms for NE's role in PTSD. Indications are that increased NE would enhance fear-related arousal and attention mediated by the LC due to NE's stimulating effect on that structure. A relatively high level of NE also tends to aid in the encoding of fearful or stressful memories through the amygdala (Southwick et al., 2005). A recent study (van Stegeren et al., 2005) using functional magnetic resonance imaging (fMRI) and blocking NE activity with propranolol provided some support for NE's role in mediating amygdala activity related to memory formation. Those receiving placebo (and therefore experiencing a relatively normal NE level) during exposure to stress-provoking visual images showed significantly greater amygdala activation than subjects receiving propranolol. When memory was tested at two weeks, the placebo also resulted in a larger difference between the memory of stressful vs. non-stressful

images when compared to the propranolol condition. This implies that the effect of the propranolol's blocking of NE activity was to inhibit the normal, enhanced memory formation for the stressful images.

Also, a couple of clinical studies (Yehuda, Southwick, Giller, Xiaowan, & Mason, 1992; Lemieux & Coe, 1995) have explicitly linked elevated NE to increased endorsement of intrusive memories on the Impact of Events Scale (IES; Horowitz, Wilner, & Alvarez, 1979). One hypothesis for this result is that high levels of catecholamines such as NE contribute to the intrusion of these traumatic memories by their interference with the PFC. This interference might reduce the capacity of the PFC to inhibit such thoughts (Southwick et al., 1999).

Neuropeptide Y

Overview of neuropeptide Y. Neuropeptide Y (NPY) is an amino acid peptide commonly found throughout the mammalian nervous system. Although it has not been studied as extensively as the catecholamines or the neurochemicals directly related to the HPA axis, evidence suggests that it does play an important role in the human stress response. NPY is secreted in conjunction with high-level activation of the sympathetic nervous system (Rasmusson & Friedman, 2002), may be involved with inhibiting the consolidation of fear memories through both the amygdala and hippocampus, and appears to provide an anti-anxiety balance to CRF in the body's stress response system via the LC and the amygdala (Charney, 2004).

Some initial evidence seems to indicate that this neurotransmitter also aids in the adaptive management of extreme stress levels. Morgan and colleagues (2000) studied hormone levels

among a group of soldiers participating in an intense survival training meant to approximate many types of wartime stress. The researchers reported positive correlations between NPY levels and successful behavioral performance under stress. Results also revealed negative associations between NPY and dissociative symptoms.

PTSD, NPY, and possible mechanisms. NPY has been shown to be low among subjects with PTSD (Rasmusson et al., 2000). Further, there are indications that a deficit of NPY may be linked with other conditions that have been associated with PTSD. For instance, sleep disturbance is a common comorbid problem with PTSD (Mellman, 1997) and in animal models NPY has been shown to interact with glutamate to affect circadian phase shifts in sleep (Biello, Golombek, & Harrington, 1997). Another animal study (Greber, Schwarzer, & Sperk, 1994) indicated NPY inhibits the release of glutamate, another neurotransmitter which has been implicated as a contributing factor in stress-related atrophy of the hippocampus (McEwen, 2000). When taken together with research that has associated PTSD with smaller hippocampal volume (Bremner et al., 1997b), this suggests the possibility that low NPY may contribute to PTSD-related reductions in the volume of the hippocampus. However, research has yet to explicitly confirm this hypothesized connection.

Gender Differences Related to the HPA Axis, NE, and NPY

Historically, research into the neurobiology of PTSD and stress has focused on males (Andreano, Arjomandi, & Cahill, 2008; Rasmusson & Friedman, 2002). This has begun to change in recent years, though, with increasing investigation of the role that female sex hormones and the menstrual cycle play in modifying the body's stress hormone response. The

early evidence indicates that these female-specific biological systems may interact with the HPA axis, NE, and NPY in PTSD.

HPA axis-related gender differences. A few studies have focused on estrogen and its relationship to hormones in the HPA system. In both a study of female animals (Young, Altemus, Parkison, & Shastry, 2001) and a small ($N = 12$) study of perimenopausal women (Komesaroff, Esler, & Sudhir, 1999), indications were that administering exogenous estrogen reduced ACTH and thereby cortisol under conditions of stress. On the other hand, an investigation of nine women (age range: 44-48) undergoing ovariectomies (De Leo, la Marca, Talluri, D'Antona, & Morgante, 1998) also showed reduced ACTH even though estrogen was very low post-surgery. However, in this study all subjects received the surgery and subsequent hormone testing, whereas Komesaroff et al. (1999) randomized assignment to treatment or placebo conditions. Cucinelli and colleagues (2002), although not focusing on PTSD or the stress response specifically, also found that administration of estrogen was associated with a significant ACTH decrease, this time in 20 postmenopausal women. Possible mechanisms for estrogen's effects on ACTH include changes to glucocorticoid negative feedback or alterations in the overall stimulation of the entire HPA axis (Charney, 2004).

Menstrual cycle and the HPA axis. More recently, studies have begun to compare different phases of the menstrual cycle in an attempt to elucidate the interactions between stress response hormones and female sex hormones. Goldstein et al. (2005) measured skin and brain arousal (using fMRI) in response to negative visual stimuli among twelve premenopausal women during both the early and late follicular phases of the menstrual cycle. Estrogen was higher in

the late follicular phase, which coincided with attenuation in both skin arousal and arousal in brain structures related to HPA function. Although there was no separate comparison group, the results support a role for estrogen in destimulating the HPA axis.

Other studies have investigated hormonal changes and stress-related memory. Andreano and colleagues (2008) looked at salivary cortisol and long-term memory across three distinct phases of the menstrual cycle. The study involved 64 women experiencing normal, natural cycling, alternatively exposed to warm water (control/neutral stress) or ice water (cold-pressor stress) immediately after reading a narrative passage. One week later the subjects were tested on their recall of the passage. The authors reported a positive relationship between cortisol levels and ability to recall for those originally exposed to the water conditions during the mid-luteal phase—when both estrogen and progesterone are elevated. Those subjects presenting during the early or late follicular phases showed no such correlation. This result is consistent with research showing that cortisol generally enhances memory consolidation (e.g., Abercrombie, Kalin, Thurow, Rosenkranz, & Davidson, 2003) and it informs studies reporting that the effect of cortisol on memory differs for men and women (Andreano & Cahill, 2006; Jackson, Payne, Nadel, & Jacobs, 2005). Neither of these reports showing differential gender effects on memory for glucocorticoids (Andreano and Cahill, 2006; Jackson et al., 2005) controlled for the phase of menstrual cycle among the female subjects. In light of the Andreano et al. (2008) study, it appears that cyclical fluctuations in gonadal hormones such as estrogen could have led to the reported gender differences in the effects of cortisol on memory. This suggests a role for sex

hormones in modulating the nature of women's HPA stress response—at least with respect to memory.

Fear conditioning. The precise mechanism for how the neurobiology of memory is involved in PTSD remains under investigation. However, classical fear conditioning has been proposed as one plausible method whereby memory might be connected to PTSD and other anxiety disorders (Charney, 2004; Garakani, Mathew, & Charney, 2006). Fear conditioning is theorized to involve memory-related structures such as the amygdala (numerous proposed functions, including fear acquisition and associating the unconditioned stimulus [UCS] with the conditioned stimulus [CS]), the prefrontal cortex (long-term, extinction-related memory), and the hippocampus (consolidation of fear-related memories into long-term memory) (Charney, 2004)—all of which appear to be functionally affected by cortisol (Rasmusson et al., 2001).

Milad and colleagues (2006) recently investigated whether fear conditioning itself was affected by the menstrual cycle. They studied 42 subjects (66.7% female) randomly assigned to two conditioning contexts where visual stimuli (the CS) were first paired with a mild (non-painful) electrical shock (the UCS) and then extinguished over a period of two days. Retention of the extinguishment memory was measured by skin conductance when the subjects were exposed to the CS on the second day. The researchers reported that among women in the late follicular phase of their cycle, when estrogen levels are elevated relative to the early follicular phase, retention of the extinction learning was at a significantly lower level compared to men and to women in the early follicular phase. This finding is important, because problems with

successful extinction can result in ongoing problems with anxiety (Garakani et al., 2006), such as is observed with PTSD and other anxiety disorders.

None of these studies were conducted with subjects diagnosed with PTSD, so any explicit role for sex hormones in mediating the HPA response within this disorder remains to be discovered. However, the evidence for estrogen's effects on the HPA system in healthy subjects and its impact on memory indicate that, in light of the gender differences in PTSD, further research into the interactions among these factors is warranted.

Gender differences in the NE system. A few studies have examined gender differences in NE. Luine (2002) found that when female and male rats were exposed to identical regimens of chronic, repeated restraint stress, only female rats exhibited higher levels of NE in the hippocampus. Conversely, two human studies reported elevated NE among males. Frankenhaeuser et al. (1978) compared male ($N = 19$) and female ($N = 30$) high school students exposed to either a normal (typically non-stressful) or a stressful academic situation. The analysis revealed that in the stress condition males consistently excreted higher levels of urinary NE than females. More recently, a study of 315 (65% female) healthy nursing students randomly selected from a volunteer pool (Deane, Chummun, & Prashad, 2002) also found elevated NE among males. The authors assert that the male and female subsamples were homogeneous on characteristics such as age and level of physical activity. However, no demographic statistics other than age were supplied.

Menstrual phase and NE. NE and other catecholamines also appear to be affected by menstrual phase and gonadal hormones. However, much of this evidence comes from animal

studies or research into the relationship between the catecholamines and the general physiological states (e.g., blood pressure) that have been linked to both the sympathetic stress response and risk of cardiovascular disease. For example, in a study of female sheep (Komesaroff et al., 1998) experiencing either a metabolic or an audio-visual stressor, NE response was found to be attenuated during the follicular phase relative to the luteal phase. A small, early human study (Goldstein, Levinson, & Keiser, 1983) likewise reported that plasma NE was significantly elevated in the luteal phase among six hospitalized, but healthy women who were followed throughout their ovulation cycle.

Cardiovascular risk studies. In studies investigating cardiovascular disease risk, which often assess similar physiological arousal metrics as those examined in anxiety and PTSD studies, both Girdler, Pedersen, Stern, and Light (1993) and Tersman, Collins, and Enroth (1991) looked at the effect of menstrual phase on blood pressure. The former (Girdler et al., 1993) found that blood pressure did not significantly vary between the follicular and luteal phases among the 30 subjects when exposed to either cold pressor or a speech stress. Tersman and colleagues (1991) also reported no effect of menstrual phase on blood pressure among fifteen females exposed to a mental arithmetic test. However, when those same women were given a cold pressor test, a significant blood pressure increase was detected during the luteal phase. These results contrast with Sita and Miller (1996), who looked at both menstrual phase and level of endogenous estrogen with respect to heart rate and blood pressure. Female undergraduates (ages 18-35) were subjected to multiple stress tasks in counterbalancing order, with blood samples taken after each task. Cold pressor exposure during the luteal phase was associated with

both elevated estrogen and attenuated heart rate and blood pressure. Average premenopausal estrogen levels are generally higher during the luteal phase relative to the follicular phase (Rasmusson & Friedman, 2002).

Three other cardiovascular risk studies randomized women to either estrogen or placebo administration, presented a stressful stimulus, and measured any resulting catecholamine response. Ceresini et al. (2000) subjected postmenopausal women ($N = 10$; mean age = 53.9 years) to mental stress using an arithmetic test and found no significant difference in plasma NE between the estrogen and placebo conditions. They did, however, find that estrogen administration was associated with a significant decrease in the related catecholamine epinephrine. In the Komesaroff et al. (1999) study of perimenopausal females ($N = 12$), those receiving estrogen showed a significantly reduced NE response after stress from a mental arithmetic task compared to placebo. Sofowara, Singh, He, Wood, and Stein (2005) detected a significantly increased NE response to mental stress among 19 postmenopausal subjects (mean age = 54) for the estrogen condition relative to control, but no NE difference after cold pressor stress. However, these researchers only administered estrogen for 36 hours prior to stress exposure compared to 3 weeks for Ceresini et al. (2000) and 8 weeks for Komesaroff et al. (1999).

There appears to be conflicting evidence regarding estrogen's relationship to the sympathetic response from studies linking cardiovascular reactivity and menstrual phase. Perhaps this ambiguity might be mitigated with increasing use of newer, more discriminatory NE measures such as microneurography (Rasmusson & Friedman, 2002). However, considering

these varying results together with NE's association with PTSD, it seems reasonable that fluctuating gonadal hormones in women may be a contributing factor to the gender difference in PTSD.

Gender differences related to NPY. Research into any NPY gender differences are in the very beginning stages. Zukowska-Grojec (1995) reported that when ovariectomized women were compared to age-matched men, females exhibited a lower plasma NPY increase than males in response to exercise stress. When the women later received estrogen replacement therapy, their stress-induced NPY increase disappeared--suggesting that estrogen inhibits NPY. More recently, Karl, Duffy, and Herzog (2008) studied male and female mice that were NPY-deficient. Both sexes showed increased anxiety-like behavior in locomotion and exploration tasks, but this effect was stronger among the male mice. This could mean that when an NPY deficiency exists, males are more detrimentally affected than females. However, more human studies are necessary to know whether such a distinction will hold between men and women.

There is also evidence that menstrual cycle changes in gonadal hormones may affect NPY. While one rodent study investigating depression (Jimenez-Vasquez, Overstreet, & Mathe, 2000) failed to find any association between estrous cycle phase and NPY, a human study looking at cardiovascular disease risk did report an NPY-menstrual cycle relationship (Lewandowski et al., 1998). The latter researchers found that among eleven normally menstruating women, those in the follicular stage experienced significant increases in plasma NPY compared to basal levels when exposed to exercise stress. On the other hand, measurements indicated no significant NPY increase from basal levels for women in the luteal

phase. As discussed earlier, the average level of estrogen is higher during the luteal phase. This could mean that Lewandowski and colleagues' (1998) results are consistent with an inhibitory function of estrogen with NPY. However, another possibility is that testosterone is also helping regulate NPY levels. In rats, removal of the testes led to a reduction in NPY response to a cold pressor trial (Zukowska-Grojec, 1995). This could imply that testosterone facilitates NPY release. Indeed, Rasmusson and Friedman (2002) suggest that a facilitory role of testosterone with NPY is an alternative explanation for the Lewandowski et al. (1998) findings since testosterone levels peak in the late follicular phase immediately prior to ovulation.

With few human studies investigating sexual dimorphisms or menstrual effects on NPY and stress, it is difficult to draw firm conclusions for the interplay of these factors. Preliminary evidence indicates women have a somewhat reduced NPY response when exposed to stress and that shifts in hormone levels due to menstruation may play a role in modulating this differential response.

Implications for the PTSD gender difference. Thus far, there is a paucity of research into any direct connection between these neurophysiological gender differences and the human PTSD gender disparity. Implications must be tentatively drawn from the frequently conflicting results of the reviewed studies. With regard to gender differences in the HPA axis and cortisol, indications are that estrogen's inhibition of HPA activity may foster females experiencing periodic enhancement of stressful memories along with inhibited extinction of fear learning; both of which would be consistent with higher rates of PTSD. The NE gender difference results were highly ambiguous. However, if estrogen is eventually found to enhance the NE response to

stress, women could be at higher risk of PTSD-related hyperarousal and intrusive traumatic memories. Likewise, confirmation of estrogen's inhibitory role and/or testosterone's excitatory role with NPY would point to females having increased consolidation of fear memories and reduced attenuation of the NE-related sympathetic response.

Neurophysiological Changes Associated with Early-Life Stress

Some researchers have investigated the long-term neurophysiological consequences of trauma suffered in formative, developmental years. Such early trauma, along with other highly stressful circumstances like separation from parents, has sometimes been termed "early-life stress" (ELS; Heim, Plotsky, & Nemeroff, 2004, p. 641). Indications are that ELS is associated with longstanding, persistent changes in physiological components of the stress response.

Evidence for neurophysiological effects of ELS. Prospective animal studies with rodents (Jimenez-Vasquez, Mathe, Thomas, Riley, & Ehlers, 2001; Liu, Caldji, Sharma, Plotsky, & Meaney, 2000; Plotsky & Meaney, 1993) have shown evidence of HPA axis, NE, and NPY disruptions that persist into adulthood after exposure to ELS. Research using non-human primates (Rilling et al., 2001; Gilmer & McKinney, 2003) has similarly yielded ELS-related changes to cortisol and NE, as well as behavioral alterations. However, these animal ELS models typically involve stressors such as maternal separation or restraint stress that may not correspond directly to the subset of human ELS trauma (e.g., child sexual abuse) which would fulfill the DSM criteria for a traumatic event.

Researchers have used retrospective methods to investigate early trauma among humans. Concentrating on hypothesized long-term effects to the HPA axis, Heim, Newport, Bonsall,

Miller, and Nemeroff (2001) examined the relationship of childhood abuse-related ELS to ACTH and cortisol response in 66 women (mean age = 27.8 years). The women were categorized into four groups: 20 with no history of child abuse (comparison group), 20 with a history of child abuse but no current diagnosis of major depressive disorder (MDD), 15 with both a child abuse history and MDD, and 11 with MDD but no history of childhood abuse. Relative to the comparison group, CRF stimulation resulted in an augmented ACTH response for women with child abuse and no MDD. In contrast, the same test attenuated ACTH response in the two groups with MDD. Another test, $ACTH_{1\text{-}24}$, which allows measurement of the adrenal cortex's response to ACTH, yielded reduced plasma cortisol for the childhood abuse/no MDD group at both baseline and after ACTH stimulation. These results suggest an association between child abuse-related ELS and sensitization of the pituitary ACTH response coupled with desensitization of the ACTH receptors in the adrenal cortex.

Similarly, Shea and colleagues (2007) found that childhood abuse was correlated with lower basal cortisol levels among 66 pregnant women (mean age = 31.9 years). Child abuse experience was measured using the Childhood Trauma Questionnaire (Bernstein et al., 2003). Half of the subjects were depressed, but this relationship held even after controlling for antidepressant use status. Likewise, a study of 20 men with personality disorders reported a positive correlation between Childhood Trauma Questionnaire scores and CRF levels in the cerebrospinal fluid (Lee, Geracioti, Kasckow, & Coccaro, 2005). In a sample of women ($N = 39$, mean age = 31.2 years) with borderline personality disorder, Rinne et al. (2002) found that the 24 endorsing a history of childhood abuse exhibited an increased ACTH response to a combined

dexamethasone suppression and CRF test compared to the remainder who did not endorse child abuse. However, this study differed somewhat from the previous two in that the cortisol response for those with child abuse was also elevated, rather than reduced.

Bremner et al. (1997b) compared the hippocampal volumes of 17 PTSD patients with a history of childhood abuse to 17 healthy, non-abused subjects using MRI. The comparison group subjects were matched to the PTSD subjects on several factors including age, sex, education, etc. The PTSD/child abuse group had 12% lower volumes in the left hippocampus than the controls. Vythilingam and colleagues (2002) also used MRI to compare hippocampal volumes. Assessing 32 women with MDD and fourteen healthy women as controls, the authors reported that those MDD subjects with a self-reported history of childhood physical or sexual abuse had smaller hippocampi. Specifically, the size of the structure was reduced by 15% compared to the controls and 18% compared to women with MDD but no history of childhood abuse.

Of course, the retrospective design of these human studies makes it impossible to conclude that ELS caused the neurobiological changes that have been mentioned. However, taken together with the prospective animal studies, a reasonable case can be made that certain types of ELS result in long-term abnormalities in the body's stress response system.

Implications for the gender differential in PTSD. As already discussed in the section on differential exposure to trauma, females disproportionately indicate that they have experienced child sexual abuse when compared to males (e.g., Costello et al., 2002; Goldberg & Freyd, 2006; Perkonigg & Wittchen, 1999; Walker, Carey, Mohr, Stein, & Seedat, 2004). Given

increased ELS exposure in this significant area of trauma, females may be at somewhat increased risk for ELS-related neurophysiological changes that have been associated with vulnerability to stress disorders such as PTSD. Confirming any causal role that ELS plays in such a differential risk will be difficult, however, without long-term, prospective studies assessing ELS, stress response markers, and psychiatric diagnosis beginning in childhood and progressing into adulthood.

Summary of Neurophysiology and the PTSD Gender Difference

Separate lines of research point to the complex relationship between gender, the neurophysiology of stress, ELS, and PTSD. While the interplay of these factors may indicate an important role for neurophysiology in the PTSD gender difference, significant questions remain. For example, it is difficult to reconcile that most of the neurophysiological differences noted above are relatively small in magnitude, while the majority of prevalence estimates put women at nearly twice the risk of PTSD. Until such time that a deeper knowledge of biology might explain such discrepancies, it seems reasonable to investigate other potential contributors to the wide gender PTSD gap. The discussion now turns to gender differences in diagnostic criteria for the disorder of PTSD.

Gender and PTSD Diagnostic Criteria

Differences in the Immediate Response to Trauma

The American Psychiatric Association's (APA) DSM-IV (1994) criteria for PTSD include a stipulation (criterion A2) that the traumatic event elicit a response of horror, fear, or

helplessness in the victim. It is no surprise, then, that such emotional reactions at the time of the PTE have been positively associated with PTSD (Brewin, Andrews, & Rose, 2000a; Ozer, Best, Lipsey, & Weiss, 2003; Vaiva et al., 2003). A related finding is that those who perceive that their life is threatened by a trauma are also subsequently at greater risk of PTSD (Ozer et al., 2003; Resnick, Kilpatrick, Dansky, Saunders, & Best, 1993). Some recent studies, however, have reported that there are gender differences in such peritraumatic emotional responses and perceptions of threat to life.

Breslau & Kessler (2001) analyzed responses on the DSM-IV PTSD trauma criterion (criterion A) using a large sample ($N = 2181$; age range = 18-45 years) originally selected using random-digit dialing. The results indicated that women were more likely than men to endorse criterion A2 in relation to a suffered trauma (adjusted $OR = 2.66$, 95% CI [1.92, 3.71]). Indeed, gender was the only factor significantly related to endorsing A2—including such seemingly relevant factors as the type of event experienced. A smaller, longitudinal study of criterion A2 (Brewin et al., 2000a) came to similar conclusions. The authors obtained a convenience sample consisting of victims of violent crime recruited through local police and hospitals ($N = 138$; 75% male; mean age = 36.7). Those subjects reporting that they experienced intense fear, helplessness, and/or horror at the time of the traumatic event were more likely to be identified as having PTSD at six months. Also, both fear and horror were found to be significantly correlated with female gender.

Vaiva and colleagues (2003) conducted a longitudinal study looking at the relationship between peritraumatic fright and later PTSD status among a sample of consecutively hospitalized

victims of MVA (N = 123; 68% male; mean age = 31.3). A fright reaction was determined to have occurred when a victim voiced the perception that their life was threatened and identified experiencing a subsequent lack of thoughts, words, or similar. MVA victims who were judged to have experienced such a fright response in reaction to the trauma were much more likely than those who did not identify a fright reaction to be diagnosed with PTSD (OR = 16.75). With respect to gender differences, females were reported to have a higher likelihood of indicating a peritraumatic fright reaction (χ^2 = 7.4, $d.f.$ = 1, p = .007). However, possibly due to the relatively small sample size, there was no significant PTSD gender difference. Nonetheless, a couple of relatively small studies (Bryant & Harvey, 2003; Fullerton et al., 2001) associated a peritraumatic dissociative reaction with higher risk for subsequent PTSD, and reported that females were more likely to identify such a reaction at the time of the trauma. Both studies also found that women with peritraumatic dissociation were far more likely than men with the same to later receive a PTSD diagnosis.

While these limited results are hardly conclusive, they are suggestive of a role for gender differences related to peritraumatic reactions in influencing the observed PTSD differential between males and females. Since such an emotional reaction is a necessary component of any formal PTSD diagnosis, further investigation of possible gender differences in this area is warranted.

Gender Differences in PTSD Symptom Reporting

There is some evidence that, at the statistical level, women tend to respond to symptom queries on PTSD instruments in a pattern that is unique compared to men. Before reviewing this

literature, however, a brief discussion of the organization of PTSD symptoms organization is necessary. The DSM-IV (APA, 1994) recognizes three official symptom clusters consisting of reexperiencing (also referred to as intrusion), avoidance/numbing, and hyperarousal. Multiple studies using factor analysis have suggested that four clusters would be more appropriate with numbing and avoidance separated into distinct symptom categories (see Asmundson, Stapleton, & Taylor, 2004, for a review). Although the subject PTSD instrument, the PC-PTSD screen, uses the four cluster framework, both conceptual approaches are employed in the reviewed studies that follow. Depending on how the various researchers conceptualized the constructs, avoidance and numbing will sometimes be discussed jointly and at other times as distinct subclusters.

Reexperiencing/Intrusion. Norris, Perilla, Ibanez, and Murphy (2001) used a cross-sectional, retrospective design to look at two groups that had been exposed to serious neighborhood hurricane damage. The researchers used purposive sampling in order to include approximately equal proportions of both sexes and younger, middle aged, and older subjects. The Mexican sample ($N = 200$) and the U.S. sample ($N = 270$; 50% white, 50% African-American) were interviewed using the Revised Civilian Mississippi Scale for PTSD (Norris & Perilla, 1996) six months post-disaster. This 30-item instrument is not directly linked to the seventeen individual symptoms included in the DSM-IV diagnosis of PTSD.

Results indicated that, compared to males from the same culture, US white women, US black women, and Mexican women all exhibited elevated intrusion subscale scores, as well as higher overall PTSD scores on the PTSD instrument. However, each of these groups of women

also had significantly higher scores than comparable men in the avoidance/numbing cluster, and only the African-American females from the US were not also higher on the third PTSD symptom cluster of arousal. These particular results on intrusion, therefore, appear to be representative of a generally increased response to all PTSD clusters rather than a gender disparity specific to the reexperiencing/intrusion cluster.

Amir and Sol (1999) reported that in their large convenience sample ($N = 983$) exposed to political violence, the intrusion subscale scores were higher for women than men. This result was consistent across two instruments: the IES and the PTSD Scale (Horowitz, Wilner, & Kaltreider, 1980). On the PTSD Scale, intrusion was the only subscale that was differentially elevated for females. While the IES, like the Revised Civilian Mississippi Scale for PTSD, is not directly linked to the DSM PTSD symptoms, the PTSD Scale's seventeen items do so correspond.

Avoidance and Numbing. In Breslau et al. (1999), the researchers used a three-cluster approach to reanalyze the epidemiological data used in the Breslau et al. (1998) investigation. When examining symptom differences among the subgroup exposed to assaultive violence, women were more likely than men to meet the cluster criteria for reexperiencing, avoidance/numbing, and hyperarousal. However, avoidance and numbing exhibited the largest disparity, with females nearly three times as likely as males (ratio of women to men = 2.8:1) to meet criteria for that cluster. Comparison of all three symptom cluster gender ratios revealed that the difference between the avoidance/numbing ratio and the reexperiencing ratio (1.5:1) was

significant (p = .031), while the difference between the avoidance/numbing and hyperarousal (1.8:1) ratios approached significance (p = .062).

Other evidence is also suggestive of differences between men and women in the area of avoidance and numbing. One study (Fullerton et al., 2001) found that females were significantly more likely than men to endorse that cluster of symptoms (OR = 4.71). Another (Norris et al., 2001) reported that white women, Mexican women, but not African-American women, scored significantly higher on the avoidance/numbing subscale. Within that cluster and without respect to culture, women were more likely than men to endorse two items that assessed avoiding reminders of the traumatic event and two items that addressed emotional numbing. Finally, Amir and Sol (1999) indicated females had significantly higher scores than males on the avoidance subscale of the IES, but not on the PTSD Scale.

Some recent work viewing avoidance and numbing as separate constructs has pointed to the pivotal nature of one or the other symptom in the diagnosis of PTSD. For example, in a validation study examining both the PC-PTSD screen and a diagnostic checklist for PTSD among combat veterans returning from deployment (Bliese et al., 2008), an item response theory analysis revealed that the highest information was provided by two avoidance items. In fact, on the PTSD screen a single question about avoidant behavior performed about as well as the full four item screen for identifying a probable positive PTSD diagnosis. However, this study's sample had an extremely small proportion of females (2% to 4%, depending on the particular analysis) and no information was reported on gender differences.

A couple of recent studies have illuminated Breslau and colleagues' (1999) findings with evidence that emotional numbing plays a crucial role in more severe PTSD cases. Chung and Breslau (2008) used the four-cluster model of PTSD to look at the relationship between gender and PTSD symptoms within the context of three severity classes of the disorder (pervasive disturbance, intermediate disturbance, and no disturbance). In this large (N = 1360), but non-random, sample of young adults (mean age = 21 years) from a mid-Atlantic city, no gender differences were detected within the three disturbance classes in the rate of reporting symptom clusters. Instead, results indicated a connection between being violently assaulted and higher rates of emotional numbing. Interestingly, though, women who had been violently assaulted were more likely than male violent assault victims to be judged as having pervasive disturbance. Similarly, Breslau, Reboussin, Anthony, and Storr (2005) found that assaultive victimization resulted in a more elevated risk of pervasive disturbance for women than for men, with the pervasive disturbance class exhibiting a strong association with DSM-IV PTSD. Breslau and colleagues (2005) analyzed both the sample from Breslau et al. (1998) and the sample later used by Chung and Breslau (2008). An important additional finding was that for both males and females the emotional numbing category of symptoms was rarely identified outside of pervasive disturbance, prompting these authors to speculate that emotional numbing may be a "marker" (Breslau et al., 2005; p. 1350) for a PTSD diagnosis.

Taken together, these three studies (Breslau et al., 1999; Breslau et al., 2005; Chung & Breslau, 2008) raise the question of whether there is a connection between gender, assaultive violence, and emotional numbing. Given that females are disproportionately exposed to sexually

assaultive violence and that sexual assault is among the traumata most likely to result in PTSD (Gnanadesikan, Novins, & Beals, 2005; Resnick et al., 1993), further investigation is warranted regarding gender differences in emotional numbing and particular types of violent assault.

Arousal. A couple of the studies already cited for the reexperiencing/intrusion and avoidance/numbing symptoms of PTSD also provide evidence of a gender differential in arousal symptom reporting. Norris and colleagues (2001) reported that in regression analyses of cluster subscale scores conducted across all studied cultures, gender was predictive only of arousal (not reexperiencing/intrusion or avoidance/numbing). Females were significantly more likely to score positive on the arousal cluster. Fullerton et al. (2001) likewise indicated a higher likelihood of women than men endorsing the arousal cluster (statistically significant $OR = 3.8$).

Conclusion of symptom differences. Of the reviewed studies, a number found evidence that women were more likely than men to endorse multiple, or even all PTSD symptom clusters. This could imply that the PTSD gender difference is at least partly the result of females' more frequently responding "yes" when queried about PTSD symptoms. Only one study showed uniquely increased intrusion reporting for females, and that on only one of two instruments (Amir & Sol, 1999). However, multiple studies provide limited evidence that the gender difference in symptom reporting may be more pronounced for the avoidance/numbing cluster (Breslau et al., 1999; Fullerton et al., 2001), and perhaps with the arousal cluster, as well (Fullerton et al., 2001; Norris et al., 2001). Finally, studies separately linking both female gender and assaultive violence to emotional numbing (Breslau et al., 1999) and to a pervasive disturbance class associated with DSM-IV PTSD (Breslau et al., 2005; Chung & Breslau, 2008)

give indications that gender may be interacting in complex ways with particular types of trauma to produce unique symptom reporting.

Summary and Connections to Current Study

Summary of PTSD Gender Differences

Prevalence and trauma exposure. Despite a history predominantly focused on males, the diagnosis of PTSD has generally been shown to occur disproportionately among females in both epidemiological and non-epidemiological prevalence studies. This gender disparity appears to hold across international borders and among children as well as adults. Even though women are at greater risk of PTSD, trauma prevalence studies are in relative agreement that women report fewer traumatic events than men. This indicates that the simple average number of traumas experienced for males and females is not an adequate explanation for the PTSD gender difference. Rather, a closer look at the PTE research reveals that men and women are at uniquely elevated risk for particular types of trauma. Men tend to face more accidents, combat, and physical assaults, whereas women more often deal with sexually assaultive traumas—although evidence points to growing combat exposure for women. Of the most common types of traumatic event, sexual trauma (Kessler et al., 1995; Norris, 1992) and combat (Kessler et al., 1995; Kulka et al., 1990) have been associated with higher risk for PTSD subsequent to exposure. Thus, differential exposure to particular types of trauma may account for some of the differential risk of PTSD among genders.

Neurophysiology. Neurophysiological differences between men and women constitute another potential contributing factor. Studies looking at the effects of estrogen level variation on

stress-related neurochemicals such as cortisol and NPY have shown some promise for informing the gender PTSD difference. However, at the current time there is no research directly linking differential PTSD risk with biological gender differences. Additionally, the measured changes in these neurochemicals is generally very small—making it difficult to conceive of how such modest changes could lead to the relatively substantial observed PTSD prevalence differences between men and women. Nonetheless, limited extant evidence suggests that more than one neurobiological system has the potential for influencing gender differences in the area of stressful memory acquisition and fear conditioning.

Research into ELS provides a possible connection between the neurobiology of PTSD and differential trauma exposure. With ELS associated with neurophysiological changes that are elsewhere linked to PTSD, it appears plausible that women's disproportionate exposure to child sexual abuse could result in a gender difference in ELS-related biological vulnerability to PTSD.

Peritraumatic reactions. Relatively few studies have examined gender differences and PTSD as they relate to the response to trauma and subsequent social support. Among immediate reactions to trauma, the strongest evidence indicates that the peritraumatic emotion of fear, which is associated with greater PTSD risk, is also more often reported by women than men. Likewise, peritraumatic dissociation has been linked to higher rates of later PTSD; with the limited research in this area showing that such dissociation leads to significantly higher PTSD risk for women than for men.

PTSD symptom reporting. Dissociation, as defined in the DSM-IV criteria for ASD (APA, 1994), contains an element of emotional numbing. This conceptual relationship between

dissociation and emotional numbing suggests an interesting potential connection between gender differences in both peritraumatic dissociation and symptom reporting. Perhaps a dissociative reaction at the time of the trauma is somehow connected to persistent emotional numbing as a coping mechanism. If this progression were to occur more frequently among women than men, whether across all traumas or a specific subset of traumas such as violent assault, this could provide yet another potential mechanism leading to differential PTSD risk. While not conclusive at this point, some evidence points to more frequent reporting of numbing and/or avoidance PTSD symptoms for women as compared to men. One large study (Breslau et al., 2005) found that for both genders emotional numbing was strongly associated with pervasive a disturbance level linked to PTSD. Moreover, among those subjects that have been violently assaulted, emotional numbing appears to be an especially pronounced feature with females.

The symptom cluster of arousal has also been shown to have a gender difference in a couple of studies (Fullerton et al., 2001; Norris et al., 2001). Again, women reported significantly more arousal symptoms than men. Evidence for a gender differential for intrusion symptoms can be found in two studies, as well (Amir & Sol, 1999; Norris et al., 2001), and may have some support in the above mentioned neurophysiological sex differences that suggest possible enhanced stressful memory consolidation for women under certain circumstances. However, the particular results showing elevated intrusion symptoms for females in Norris et al. (2001) appeared to be part of a larger pattern of increased female endorsement of all PTSD symptom clusters.

Areas for Possible Research

While knowledge of the neurobiological bases of PTSD and stress continues to grow, it is clear that large gaps remain. The bulk of the research, so far, has been performed using animal subjects. More work is needed to foster our biological understanding of PTSD and the human stress response. For example, research indicates that NPY may inhibit fear-related memory consolidation. One hope is that further delineation of how such memories are processed from a biological perspective would be helpful for learning the best avenues of PTSD intervention—whether psychotherapeutic or pharmacological or some combination—for a particular patient with given NPY levels.

Also, little is known about underlying neurophysiological processes involved in either peritraumatic responses or subsequent PTSD symptom reporting. For instance, what are the contributing biological factors in what appear to be unique gender responses contemporary to a traumatic event? Similarly, a dearth of physiologically-oriented research has investigated what leads men and women to endorse differing patterns of symptoms in PTSD. Perhaps future efforts will provide connections between estrogen- or menstruation-linked gender differences in the neurochemicals mentioned above and particular symptoms that are more likely to be endorsed by either females or males.

One possible direction for future work in this area is with NE. Elevated levels of this neurotransmitter have been linked to increased arousal and intrusive memories. Thus, settling the question of whether estrogen enhances or diminishes the NE stress response could open

further avenues to investigate whether differences in such gonadal hormones are related to the observed gender differences in PTSD symptoms such as hyperarousal and intrusion.

With the limited, but important evidence connecting peritraumatic emotional and dissociative reactions to elevated PTSD risk for women, further investigation would be helpful. Specifically, what are needed are prospective, longitudinal studies initially assessing immediate reactions to trauma and subsequently measuring for PTSD diagnosis. It will be crucial for such studies to have sufficiently large male and female sample sizes to provide the statistical power to detect any overall gender PTSD difference. Unfortunately, it can be challenging to achieve large samples with a prospective design due to the need to quickly identify and assess those who have experienced a trauma. Perhaps telephone interviews conducted with a large population—a design which was used in many of the reviewed epidemiological prevalence studies—would be helpful in this regard.

Whatever the specific etiology of the PTSD gender gap, clearly the result is that when the diagnostic criteria for PTSD are applied to women, they are more likely than men to meet sufficient criteria for a positive diagnosis. With various measurement instruments being the predominant method for determining the presence or absence of PTSD in most prevalence studies, it is important to ensure that the tests themselves are not biased toward diagnosing women. This issue is potentially related to the gender symptom differences in that many of these PTSD instruments are directly linked to the DSM criteria for the disorder—including the symptoms (Cusack et al., 2002). Therefore, an important possible factor for consideration when examining the PTSD gender disparity is that the instruments themselves may be biased toward

diagnosing females with PTSD. A preliminary search revealed no differential functioning or measurement equivalence studies of PTSD instruments using males and females as the comparison groups.

As suggested by Kimerling and colleagues (2007), studies testing measurement equivalence of common PTSD instruments across genders would be helpful for ruling out differential functioning, or test bias of PTSD tests for women versus men. If measurement equivalence by gender holds for the instruments that are evaluated in this fashion, one important result would be increased confidence that the reviewed gender differences are not merely artifacts of biased PTSD measurement. Such an evaluation is the primary focus of the current study. The target instrument is the PC-PTSD screen used extensively by the US military.

Primary Care-PTSD Screen

The PC-PTSD consists of four dichotomously scored (Yes/No) items meant to assess the four main symptom clusters of PTSD: reexperiencing, avoidance, arousal, and numbing. See Figure 1 on the next page for the PC-PTSD screen as it is configured within the PDHRA. All subsequent tables and figures will likewise be located on the page immediately following the page where the table or figure was mentioned in the text. Initial validity and reliability was examined using a sample of primary medicine patients (N = 188; 66% female) from one Veteran's Administration (VA) facility with the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1995) serving as the gold standard (Prins et al., 2003).

Sensitivity and specificity were highest with cutoff scores of two and three positive responses. A cutoff of two was better for sensitivity at .91, with specificity at .72.

12. Have you ever had any experience that was so frightening, horrible, or upsetting that, IN THE PAST MONTH, you….

a. Have had nightmares about it or thought about it when you did not want to? **Yes/No**

b. Tried hard not to think about it or went out of your way to avoid situations that remind you of it? **Yes/No**

c. Were constantly on guard, watchful, or easily startled? **Yes/No**

d. Felt numb or detached from others, activities, or your surroundings? **Yes/No**

Figure 1. The Primary Care Posttraumatic Stress Disorder (PC-PTSD) screen as it is configured within the Post-Deployment Health Reassessment (PDHRA).

Three affirmative responses were determined to be optimal, though, diminishing sensitivity to .78 while improving specificity to .87. When applying the cutoff of three to each gender the authors received somewhat different results. For males sensitivity was .94 and specificity was .92, yielding a diagnostic efficiency of .92. The figures for females were lower with sensitivity at .70, specificity at .84, and efficiency at .81. The researchers speculated that the lower performance at this cutoff level with women may have resulted from females' measured PTSD symptoms being differentially affected by factors such as comorbidities with PTSD and PTSD course length.

Bliese and colleagues (2008) recently validated the PC-PTSD against a modified PTSD module of the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) with a sample of active duty soldiers ($N = 352$; 96% male) returning from combat. Their analysis yielded similar results, with shifting from a two "yes" cutoff to a three "yes" cutoff diminishing sensitivity from .85 to .76 while improving specificity from .71 to .88. The researchers also found that using the avoidance question alone as a single-item screen yielded a diagnostic efficiency approximately between that provided by the two- and three positive response cutoffs. However, with only 13 females (4%) the validation sample was overwhelmingly male, leaving this study's generalizability to women open to question.

Army researchers in Europe (Bliese et al., 2004) conducted a validation study of a screen for multiple psychological issues that included the PC-PTSD as one element. The sample ($N = 528$) consisted of soldiers returning from a deployment to Iraq. Clinical interviews were conducted using the modified MINI (Sheehan et al., 1998) to provide the standard against which

the screen was validated. These authors found that a cutoff of two resulted in sensitivity of .73 and specificity of .88. A three cutoff, on the other hand, yielded sensitivity of only .46 but a specificity of .97. Three was chosen as the optimal cutoff since specificity was favored due to the need, identified prior to the study, to minimize false positives. However, the sensitivity at this cutoff was very low (.46). Also, this study provides no indication of what proportion of the sample was female.

As these three validation studies establish, current recommendations for the PC-PTSD are that a score of three (out of four) is indicative of a probable PTSD diagnosis (Bliese et al., 2008; Prins et al., 2003). However, within the USAF a positive response on any one of the four PC-PTSD items results in supplemental testing and a referral for face-to-face follow-up (Office of the USAF Surgeon General, 2008). This approach equates to using a cutoff of at least one positive PC-PTSD item. The PC-PTSD has thus far not been evaluated with a sample of returning military deployers that includes a significant proportion of females. Also, none of the published PC-PTSD validation research addresses measurement equivalency across groups, including gender.

Research Questions

The PC-PTSD is yet to be evaluated with a sample of returning military deployers that includes a significant sample size of females. Also, none of the published PC-PTSD validation research addressed measurement equivalency across gender. The questions pursued in this study arise from these issues as well as the areas for further research previously identified in this chapter. These research questions are as follows:

Research Question 1: Are the parameters of the PC-PTSD items consistent across groups of returning male and female USAF Iraq and Afghanistan deployers (the target population)?

Research Question 2: Does the PC-PTSD, in various scoring configurations including the recommended cutoff of three, differentially predict a clinical diagnosis of PTSD for men and women in the target population?

Research Question 3: Does the proposed single-item screen version of the PC-PTSD, using only the avoidance item, differentially predict a clinical diagnosis of PTSD for males and females in the target population?

Research Question 4: Is the diagnostic utility of the PC-PTSD consistent when used with men versus women in the target population?

Research Question 5: Does the PC-PTSD demonstrate good convergent and discriminant validity when used with a sizable sample of the target population?

CHAPTER III: METHOD

Sources of Data

There were two USAF datasets used as the sources for the data in this study: the PDHRA database and the military electronic medical record system. There are actually two PDHRA databases containing responses on all screening items. An older database includes responses from 2005 to 2007 on a previous version of the PDHRA. The database used for the present study is linked to the PDHRA as revised in 2008 and includes response data for 2008 and 2009. The PC-PTSD is identical in both PDHRA versions. The military electronic medical record system records data on all military medical visits for all members of the USAF.

Institutional Review Board (IRB) permission was sought from the Human Subjects Committee of the College of Social Work at the University of Tennessee using Form A (Certification for Exemption from IRB Review for Research Involving Human Subjects). The exemption was granted based on Category 4, which applies to research utilizing existing data under the condition that the investigator records the information in a manner that precludes personal identification of the subjects. Approval for the study was also requested from the Air Force Research Laboratory (AFRL) IRB at Wright-Patterson Air Force Base, Ohio. This body ruled that, under federal regulations, the current study did not constitute human subjects research due to the use of pre-existing data that had been de-identified and, consequently, no AFRL IRB approval was required.

PDHRA Database

Two post-deployment screenings containing the PC-PTSD are given to returning deployers in the USAF (and in the DoD, in general). The first occurs concurrently with an airman's return from deployment. This screening program was the first to be implemented and was termed the Post-deployment Health Assessment (PDHA). Subsequently, though, research with a matched sample of US Army soldiers returning from a year-long deployment to Iraq (Bliese et al., 2007) showed convincing evidence that prevalence of PTSD and other mental health-related problems increased during the six months postcombat.

Bliese and colleagues (2007) is the only study to date examining changes over time in mental health prevalence after exposure to combat. The researchers conducted assessments for PTSD, depression, and other problems during a seven day "reintegration program" (Bliese et al., 2007, p. 143) with over 1500 soldiers immediately upon their return from deployment. Four months (about 120 days) later, 509 randomly sampled soldiers were selected from those who took the original assessment to receive the identical assessment instruments. Results indicated PTSD prevalence significantly increased from either 1.39% to 4.81% or 2.98% to 8.42%, depending on whether a standard or liberalized cutoff of the Posttraumatic Stress Disorder Checklist (PCL; Weathers, Litz, Huska, Herman, & Keane, 1993) was used. With the one exception of relationship problems, the prevalence of all other measured mental health problems also increased significantly during this time period. In fact, the pre-publication knowledge of these results was one reason that the DoD instituted the PDHRA in 2005 to supplement the initial PDHA (Bliese et al., 2007).

Given these results, it was decided that the present study would use the later PDHRA rather than the PDHA database to assess the PC-PTSD screen. A screen assessing for PTSD would likely encounter more positive responses during the period 90 to 180 days post-deployment (which encompassed the 120 day postcombat timing used in Bliese et al., 2007) than immediately after a deployment ends.

The PDHRA process itself is largely automated, although there are personnel designated at each military medical facility who are responsible for administration of the system. At the appropriate time, a USAF computer system prompts an email to be sent to each recent deployer requesting that they complete the online survey. If the PDHRA screening is positive for any of a number of pre-identified subscales, the member will be referred for further assessment with an appropriate medical and/or behavioral health provider (Office of the USAF Surgeon General, 2008).

Medical Record Database

The USAF employs an electronic medical record system containing data for all medical and mental health diagnoses received by all USAF members when under the care of DoD providers. Specifically, each medical encounter for an individual contains codes for any diagnoses (including a code for "no diagnosis") received during the visit and any associated procedures performed by the practitioner.

In this study, medical diagnostic information was linked to subjects' PDHRA data in the following fashion: For every PDHRA respondent during 2008 and 2009, the researcher merged the PDHRA data by common study case number with the medical record database containing all

medical visits for each subject. Then, information was extracted regarding the dichotomous presence or absence of both a PTSD diagnosis and a major depressive disorder diagnosis during any medical encounter in the 90 day period after the date the subject completed the PDHRA. All other diagnoses were ignored and deleted from the study datasets.

Design

This was a non-experimental validity study of a PTSD screen using secondary data from a large population of male and female USAF veterans of Iraq and Afghanistan. The first part of the investigation evaluated whether the PC-PTSD (Prins et al., 2003) screen was measuring PTSD risk with equal validity for both men and women. IRT DIF methods were employed to detect measurement inequivalence at the item level. This measurement equivalence phase sought internal evidence of differential validity since the analysis did not make use of any criterion external to the screen itself (Camilli, 2006). Instead, only the item response data were analyzed.

The second phase focused on investigating external evidence of validity in the instrument. This part of the study looked for "differential prediction" (Camilli, 2006, p. 231) by gender through a comparison of responses on the PC-PTSD to a subsequent PTSD diagnosis. For the concurrent validity analyses, the predicted criterion was the presence or absence of a clinical diagnosis of PTSD as recorded in the military medical record database within 90 days of taking the PDHRA. Divergent validity was evaluated by performing similar comparisons between the PC-PTSD responses and the dichotomous presence/absence of a diagnosis of major depression (also within the 90 day time frame and from the same medical record database).

Predictive validity was assessed using both logistic regression methods and standard diagnostic utility statistics. Both methods were used because, although regression readily provides statistics for evaluating model fit and hypothesis testing, diagnostic utility yields statistics that are at once relatively simple to understand and practically useful for aiding a clinician in diagnostic or treatment decisions related to the scale.

Finally, since the DoD is using the PC-PTSD to screen for PTSD with all returning military deployers, this study presented a unique opportunity to assess the overall validity of the PC-PTSD with a very large sample of males and females from that target population. To that end, convergent and discriminant validity were additionally assessed using confirmatory factor analysis (CFA) methods.

Variables of Interest

Dependent Variables

PTSD status. For the predictive validity study, the dichotomous outcome variable is PTSD diagnosis status. Positive PTSD status is defined as the presence of a clinical PTSD diagnosis during any military medical encounter within 90 days of completion of the PDHRA, as recorded in the electronic medical record database. Conversely, negative PTSD status is determined when no medical encounters during the 90 days after PDHRA completion indicated a clinical PTSD diagnosis.

Major depressive disorder (MDD) status. The dependent variable (DV) for the divergent validity analysis is MDD diagnosis status. For the purpose of this study, MDD encompasses two separate diagnoses: MDD, single episode and MDD, recurrent. The reason for

including both is that they each require meeting the criteria for a major depressive episode. Since this variable is being used only to establish discriminant validity for the PC-PTSD screen, whether or not a subject has experienced one or multiple major depressive episodes is of little concern in this study. Like positive PTSD status, positive MDD status is defined as the presence of a clinical MDD, single episode or MDD, recurrent diagnosis during any military medical encounter within the 90 days after responding to the PDHRA. Negative MDD status corresponds to the case when no medical encounters during the 90 day post-screening period resulted in either a MDD, single episode or a MDD, recurrent diagnosis.

The MDD diagnosis was chosen to test for discriminant validity because, as mentioned in the introduction, depression is highly comorbid with PTSD. In fact, in the National Comorbidity Survey (Kessler et al., 1995) major depression exhibited the highest comorbidity with PTSD for women and was a close second to substance abuse among men. The rationale was that if the PC-PTSD could discriminate between PTSD and a disorder that commonly co-occurs with PTSD, this would imply strong divergent validity for the scale.

Independent Variables

Risk of PTSD. For all analyses assessing for external evidence of validity, PTSD risk as measured by the PC-PTSD screen was the primary independent variable (IV). Various cut scores and PC-PTSD responses were used as the determinant for risk, as suggested by previous research or standard USAF practice. A cut score was defined as being met whenever a subject responded positively to a predetermined number of PC-PTSD items equal to or greater than the chosen cut score. For example, a subject meeting the cut score of 3 on the PC-PTSD responded

positively to any three items or to all four items. Subjects not meeting a cut score of 3 responded positively to zero, one, or two PC-PTSD items. The actual PC-PTSD response values that were chosen to represent PTSD risk in each analysis will be delineated in the sections below detailing the regression and diagnostic utility analyses.

Gender. A subject's gender was determined by the "Sex" field in the demographics of the PDHRA database. For the purposes of this study, the terms "sex" and "gender" are used interchangeably. Gender (male or female) was used to determine reference and focal group membership in the IRT DIF analysis, as a moderator variable in the regression analyses, and as a grouping variable in the diagnostic utility analyses.

Subjects

Subjects were male and female members of the USAF who had participated in the PDHRA screening during 2008 and 2009 within 90 to 180 days of returning from deployment to Iraq and Afghanistan. The DoD requires that every returning deployer is provided the opportunity to participate in this screening within 90 to 180 days of returning from their deployment. However, the screening program is essentially voluntary and, therefore, not every USAF member returning from deployment completes the PDHRA (Office of the USAF Surgeon General, 2008). No information was available on those who opted not to complete the PDHRA.

Missing Data

Listwise deletion was employed for all missing data. The potential risk with listwise deletion is twofold: attenuated statistical power and a less representative remaining sample (Saunders et al., 2006). However, all analysis sample sizes were relatively large and contained

less than 2% of cases with missing data. Under these circumstances, the risks associated with listwise deletion are minimized (Saunders et al., 2006).

The only variable of interest that contained missing data in the source datasets was PTSD risk, measured by the PC-PTSD. More information on the specific proportion of cases with missing data that were deleted is included in later sections describing the specific analysis samples.

Exclusion Criteria

It is possible that some individuals would have deployed to Iraq and/or Afghanistan more than once and, therefore, been given multiple opportunities to respond to the PDHRA during 2008 and 2009. If so, such individuals could have also taken the PC-PTSD multiple times. Including more than one screening for a particular subject would result in the sample containing paired data. Therefore, whenever a subject was found to have completed the PC-PTSD more than once, only the most recent screening results were included in this study. Keeping the most recent screening allowed the final sample to retain as much information as possible on the total number of deployments for each subject. If any previous screening had been selected, that screening's variable for the total number of deployments would not have reflected any subsequent deployments for which the individual had also been screened. In this study, the total deployments variable is a demographic variable that is used to help determine the extent to which the male and female subgroups are similar.

No identifying information (e.g., name, date of birth, social security number) on subjects was included in the study data. Instead, each subject was given a unique case identification

number that was unrelated to any identifying information. However, subject data did include limited demographic information, such as gender, age, rank, military status prior to deployment (i.e., active duty, guard, or reserve), number of deployments, and marital status. The study sample excluded any USAF officers holding the rank of Brigadier General (O-7) or higher. This decision was made in order to prevent inadvertent identification of particular subjects. There are a very limited number of USAF general officers deployed to Iraq and Afghanistan at any one time, and their roles tend to be relatively high profile. These conditions heighten the risk that the screening results for an individual subject could be linked to their identity. Therefore, out of privacy concerns, all general officers were excluded. All other ranks were substantially more common in the deployed setting and were therefore included in the study data.

Samples

While all samples were derived from the same larger datasets, the different types of analysis and variables dictated that unique samples be employed for some analyses. Following is a discussion of the specific characteristics of each sample used in this study.

IRT DIF sample. The 9806 subjects (8381 males, 1425 females) used in the IRT DIF study constituted the entire population of USAF Iraq and Afghanistan deployers completing the PDHRA (including all items on the PC-PTSD) from 2008 through 2009 who responded during the 90-180 day post-deployment window. In order to attain this final sample, 134 cases (111 males, 23 females) that contained at least one missing response on the PC-PTSD portion of the PDHRA were removed from the total population that included missing values (N = 9940, with 8492 men and 1448 women). The overall percentage of cases with missing data that were

removed constituted 1.35% of the larger population. While a slightly larger percentage of women than men had cases deleted due to missing PC-PTSD responses (1.59% of women vs. 1.31% of men), an analysis of this difference using the Pearson chi-square in PASW indicated that it was not statistically significant ($\chi^2(1, N = 9940) = .736, p = .391$).

Regression and diagnostic utility sample. The same sample was used for both the regression and diagnostic utility analyses. It constituted a smaller, but still substantial subset ($N = 6999$) of the total population used in the IRT DIF study. Of these, 5785 were men and 1214 were women. The smaller magnitude of the regression sample when compared to the IRT DIF sample was the result of merging the PDHRA data with the electronic medical record data. Cases were only included in the predictive validity sample if they had at least one medical diagnostic visit within the 90 days after completing the PDHRA. This resulted in the exclusion of 2835 subjects (2621 men and 214 women) that had been included in the IRT DIF sample.

There were a couple of purposes for applying this data criterion. The first was to ensure that only those subjects with an opportunity for a clinical PTSD or MDD diagnosis were included in the study sample. The second purpose was to accommodate an important assumption: that the symptoms being measured at the time of screening were the same (or reasonably similar to) the symptoms being considered by the clinician at the time of the medical visit. The greater the length of time between the screening date and the subsequent diagnostic medical encounter, the more tenuous this assumption becomes.

The resulting sample ($N = 7105$, males = 5871, females = 1234), however, still contained missing responses on the PC-PTSD items. As with the IRT DIF sample, all cases with missing

data on these items were removed to reach the final study sample of 6999. Overall, 106 (1.49%) cases with missing responses were removed. The gender breakdown of the removed cases was 86 (1.46%) men and 20 (1.62%) women. Again, the difference in the proportion of deleted cases was not statistically significant ($\chi^2(1, N = 7105) = .169, p = .681$).

CFA sample. The sample used for the CFA validity analysis was based on the sample used for the logistic regression and diagnostic utility analyses. As detailed above, the regression sample had already been purged of cases with missing data on the PC-PTSD. However, since the CFA analysis also used data from a depression screen on the PDHRA, any cases with missing data on the two depression screen items were deleted. This procedure resulted in 36 (29 men and 7 women) cases being deleted, reducing the sample size to 6963 (5756 males, 1207 females). Only .50% of male cases and .58% of female cases were removed and the gender difference was not statistically significant ($\chi^2(1, N = 6999) = .111, p = .739$).

Data Analyses

Measurement Equivalence by Gender

Equivalence or invariance of measurement is said to exist when members of two subgroups that possess equal trait or ability levels on a latent construct are observed to also receive the same score on a given instrument measuring that construct (Drasgow & Kanfer, 1985). Of interest in the present study was whether measurement invariance holds for the PC-PTSD across male and female veterans of Iraq and Afghanistan.

Differential item functioning (DIF) utilizing IRT is a common method for determining measurement equivalence (Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993). IRT

relates an underlying construct to observed measurements and provides item response functions that aid in the graphical analysis of the items. IRT also uses measurement invariance definitions that do not require the latent trait score distributions among comparison groups to be identical (Raju et al., 2002). The latter characteristic is especially helpful for detecting any measurement inequivalence across groups. IRT DIF procedures were used in the current study for assessing the measurement equivalence of the PC-PTSD screen across gender groups.

IRT and DIF. DIF research is an outgrowth of item bias research which began in the 1960s with studies investigating whether intelligence and selection tests were biased against minority subjects and in favor of majority subjects (Angoff, 1993). Item bias, now commonly referred to as DIF, can be defined as the case when an item behaves differently when used with two population subgroups that have the same trait or proficiency level (Angoff, 1993).

IRT has been used extensively in educational and psychological measurement for years. This theory is the foundation for many well-known tests such as the Graduate Record Examination (GRE) and the Armed Services Vocational Aptitude Battery (ASVAB) (Embretson & Reise, 2000). Researchers have begun to extend the application of IRT DIF methods beyond its traditional basis in educational assessment to personality and clinical scales (see recent examples in Baker, Caison, & Meade, 2007; Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006; Hays, Liu, Spritzer, & Cella, 2007).

Assumptions of IRT. There are several IRT models, but each model provides a basic statistical characterization of the relationship between a respondent's unique attributes and that individual's response to a given item. Importantly, each item also possesses its own specific

characteristics, or item parameters. IRT models have been devised to handle test items that yield both dichotomous (e.g., correct or incorrect) and polytomous (e.g. partial-credit scoring and Likert rating scales) data (Yen & Fitzpatrick, 2006).

The two foundational ideas of IRT are that (1) certain latent traits are predictive of an examinee's performance on a given item and (2) an item characteristic curve (ICC) describes the relationship between the examinee's latent traits and their performance on the item. (Hambleton, Swaminathan, & Rogers, 1991). An examinee's trait or ability can be referred to as a "person parameter" (Yen & Fitzpatrick, 2006, p. 112). It is assumed that the ICC, also termed the item response function or trace line, is an accurate representation of the relationship between the person parameters and the responses on an item with particular characteristics. When viewed through the lens of IRT, the definition of DIF can be modified: DIF is inferred when, after an item is presented to two or more groups, the resulting item response functions are different between those groups (Lord, 1980; Steinberg & Thissen, 2006).

Three additional assumptions are considered fundamental to most commonly used models within the IRT framework: dimensionality, local independence, and parameter invariance (Hambleton et al., 1991). Dimensionality involves the idea that a certain examinee trait or traits constitute the "'dominant' component or factor that influences test performance" (Hambleton & Swaminathan, 1985, p. 17). If a model assumes that only one trait or ability serves this purpose, the model is termed unidimensional. The local independence assumption holds that, after controlling for relevant abilities, there is no statistical relationship between examinees' responses to various items within a test. This assumption states that responses on

one test item do not influence the responses on a different test item. Rather, it is the latent trait(s) that are the determining factors in both item responses (Hambleton & Swaminathan, 1985; Hambleton et al., 1991). Parameter invariance is the IRT assumption that item parameters are independent of the distribution of ability in the tested population, and that person parameters are likewise independent of the particular group of items used in a test (Hambleton et al., 1991).

Common general IRT models. An important distinguishing factor between many IRT models is the nature and number of item parameters that the model assumes will contribute to a person's response (Hambleton et al., 1991). In the simplest models, only one item parameter or characteristic is assumed to affect examinee performance: item difficulty or location (typically denoted *b*). An IRT model utilizing this assumption is commonly called a Rasch model or a one-parameter logistic (1PL) model (Embretson & Reise, 2000). In IRT, item difficulty can be thought of as the location on the scale of ability (θ) where there exists a 50% chance of a successful response to the item. Therefore, as item difficulty increases, an examinee's ability or trait level must also increase in order to maintain the same probability of passing the item (Hambleton et al., 1991).

The two-parameter logistic (2PL) model adds an item discrimination parameter (usually symbolized *a*). Discrimination affects the slope of an ICC. Items possessing higher discrimination (i.e., steeper slopes) have greater usefulness for sorting respondents into separate trait- or ability-level categories (Hambleton et al., 1991).

A third item characteristic that can be added to an IRT model is a pseudo-chance or pseudo-guessing parameter (designated *c*). The three-parameter logistic (3PL) model uses *c* to

represent the probability that examinees with very low ability (approaching infinity) will successfully respond to the item. For the 3PL model ICC, *c* becomes the lower asymptote of the function (Yen & Fitzpatrick, 2006).

While the 3PL model can be applied to dichotomous response data, the 2PL model is more frequently used (Edelen et al., 2006). Additionally, it is difficult to conceptualize the meaning of "guessing" on an item within an instrument that assesses a trait rather than an ability or aptitude. Therefore, in the current study the PC-PTSD data were fitted to the 2PL model. Using the 2PL, the trace line (*T*) modeling the probability of an affirmative response ($x_i = 1$) to item *i* is represented in the following equation:

$$T(x_i = 1) = \frac{1}{1 + \exp\left[-a_i(\theta - b_i)\right]}$$

(Edelen et al., 2006; Thissen, Steinberg, & Wainer, 1993) where θ is the level on the latent construct that the test is assumed to be measuring and b_i and a_i are the difficulty and discrimination parameters for item *i*, respectively.

Importantly, when there is a good fit between an item response model and the test data to which the model is applied and the tested sample is of sufficient size, the resulting estimates of the item parameters are not dependent upon the population subgroup that was tested (Camilli & Shepherd, 1994; Hambleton & Swaminathan, 1985). Thus, comparison of item characteristics between groups is possible under IRT. These attributes make IRT a very attractive theoretical framework for DIF research.

As touched on in the previous paragraph, sample size is an important consideration when using IRT. There does not appear to be any research addressing this issue in specific relation to

the 2PL model. However, Reise and Yu (1990) recommended a sample size of at least 500 for the graded response model. Since one perspective on the 2PL model with dichotomous responses is that it is a special case of the graded response model (Raju et al., 2002), 500 may be a reasonable starting point absent relevant research findings. Other general recommendations include that the analysis utilize a sample size sufficiently large to make the parameter estimate standard errors small enough to meet the needs of the researcher's specific research question (Embretson & Reise, 2000). Sample sizes for both men and women in this study were substantially larger than 500 each.

IRT Methods for DIF. Several approaches have been devised for detecting DIF using IRT. Examples include *b*-difference methods (Camilli & Shepherd, 1994; Thissen et al., 1993), Lord's chi-square (Lord, 1980), and Raju's area under the curve (Raju, 1988). The current study employed the likelihood ratio test (LRT; Thissen et al., 1993) method. LRT has been widely used in DIF research because it is flexible enough to handle several IRT models as well as both dichotomous and polytomous scoring regimes (Camilli, 2006; Embretson & Reise, 2000). Also, a separate equating procedure is not required since the method itself places the item parameters for the comparison groups onto the same metric by estimating the parameters for both groups simultaneously (Kim & Cohen, 1995).

Kim and Cohen's (1995) comparison study indicated that LRT performed about as well as Raju's area measure and Lord's chi-square on DIF detection. In a Monte Carlo study comparing DIF detection procedures, Finch (2005) found that LRT performed particularly well in detecting DIF in tests containing relatively few items. This is the case for the 4-item PC-

PTSD being evaluated in this study. Finally, the computer program MULTILOG 7.0 is available for conducting LRT DIF analyses, in addition to other IRT applications. In this study, MULTILOG 7.0 was implemented using the newer, more user-friendly interface as reflected in the updated user's guide (du Toit, 2003).

DIF procedure. In the IRT portion of this study, the LRT method was used to test for gender DIF in the items using the 2PL. Females were designated as the focal group and males constituted the reference group. Any items whose parameters significantly differed between the male and female groups as evidenced by the likelihood ratio test were determined to have DIF. Based on the review of PTSD symptom differences by gender, it was hypothesized that, if DIF were detected, it would occur on the avoidance and numbing questions of the PC-PTSD. The expected result would be that the *b* parameter on these items would be significantly lower for females than for males.

Several steps are required to implement LRT DIF. First, prior to engaging in any application of IRT to a particular data set, it is important to assess that the proposed IRT model fits adequately and that the appropriate IRT assumptions are met (Hambleton et al., 1991). Next, the LRT approach requires that DIF-free "anchor" items must be identified (Thissen et al., 1993). These anchors are then used in the actual LRT DIF detection analysis. Finally, certain plots can be used to aid in the interpretation of any DIF.

Checking IRT assumptions. Verifying the appropriateness of IRT assumptions can involve both statistical assessments of model-data fit and seeking direct evidence that particular

IRT assumptions are tenable in relation to the data (Hambleton et al., 1991). In this study both approaches were employed.

Model fit. Methods for analyzing the overall fit of an IRT model to the data typically use a goodness-of-fit statistic which is compared to a chi-square distribution. However, several problems have been identified with this approach. For one, it has been shown that larger sample sizes result in excessive identification of misfit (Hambleton et al., 1991; Stone, 2000). In this application, these goodness-of-fit statistics also often use expected frequencies derived from estimated, unknown trait level theta (θ) parameters. This can be particularly problematic in tests with few items, as was the case in this study. Fewer items are linked with higher variability in θ estimates, reducing the reliability of the associated goodness-of-fit statistic (Stone, 2000). Therefore, as recommended by Hambleton and colleagues (1991), evaluations of specific IRT assumptions were used to provide broader evidence of satisfactory IRT model fit.

This study employed nested model comparison methods for testing overall fit of the IRT model to the data. Despite the limitations mentioned above, M. Orlando Edelen, who has conducted a number of studies related to IRT DIF methods (e.g., Edelen et al., 2006; Hepner, Morales, Hayes, Edelen, & Miranda, 2008; Orlando & Marshall, 2002; Orlando & Thissen, 2000), stated in personal communication (November 2, 2009) that model comparison remains the most widely-used approach for this purpose. First, the model was estimated twice (with data from males and females together): once using the 1PL model (*b* is the only item parameter) and then using the 2PL (two item parameters—*a* and *b*). Then, since the only difference in these model runs was the presence of *a* (slope) in the 2PL, the models were compared using negative

two times the log likelihoods. This statistic is distributed as a chi-square with one *d.f.* per item included in the models (Embretson & Reise, 2000). For the four-item PC-PTSD used in this study, for example, the *d.f.* difference for this model comparison would be equal to four. As mentioned previously, it was expected that the 2PL would be the appropriate model for these data. The purpose of this model comparison, then, was to test whether the data were more appropriately modeled with the 2PL than the simpler 1PL model. A similar comparison was conducted between the 2PL and the 3PL parameterizations of the data.

In addition to evaluating model fit at the level of the test, a method was also employed to assess model fit at the item level. Using the $S\text{-}\chi^2$ statistic, the observed and expected frequencies were compared at every level of the total positive responses on the scale (Bjorner et al., 2007). This statistic has been shown to minimize Type I error when compared to item fit methods that contrast predicted and observed frequencies across levels of θ, which is an estimated value based on the tested latent construct (Orlando & Thissen, 2000). The $S\text{-}\chi^2$ item fit statistic was implemented by calling the IRTFIT macro (Bjorner et al., 2007) while using the SAS 9.2 statistical program.

Unidimensionality. This assumption was evaluated using tetrachoric correlations to obtain an interitem correlation matrix. Eigenvalues from this matrix were plotted in a scree plot and examined for indications of a dominant factor (Hambleton et al., 1991; Lord & Novick, 1968; Reise & Waller, 1990; Teresi, Kleinman, & Ocepek-Welikson, 2000). If the first eigenvalue is much larger than the second, and the second and all remaining eigenvalues are of similar magnitude, a single dominant dimension is indicated (Hambleton et al., 1991; Lord &

Novick, 1968). While Hattie (1985) points out that such an approach cannot be definitive due to a lack of empirically-determined guidelines for judging unidimensionality, Embretson and Reise (2000) indicate that it may be useful as evidence of single-factor dominance. Indeed, Lord and Novick (1968) state that the presence of a single common factor resulting from a tetrachoric item intercorrelation constitutes "a sufficient but not a necessary condition for the unidimensionality of the latent space" (Lord & Novick, 1968, p. 382). Tetrachoric correlations were obtained by conducting an exploratory factor analysis (EFA) using Mplus 4.0 (Muthen & Muthen, 2006). This statistical program uses tetrachoric correlation when implementing EFA with dichotomously scored responses (Muthen & Muthen, 2004.)

CFA is frequently used to test the IRT unidimensionality assumption, as well (Embretson & Reise, 2000). However, with so few items in the PC-PTSD there were insufficient degrees of freedom available to allow the use of this approach with this dataset. While a model with a single latent factor could be fit to the data, a two-factor model could not. This precluded the requisite model comparisons.

Local independence. No separate analyses verifying local independence were conducted in this study. This is because of the relationship between unidimensionality and local independence. The condition of local independence is said to be met whenever all latent factors influencing test performance have been specified. Such a circumstance can occur with either unidimensional or multidimensional data. However, since meeting the assumption of unidimensionality implies that there is only one dominant latent factor in the data, local independence is likewise satisfied (Hambleton et al., 1991; Lord & Novick, 1968; Reeve, 2000).

Parameter estimate invariance. While this, too, is an important assumption for IRT, no separate test of this assumption was conducted in this study. In fact, an IRT DIF study by definition looks for possible violations of this property. For example, one of the methods for testing the parameter invariance assumption is to compare the estimates of the item parameters obtained after giving the test of interest to more than one subgroup of the test's target population (Hambleton et al., 1991). These subgroups could be characterized by differences in ethnicity, past test performance, geography, or gender. If significant variation is found for any of the parameters across groups, it is judged that this assumption has been violated. Therefore, performing a separate parameter invariance test in a DIF study would be redundant.

A literature search yielded no sources that explicitly discussed this possible redundancy, but did reveal that the great majority of recent published IRT DIF studies make no mention of testing this assumption. However, M. O. Edelen (personal communication, March 4, 2010) confirmed that, indeed, a DIF study is a specific type of test for the parameter invariance property and that the assumption is rarely explicitly tested in modern IRT research. A major reason for this is that there is no accepted global test at this time, and thus it is only practical to conduct tests for specific instances where the researcher suspects the presence of a violation of the assumption.

Identifying anchor items. In LRT, each model also contains a set of anchor items assumed to contain no DIF and accordingly forced to have equal *a*s and *b*s. Since in each model comparison these anchor items are common to each model, they serve the purpose of placing the models on the same metric (Thissen et al., 1993). While multiple anchor items are typical and

lend increased DIF detection power, it has been shown that a single item can adequately comprise the "anchor set" (Thissen, Steinberg, & Wainer, 1988). Several methods have been suggested in order to determine which item or items to include in this internal anchor.

The Mantel-Haenzsel (MH) test can be used to identify items with DIF and subsequently remove those items from the anchor; leaving only those items demonstrating no MH-determined DIF in the anchor (Edelen et al., 2006; Thissen et al., 1993). Another suggested anchoring method is to use the LRT method itself in an iterative procedure to successively identify and remove items that show DIF. The iterative procedure would cease after the LRT method indicates no further DIF in any of the remaining potential anchor items (Thissen et al., 1993). However, this purification method can be very labor-intensive (Kim & Cohen, 1995) and may merely determine the presence of no more DIF than is common to all items (Thissen et al., 1993).

Alternatively, Flannery, Reise, and Widaman (1995) used regression modeling to identify anchor items and possible DIF items. The item parameters in question are first independently estimated for each group. Then for each freely estimated item parameter, the values for the focal group are regressed on the values for the reference group. The resulting standardized residual values are used to analyze the items. Large standardized residuals indicate possible DIF in an item, whereas small standardized residuals demonstrate that item's candidacy for inclusion in an anchor.

Another method approaches this problem graphically. Angoff's (1982) method proposed plotting the individual estimated parameters for the focal group against the same for the

reference group. When, for example, the difficulty on a particular item is nearly identical between two groups, such a plot will appear approximately linear with a correlation typically above .98. The researcher would inspect the plot and any items that appear as outliers or that depart from the line to a greater degree than the other items would be examined for DIF. Accordingly, those items that most closely approximate the linear relationship are candidates for anchoring.

Two anchor-identification methods were used in this study: Flannery and colleagues' (1995) regression method and Angoff's (1982) graphical method. Only items that were indicated as DIF-free by both methods were included in the anchor. The rationale was that using two methods would increase the confidence that the chosen anchor items could be assumed to have equal *a* and *b* parameters across gender groups. Together, these methods were used to identify at least one anchor item. The number of anchor items minimally necessary for detecting DIF is a critical issue for assessing DIF in the four-item PC-PTSD. A simulation study using LRT DIF methods found that one anchor item was sufficient to be effective in this regard—although more anchors were related to increased detection power (Wang & Yeh, 2003).

LRT DIF detection. This approach compares the goodness-of-fit of two models. In the augmented (A) model one or more item parameters were free to vary between the male and female groups. In the compact (C) model the item parameters that were free to vary in the A model were forced to be equal across the groups. Both models were estimated independently and a log-likelihood was obtained for each. Then, the value for the LRT statistic was obtained

by calculating negative two times the difference in the log-likelihoods (*LL*) of the C and A models.

$$G^2(d.f.) = -2(LL_C - LL_A)$$

In large samples, this statistic is distributed as a chi-square (χ^2). The degrees of freedom (*d.f.*) value is equal to the difference in the number of parameters in the A and C models (Camilli, 2006; Embretson & Reise, 2000; Thissen, 2001).

In order to determine DIF in both the *b* and *a* parameters individually in the context of a 2PL model, three nested models are compared hierarchically (Edelen et al., 2006; Reeve, 2000; Thissen, 2001). Model 1 (the most unconstrained model under this DIF detection protocol) permitted free estimation of both *b* and *a* parameters across groups. Model 2 constrained only the slope parameter *a* to equality across the compared groups, freely estimating the location parameter *b*. Model 3 (identical to model C above) constrained both the *b* parameter and the *a* parameter to equivalence.

The significance of any DIF in the *a* parameter was tested by calculating negative two times the difference in the log-likelihoods of Model 2 and Model 1 as in the formula above. Rejection of the null hypothesis is indicative of *a*-DIF. Significance of *b*-DIF was similarly determined using the same procedure for Model 3 and Model 2. The order of the model comparisons proceeded logically. First, because hypothesis testing regarding slopes are only meaningful in the context of equal guessing parameters or asymptotes, the first comparison was performed under the circumstances of assumed equality among *c* parameters across groups (as is always the case under the 2PL). Likewise, testing for location parameter equality makes sense

only under the condition of equivalent trace line slopes. Therefore, the assumption was that the *a* parameters are equal when testing for *b*-DIF (Thissen, 2001; Thissen et al., 1993). In both cases, the *d.f.* value was obtained by subtracting the number of parameters freely estimated in the augmented model minus the freely estimated parameters in the compact model. Since each successive model is nested with one additional parameter constrained to equality, for each model comparison the *d.f.* value was one (Thissen et al., 1993).

These hierarchical model comparison IRT methods used in this study are sensitive to sample size (Reeve, 2000). Due to this and the fact that multiple comparisons will be conducted, a conservative alpha (α) of .01 will be used for all IRT DIF hypothesis testing to protect against Type I error (Reeve, 2000; Teresi et al., 2000).

DIF interpretation. After the DIF detection process is complete, a final parameterization can be run using MULTILOG 7.0 that incorporates any identified, significant DIF. As discussed in Edelen et al. (2006), the resulting estimates can be used to aid in interpretation of the DIF. DIF magnitude is visually inferred by placing the item ICCs corresponding to a positive response for both groups on the same plot. Also, all item ICCs can be summed to create a separate test response function (TRF) for both males and females. Placing each group's TRFs on the same plot allows a visual indication of the magnitude of the effect that the item-level DIF had on the overall test total score. Excel can be used to create these plots using the 2PL equation.

External Evidence of Equivalent Validity by Gender

Logistic regression. Logistic regression was used to test whether the PC-PTSD was differentially predictive of a later clinical PTSD diagnosis. This predictive validity analysis was

performed multiple ways using various configurations of the PC-PTSD score as the IV: 1) a continuous variable using the raw score (0 to 4), 2) a dichotomous variable indicating whether or not the cutoff value of three was met, 3) a dichotomous cutoff of two, 4) a dichotomous cutoff of one (the trigger for further evaluation and referral in the USAF), and 5) a dichotomous variable using only the avoidance item (the avoidance-only screen). In each of these analyses, the DV was the dichotomous indicator of clinical PTSD diagnosis status. Differential prediction for men and women was assessed by testing for interactions between gender and each PC-PTSD scoring method in each model predicting PTSD diagnosis status.

One additional regression analysis was conducted to assess the divergent validity of the PC-PTSD. In this case the primary IV was the recommended PC-PTSD cutoff of three and the DV was the dichotomous MDD diagnosis status variable. Depression was chosen for the divergent validity comparison because it has been shown to be among the most highly comorbid conditions with PTSD (Kessler et al., 1995).

Link function. The log-log link was chosen for these data due to the strong positive skew of the frequency distribution of the PTSD status variable. Relatively few of the subjects were diagnosed with PTSD, so the data contain far more zeros (denoting no PTSD diagnosis) than ones (denoting the positive identification of PTSD). Under these circumstances, the log-log model is preferable due to its asymmetrical density function and a probability function that approaches one slowly while approaching zero rapidly. (Hardin & Hilbe, 2007; Simonoff, 2003). All regression analyses using the log-log model were conducted using the Stata 11.0 statistical program.

Modeling procedure. For each of the six logistic regression analyses mentioned above, the same modeling procedure was used as detailed in Orme and Combs-Orme (2009). Three IVs were entered simultaneously: the PC-PTSD score as configured for that particular analysis and sex as main effect IVs, and the cross-product of PC-PTSD score and sex as a moderator IV to test for interaction. If the interaction was found to be significant, the analysis stopped and the interaction was interpreted. If the interaction term was not significant, it was removed from the model and the model was run again with only the appropriate PC-PTSD score main effect IV.

Significance level. As with the DIF portion of the study, the use of multiple analyses led to concern for elevated risk of a Type I error. Therefore, a conservative alpha value of .01 was chosen for all regression analyses. Although it was expected that any detected differential prediction would result in an odds ratio for the interaction term that was greater than zero, two-tailed significance tests were used because a result in either direction was deemed important. For the same reason, two-tailed tests were used for any main effect models despite the expectation that higher scores and met thresholds on the instrument would be positively related to either diagnosis.

Model evaluation. Although there are no commonly accepted measures of overall model fit for logistic regression comparable to R^2 used in linear regression (Orme & Combs-Orme, 2009), each analyzed model was evaluated in a number of ways. Unstandardized deviance residuals were examined to check for poor model fit at the case level. The Cook's *D* (distance) values were plotted to determine whether any outliers exerted undue influence on the results.

Also, multicollinearity statistics were obtained on each model by running the model through PASW 17.0 (formerly SPSS) linear regression (Orme & Combs-Orme, 2009).

Diagnostic utility. This analysis compared the diagnostic accuracy of the PC-PTSD in identifying those that are at elevated risk for a PTSD diagnosis among both male and female deployers. It is important to note that the term "diagnostic," as used here in "diagnostic utility" is only used to signify the type of analysis used. Strictly, the PC-PTSD screen can only identify risk of PTSD or probable PTSD, not diagnose. Using the clinical PTSD diagnosis obtained from the medical record database as the "gold standard," the standard diagnostic utility statistics of sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic efficiency were calculated for each gender group. In addition, gender-specific positive and negative diagnostic likelihood ratios were obtained.

Diagnostic criterion. The use of a clinical diagnosis not explicitly obtained using a standardized, structured interview could be problematic. The reliability of psychiatric diagnoses using unstructured interviews has long been questioned (Beck, Ward, Mendelson, Mock, & Erbaugh, 1962) and structured diagnostic interviews have been shown to improve concordance rates among providers (Helzer et al., 1977). Prins et al. (2003) compared the clinical provider PTSD diagnosis obtained via examination of medical records to the gold standard of the CAPS structured clinical interview. The clinically-derived diagnosis matched CAPS results—indicating the subject was positive for PTSD—61% of the time while missing the diagnosis in 39% of cases. This equates to 61% sensitivity for the clinical diagnosis. MacGruder and colleagues' (2005) also used the CAPS with 745 veterans that agreed to a clinic interview out of

a random sample of 1198 veterans having a minimum of one visit to primary care during 1999. Medical record reviews yielded a sensitivity of 46.5% and specificity of 96.6% for the clinical PTSD diagnosis compared to the CAPS research identification.

Despite these problems, Ouimette, Wade, Prins, and Schohn (2008) used a clinical database similar to that proposed for use in the present study as the source for the reference diagnosis in their study comparing the performance of two PTSD screens. Due to the limitations of using this less reliable PTSD identification method, these researchers deemphasized findings related to optimal cutoff values that could be compared to similar results in other studies. Instead, they primarily analyzed the differences between the screens with regard to prediction of PTSD. Likewise, the purpose of the proposed study was to compare the relative performance of the PC-PTSD with males and females. The non-research, clinician diagnosis should therefore be sufficient, though not optimal, for making this comparison.

Sensitivity and specificity. Sensitivity, also known as the true positive rate, is the probability of those with the tested condition receiving a positive result on the test in question. Specificity, on the other hand, consists of the proportion of examinees without the condition manifesting a negative test result. Specificity is also known as the true negative rate (Akobeng, 2006a; Baldessarini, Finklestein, & Arana, 1983).

Predictive values. Predictive values are used to provide the probability that the results of the test directly correspond with the actual diagnostic status. Positive predictive value (*PPV*) is found by dividing the number of true positives by all those who tested positive for the condition. Negative predictive value (*NPV*) is similarly calculated with the true negatives in the numerator

and the total of those testing negative in the denominator (Akobeng, 2006a; Baldessarini et al., 1983).

Diagnostic efficiency. A related statistic, diagnostic efficiency or power, can also be calculated by dividing the combined true positive and true negative outcomes by the total of all of those tested. However, this statistic tends to overestimate a given test's power (Baldessarini et al., 1983).

Diagnostic likelihood ratio. While *PPV* and *NPV* are important indicators of test performance, they will vary according to the prevalence of the condition in a particular population (Akobeng, 2006a). Alternatively, a diagnostic likelihood ratio (*DLR*) can be obtained that combines both sensitivity and specificity (Akobeng, 2006b). This measure has a couple of important advantages. The first is that the DLR is independent from the prevalence of a condition within a population (Furukawa & Goldberg, 1999; Furukawa, Goldberg, Rabe-Hesketh, & Ustun, 2001; Pepe, 2003; Zhou, Obuchowski, & McClish, 2002). Pepe (2003) explains that this independence is derived from the fact that the *DLR* is merely a function of probabilities for diagnostic classification. A second advantage involves reduced sensitivity to variation in population factors such as comorbidity and severity of the index condition (Furukawa & Goldberg, 1999; Furukawa et al., 2001). This was a useful property for the current study because, as mentioned in the above literature review, comorbidity was speculated to be a factor in the observed gender difference in diagnostic efficiency reported in the Prins et al. (2003) PC-PTSD validation study. Within the research context, the *DLR* measure has previously been used to compare the relative diagnostic utility of a screen across populations (Furukawa,

Andrews, & Goldberg, 2002), between two PTSD screens within a single population (Ouimette et al., 2008), and between two versions of the same health screen across multiple care centers (Furukawa et al., 2001).

Also, in clinical practice *DLR*s can be combined with population-specific prevalence information (also known as the pre-test probability of having the condition) using a straightforward procedure in order to estimate post-test probabilities to aid clinicians in the diagnostic decision process (Akobeng, 2006b; Sackett & Straus, 1998). This approach is more common among medical providers than among behavioral health providers. However, as mentioned above, a positive PC-PTSD screen in the USAF can result in referral to either type of clinician.

Likewise, the raw value of the *DLR* can be helpful in interpreting the results for a particular respondent. For instance, a *DLR* for a positive test with a value greater than 10 indicates a high likelihood that that tested condition is present in a respondent that tests positive. Conversely, if the *DLR* for a positive test is less than 0.1, a respondent who has tested positive is very unlikely to have the index condition (Akobeng, 2006b; Ouimette et al., 2008).

There are two types of *DLR*s: the likelihood ratio for a positive test (*DLR*+) and the likelihood ratio for a negative test (*DLR*-). The *DLR*+ is defined as the probability of a person with the index condition testing positive divided by the probability of a person without the condition testing positive. In similar fashion, the *DLR*- is the probability of a person with the condition testing negative divided by the probability of a person without the condition receiving a negative test result (Akobeng, 2006b,). Both formulae are given as follows:

$$DLR+ = \frac{Sensitivity}{1 - Specificity}$$

$$DLR- = \frac{1 - Sensitivity}{Specificity}$$

Once the *DLR*s were obtained for both males and females under each tested PC-PTSD cutoff score and the single-item avoidance screen, the corresponding male and female *DLR*s were tested to determine whether they were significantly different from one another. This was accomplished using confidence intervals (*CI*), which are calculated as follows for both *DLR*s:

$$DLR+\ CI\text{: } \exp\left(\ln\left(\frac{Sens}{1 - Spec}\right) \pm z_{1-\alpha/2}\sqrt{\frac{1 - Sens}{s_1} + \frac{Spec}{r_1}}\right)$$

$$DLR-\ CI\text{: } \exp\left(\ln\left(\frac{1 - Sens}{Spec}\right) \pm z_{1-\alpha/2}\sqrt{\frac{Sens}{s_0} + \frac{1 - Spec}{r_0}}\right)$$

where *Sens* and *Spec* represent sensitivity and specificity, respectively, and *z* is the (1 - α/2)th percentile of the standard normal distribution (Zhou et al., 2002). For the *DLR*+ *CI*, s_1 stands for the number of true positives and r_1 is the number of false positives. For the *DLR*- *CI*, s_0 corresponds to the number of false negatives and r_0 denotes the number of true negatives.

Again, a conservative *p*-value of .01 was used to protect against the increased risk of Type I error inherent with multiple comparisons. Since the *DLR* is theoretically independent of prevalence, if the 99% *CI* shows that a particular set of male/female *DLR*s are statistically significantly different, this would be another indication that the PC-PTSD is performing

differently based on gender. All diagnostic utility statistics were computed using Microsoft Excel.

Procedure. The gender-specific diagnostic utility statistics were calculated for cutoffs of one, two, three, and four, as well as for the single item avoidance screen suggested by Bliese et al. (2008). This analysis used the same sample as that used in the logistic regression analyses above. Separate cross tabulations were created for both sexes comparing each studied dichotomous cutoff score of the PC-PTSD with PTSD diagnostic status. This resulted in a total of ten contingency tables (five per gender) that supplied the data necessary for calculating all of the required diagnostic utility statistics.

Validity using CFA.

The final analysis examined the convergent and divergent validity of the PC-PTSD using CFA. This analysis did not assess for any gender differences in the PC-PTSD, but instead examined the overall validity of the PC-PTSD with a large group of the target population for the PDHRA screening: returning USAF deployers. A two-factor model was tested with the four PC-PTSD items loading on one factor (Factor 1 - PTSD) and the two-item depression screen used in the PDHRA loading on the other factor (Factor 2 – depression). Due to the high comorbidity of PTSD and depression (Kessler et al., 1995), Factors 1 and 2 were allowed to correlate in the model. As in the logistic regression analysis, depression was chosen as the criterion for measuring discriminant validity due to the high comorbidity between that diagnosis and PTSD.

The sample for the CFA analysis ($N = 6963$; males = 5756, females = 1207) was a subset of the sample used for the logistic regression and diagnostic utility analyses ($N = 6999$). The slightly smaller sample size was the result of 36 cases from the larger sample with missing values on at least one of the two depression queries (0.5%).

Depression screen. The two-item depression screen used in the PDHRA consists of two items from the Patient Health Questionnaire (PHQ; Spitzer, Kroenke, & Williams, 1999). The screen as implemented in the PDHRA asks, "Over the past month, have you been bothered by the following problems?" The two items are, "Little interest or pleasure in doing things," and "Feeling down, depressed, or hopeless." Responses to each item are given on a four-category Likert scale ("Not at all," "Few or several days," "More than half the days," or "Nearly every day"). These responses were respectively scored 0, 1, 2 and 3.

The two-item PHQ has been shown to have favorable diagnostic characteristics in a couple of validation studies with relatively large sample sizes (Arroll, Goodyear-Smith, Kerse, Fishman, & Gunn, 2005; Whooley, Avins, Miranda, & Browner, 1997). In fact, Whooley et al. (1997) demonstrated that this two-question screen performed about as well at detecting depression as accepted depression questionnaires with many more items, including the abbreviated versions of the Beck Depression Inventory (BDI; Beck & Beck, 1972) and the Center for Epidemiologic Studies Depression Scale (CES-D; Andersen, Malmgren, Carter, & Patrick, 1994). The purpose of including this brief depression scale in the CFA analysis is not to validate yet another screen included in the PDHRA. Rather, the two-item PHQ is included in the model in order to aid in the testing of convergent and divergent validity within the PC-PTSD.

Procedure. The CFA analysis was conducted using Mplus 4.0 (Muthen & Muthen, 2006). All path weights were freely estimated, no error terms were allowed to correlate, and the variance of both latent factors was fixed at a value of 1 in order to fix their metric. Also, no cross-loadings were permitted. Robust weighted least squares is the default estimator for CFA involving factor indicators with categorical responses in Mplus, and this option was left unchanged for this analysis. Also, as is appropriate for dichotomous and polytomous indicators, tetrachoric or polychoric correlation methods were used (Brown, 2006; Muthen & Muthen, 2004).

Model fit was evaluated using both the root mean square error of approximation (*RMSEA*) and the comparative fit index (*CFI*). Recommendations for the *RMSEA* provide information at multiple levels: Values of .06 or lower are indicative of a well-fitting model with large samples, around .08 indicates a medium degree of fit, and .10 or higher suggests less than adequate fit. For the *CFI*, good fit is indicated by values near or above .95 (Byrne, 2001).

Convergent validity was assessed by examining the extent to which the indicators loaded onto the expected factors. Divergent or discriminant validity was judged using the correlation between the latent factors. Discriminant validity is considered adequate when this correlation is less than or equal to .80 (or .85) (Brown, 2006). A conventional $\alpha = .05$ was used to determine whether estimates were significantly different from zero.

CHAPTER IV: RESULTS

Sample Characteristics

Each of the three samples used in this study were analyzed to determine the degree to which the male and female comparison groups were similar on available background characteristics. For continuous characteristic variables an independent samples *t*-test was used as implemented in PASW. With categorical variables, in most cases the analyses were performed using Pearson's chi-square independence test (also within PASW). However, there were some tables larger than two-by-two that contained cells with expected frequency values that were smaller than five. In such instances, the SAS program was used to conduct Fisher's exact test. Due to the large sample sizes and alpha level of .01 was chosen. All probabilities were two-tailed.

IRT DIF Sample

All the results of the sample characteristic comparisons are included in Table 1. Within this sample, there were no differences detected between men and women with regard to the number of months they were deployed, the deployment location (Iraq, Afghanistan, or both), or the categories of rank which they had attained (enlisted, officer, or other). However, several statistically significant differences were detected. Males had higher mean age ($M = 30.88$, $SD = 7.82$; $t(9806) = 4.85$, $p < .001$) than females ($M = 29.82$, $SD = 7.56$), as well as a larger mean total number of deployments ($M_{male} = 1.87$, $SD = 1.42$; $M_{female} = 1.45$, $SD = 1.14$; $t(9806) = 12.16$, $p < .001$). Also, the mean number of days between the return from deployment and

Table 1

Sample Characteristics for the Differential Item Functioning Sample

Characteristic Variables	Males (N = 8381)	Females (N = 1425)	Test Results
Age (mean ± *SD*)	30.88 ± 7.82	29.82 ± 7.56	$t(9806) = 4.85$, $p < .001$
Months deployed (mean ± *SD*)	5.90 ± 3.79	5.75 ± 2.65	$t(9806) = 1.84$, $p = .066$
Total number of deployments (mean ± *SD*)	1.87 ± 1.42	1.45 ± 1.14	$t(9806) = 12.16$, $p < .001$
Days between deployment and screen (mean ± *SD*)	119.68 ± 24.72	117.18 ± 23.72	$t(9806) = 3.60$, $p < .001$
Deployment location, *N* (%)			$\chi^2(2, N = 9806) = 2.87$, $p = .238$
Iraq	6273 (74.8)	1070 (75.1)	
Afghanistan	1984 (23.7)	342 (24.0)	
Both	124 (1.5)	13 (.9)	
Marital status at time of screening, *N* (%)			$p < .001$, Fisher's exact test
Never married	2475 (29.5)	478 (33.5)	
Married	5199 (62.0)	658 (46.2)	
Separated	51 (.6)	19 (1.3)	
Divorced	650 (7.8)	268 (18.8)	
Widowed	6 (.1)	2 (.1)	
Rank category, *N* (%)			$\chi^2(2, N = 9803*) = 6.56$, $p = .038$
Enlisted	7001 (83.6)	1152 (80.9)	
Officer	1344 (16.0)	267 (18.8)	
Other	34 (.4)	5 (.4)	
Military status prior to deployment, *N* (%)			$p < .001$, Fisher's exact test
Active duty	7384 (88.1)	1329 (93.3)	
Reserve	144 (1.7)	20 (1.4)	
National Guard	717 (8.6)	70 (4.9)	
Civilian	119 (1.4)	5 (.4)	
Other	17 (.2)	1 (.1)	

Note. N = 9806. All *p* are two-tailed. *SD* = standard deviation.

*Missing data in rank field for 3/9806 cases (.03%).

the completion of the PDHRA screen was somewhat higher for men (M = 119.68, SD = 24.72; $t(9806) = 3.60$, $p < .001$) than women (M = 117.18, SD = 23.72). Statistically significant differences were also found for marital status at the time of taking the screen ($p < .001$, Fisher's exact test), with the largest disparities occurring those identifying themselves as married (62.0% for men compared to 46.2% of women) and divorced (18.8% among women and 7.8% of men). Women also tended to more often identify their military status prior to their deployment ($p < .001$, Fisher's exact test) as "active duty" (93.3% compared to 88.1% of men) while men had a higher proportion of national guard (8.6% vs. 4.9% among females).

These results indicate that the male and female comparison groups were not equivalent on certain background characteristics. It is likely, however, that the identified statistically significant differences have varying levels of practical significance—especially considering the large sample size. For example, when the average number of days between returning from deployment and completing the post-deployment screening approaches four months for both genders, it is not clear whether an average difference of less than three days would have much actual impact on PC-PTSD responses. In contrast, the marital status differences appear much more substantial with the proportion of divorced women being more than double the proportion of divorced men.

Regression and Diagnostic Utility Sample

See Table 2 for detailed information on each sample characteristic variable. All background variables for this sample were the same as for the IRT sample, except for one. Since

Table 2

Sample Characteristics for the Logistic Regression/Diagnostic Utility Sample

Characteristic Variables	Males (N = 5785)	Females (N = 1214)	Test Results
Age (mean ± *SD*)	30.63 ± 7.44	29.68 ± 7.49	$t(6999) = 2.90$, $p = .004$
Months deployed (mean ± *SD*)	6.24 ± 4.03	5.89 ± 2.63	$t(6999) = 3.75$, $p < .001$
Total number of deployments (mean ± *SD*)	1.84 ± 1.39	1.41 ± 1.11	$t(6999) = 11.72$, $p < .001$
Days between deployment and screen (mean ± *SD*)	118.00 ± 23.52	116.19 ± 23.39	$t(6999) = 2.44$, $p = .015$
Days between screen and diagnostic visit (mean ± *SD*)	53.35 ± 27.50	60.70 ± 25.06	$t(6999) = -9.14$, $p < .001$
Deployment location, *N* (%)			$\chi^2(2, N = 6999) = 3.26$, $p = .196$
Iraq	4338 (75.0)	917 (75.5)	
Afghanistan	1361 (23.5)	287 (23.6)	
Both	86 (1.5)	10 (.8)	
Marital status at time of screening, *N* (%)			$p < .001$, Fisher's exact test
Never married	1693 (29.3)	395 (32.5)	
Married	3589 (62.0)	572 (47.1)	
Separated	43 (.7)	16 (1.3)	
Divorced	457 (7.9)	230 (18.9)	
Widowed	3 (.1)	1 (.1)	
Rank category, *N* (%)			$\chi^2(2, N = 6998*) = 7.25$, $p = .027$
Enlisted	4856 (83.9)	981 (80.9)	
Officer	902 (15.6)	227 (18.7)	
Other	27 (.5)	5 (.4)	
Military status prior to deployment, *N* (%)			$p = .703$, Fisher's exact test
Active duty	5642 (97.5)	1193 (98.3)	
Reserve	40 (.7)	7 (.6)	
National Guard	90 (1.6)	13 (1.1)	
Civilian	10 (.2)	1 (.1)	
Other	3 (.1)	0 (0.0)	

Note. N = 6999. All p are two-tailed. *SD* = standard deviation.

*Missing data in rank field for 1/6999 cases (.01%).

this sample was the only sample used to compare the screen results to a diagnosis obtained during a clinical medical visit, a variable was added totaling the days between the date of the screen and the date of the diagnostic visit.

Similar to the IRT sample, of which this sample was a subset, no statistically significant differences were identified between groups of men and women on deployment location or rank classification. One difference, though, was that in this sample no difference was detected across gender with respect to the mean number of days between returning from deployment and the screening date. On average, though, males spent more months deployed ($M_{male} = 6.24$, $SD = 4.03$; $M_{female} = 5.89$, $SD = 2.63$; $t(6999) = 3.75$, $p < .001$), which was also a change from the IRT sample. Men also tended to be older ($M_{male} = 30.63$, $SD = 7.44$; $M_{female} = 29.68$, $SD = 7.49$; $t(6999) = 2.90$, $p = .004$), and to have been deployed more often ($M_{male} = 1.84$, $SD = 1.34$; $M_{female} = 1.41$, $SD = 1.11$; $t(6999) = 11.72$, $p < .001$). Women, on the other hand, had a higher mean number of days between screen completion and the diagnostic medical visit ($M_{female} = 60.70$, $SD = 25.06$; $t(6999) = -9.14$, $p < .001$) when compared to men ($M_{male} = 53.35$, $SD = 27.50$). There was again a statistically significant relationship between gender and marital status ($p < .001$, Fisher's exact test), with males more likely to be married (62.0% of men; 47.1% of women) and females more likely to be divorced (18.9% among women and 7.9% of men).

An interesting difference between this sample and the IRT sample is that in this instance no statistically significant relationship was detected between gender and military status prior to deployment ($p = .703$, Fisher's exact test). A closer inspection of the proportions in each category for both samples reveals that the percentage of male and female in the active duty

classification increased from that in the IRT sample (93.3% and 88.1%, respectively) to around 98% for both genders (97.5% for men; 98.3% for women) in the current sample. This change in proportion probably occurred because of the requirement that all cases in the regression/diagnostic utility sample have a military medical diagnostic visit within 90 days after the date of the PDHRA screen. Immediately after deployment, most non-active duty members of the military return to local communities that may or may not be near a military facility large enough to have military medical services. For example, while an Air National Guard base may have some limited medical capabilities, many of the guard members attached to that base tend to reside hours from the installation and typically only come to the base one weekend per month. This is in contrast to active duty military personnel who, after deployment, almost exclusively return to live locally at or near a military base that houses its own military medical facility. It is therefore likely that non-active duty subjects had a much lower frequency of attaining military medical care in the 90 days after the PDHRA screen, and so were disproportionately excluded from this sample.

CFA Sample

Table 3 contains the gender breakdown of the sample characteristics for the CFA sample. The pattern of background differences (and lack of differences) that emerged across gender groups for this sample was the same as that found for the regression/diagnostic utility sample. This was expected, to some extent, because the CFA sample was attained by removing 36 cases from the former sample due to missing responses on the depression screen used in the CFA analysis. As in the regression/diagnostic utility sample, no statistically significant differences

Table 3

Sample Characteristics for the Confirmatory Factor Analysis Sample

Characteristic Variables	Males (N = 5756)	Females (N = 1207)	Test Results
Age (mean ± *SD*)	30.35 ± 7.43	29.69 ± 7.49	$t(6963) = 2.80$, $p = .005$
Months deployed (mean ± *SD*)	6.23 ± 4.03	5.90 ± 2.63	$t(6963) = 3.58$, $p = .000$
Total number of deployments (mean ± *SD*)	1.84 ± 1.39	1.41 ± 1.11	$t(6963) = 11.58$, $p = .000$
Days between deployment and screen (mean ± *SD*)	118.02 ± 23.50	116.14 ± 23.38	$t(6963) = 2.53$, $p = .012$
Deployment location, *N* (%)			$\chi^2(2, N = 6963) = 2.85$, $p = .241$
Iraq	4318 (75.0)	911 (75.5)	
Afghanistan	1355 (23.5)	286 (23.7)	
Both	83 (1.4)	10 (.8)	
Marital status at time of screening, *N* (%)			$p < .001$, Fisher's exact test
Never married	1685 (29.3)	392 (32.5)	
Married	3570 (62.0)	570 (47.2)	
Separated	43 (.7)	16 (1.3)	
Divorced	455 (7.9)	228 (18.9)	
Widowed	3 (.1)	1 (.1)	
Rank category, *N* (%)			$\chi^2(2, N = 6962^*) = 7.76$, $p = .021$
Enlisted	4832 (84.0)	974 (80.8)	
Officer	897 (15.6)	227 (18.8)	
Other	27 (.5)	5 (.4)	
Military status prior to deployment, *N* (%)			$p = .586$, Fisher's exact test
Active duty	5615 (97.6)	1187 (98.3)	
Reserve	38 (.7)	7 (.6)	
National Guard	90 (1.6)	12 (1.0)	
Civilian	10 (.2)	1 (.1)	
Other	3 (.1)	0 (0.0)	

Note. $N = 6963$. All p are two-tailed. *SD* = standard deviation.

*Missing data in rank field for 1/6963 cases (.01%).

were uncovered between males and females on the average number of days between the return from deployment and the screen date, the location of deployment, rank category, or military status before deployment. Observed gender differences did rise to statistical significance with regard to age (M_{male} = 30.35, SD = 7.43; M_{female} = 29.69, SD = 7.49; $t(6963) = 2.80$, $p = .005$), months deployed (M_{male} = 6.23, SD = 4.03; M_{female} = 5.90, SD = 2.63; $t(6963) = 3.58$, $p = .000$), total number of deployments (M_{male} = 1.84, SD = 1.39; M_{female} = 1.41, SD = 1.11; $t(6963) = 11.58$, $p = .000$), and marital status ($p < .001$, Fisher's exact test).

Measurement Equivalence by Gender

Prior to detailing the findings of this analysis, the presentation will begin with the results of the assumption testing and model fit analyses. This will be followed by the search for suitable anchor items.

Model Fit Analyses

Model comparison. The nested model comparison between fitting the 2PL versus the 1PL to the data was statistically significant ($G^2(4) = 36.3$, $p < .001$). This indicated that, in a relative sense, the 2PL fit the data better than the 1PL. As expected, a similar comparison between the 3PL and 2PL models was not statistically significant ($G^2(4) = 2.3$, $p = .681$), meaning that fitting the 3PL to the data did not result in a significantly better fit than the 2PL. Therefore, the 2PL model was used in all subsequent IRT analyses of the study data.

Item fit. Using alpha equal to .05, two-tailed, the item fit analysis using IRTFIT (Bjorner et al., 2007) indicated misfit for two items: item 2/avoidance ($S\text{-}\chi^2(1) = 3.92$, $p = .048$) and item 3/arousal ($S\text{-}\chi^2(1) = 8.17$, $p = .004$). Items 1 and 4, corresponding respectively to

reexperiencing and numbing symptoms, showed adequate fit (item 1: $S\text{-}\chi^2(1) = 3.73$, $p = .054$; item 4: $S\text{-}\chi^2(1) = 1.32$, $p = .250$). However, at this level of α item 1 was very near to exhibiting lack of fit. These results suggested that the 2PL model does not fit well with these data (at least for two of the items).

As pointed out in Orlando and Thissen (2000), the sensitivity of this chi-square statistic is likely to increase with larger samples. Since their study was conducted with a sample size of 1000, however, it was unclear to what degree this sensitivity would have affected the item fit analysis results of the current study ($N = 9806$). To date, this statistic has yet to be tested in a simulation study with sample sizes in this range (M. O. Edelen, personal communication, April 26, 2010). Although the lack of fit indicated by the item fit analysis is a concern, the possibility cannot be ruled out that the statistic is oversensitive to misfit—that is, identifies miniscule inconsistencies as problematic—at this sample size. Using an alpha level of .01 to attempt to account for the effect of sample size resulted in an indication of misfit for only item 3.

Assumption Testing

Only the unidimensionality assumption was explicitly tested. As detailed in the chapter on methods, the assumption of local independence is satisfied if the unidimensionality assumption is met. Also, the invariance assumption was not assessed because there is no general test for parameter estimate invariance.

The EFA using tetrachoric correlations was indicative of unidimensionality in the PC-PTSD response data. Eigenvalues were obtained from the inter-item tetrachoric correlation matrix. The first eigenvalue (3.37) was much larger than the remaining three (.27, .25, .11),

which constitutes evidence consistent with single-factor dominance and, therefore, unidimensionality (Hambleton et al., 1991; Lord & Novick, 1968). The scree plot of the eigenvalues can be inspected in Figure 2.

Anchor Identification

Regression method. The standardized residuals resulting from the regression of separately estimated male and female parameters onto one another indicated that items 3 (arousal) and 4 (numbing) were likely candidates to serve as anchors during the IRT DIF procedure. Item 4 had the lowest standardized residual for the *a* parameter (.29) while that for item 3 was lowest among all the *b* parameters (.30).

Conversely, items 1 (reexperiencing) and 2 (avoidance) were identified as suspected DIF items given their relatively high standardized residuals. The highest *a* parameter standardized residual was observed on item 1 (1.00) and the highest *b* parameter residual occurred on item 2 (1.02). Therefore, Flannery and colleagues' (1995) anchor-identification method suggested that items 3 and 4 should comprise the anchor for this DIF analysis, with items 1 and 2 the targets of the analysis. As mentioned previously, throughout the remainder of the IRT DIF analysis the working assumption will be that the anchor items are free of DIF.

Graphical method. Next, the separate *a* and *b* parameter estimates for each gender group were plotted against one another. Item 1 had the greatest deviation from the fit line on the plot for *a*, while item 4 was the closest to the same fit line. Likewise, on the *b* plot item 2 was the most deviant from the fit line and item 3 was the least. Thus, items 3 and 4 were

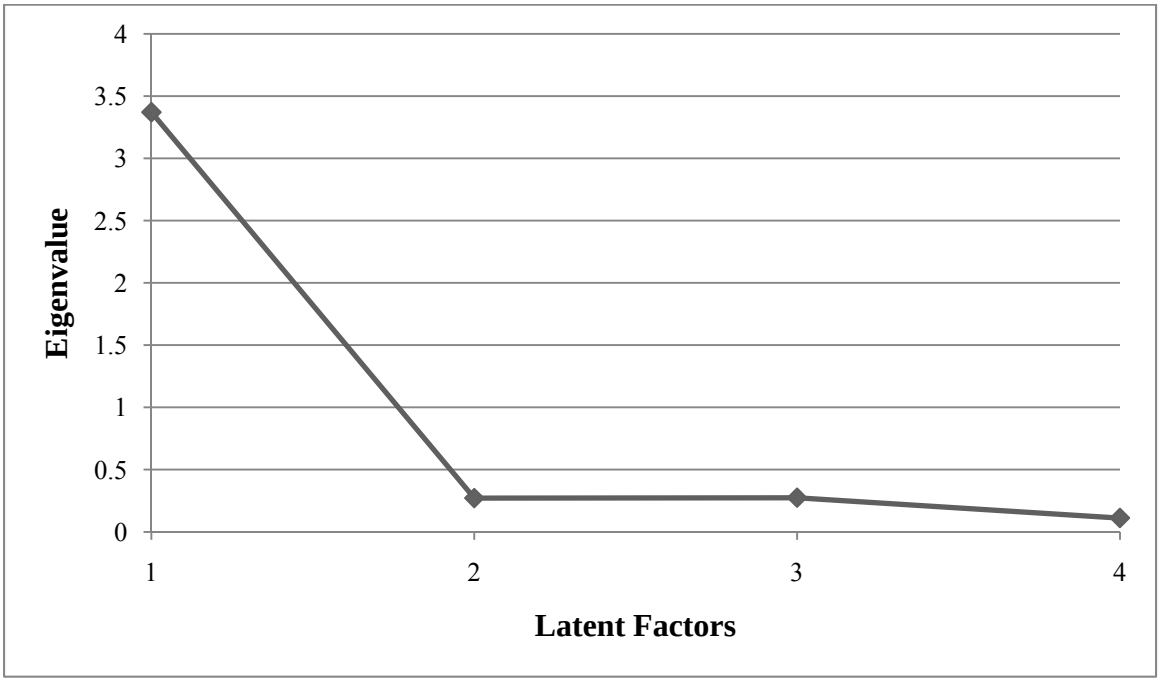

Figure 2. Scree plot of the eigenvalues for testing the unidimensionality assumption with the sample used for IRT DIF analysis.

recommended as anchors and items 1 and 2 are candidates for DIF. These results confirmed those obtained using the regression method.

IRT DIF Results

MULTILOG initially failed to achieve convergence in its estimation of the item parameters for this analysis, yielding unexpected negative chi-square difference values for G^2. However, convergence was achieved by raising the maximum number of estimation cycles in the program.

As noted in Table 4, no significant DIF was detected in either items 1 or 2 of the PC-PTSD using the model comparison method. In the likelihood ratio comparison of Model 2 and Model 1 assessing DIF in the *a* parameters for both items 1 ($G^2(1) = 0$, $p = 1.000$) and 2 ($G^2(1) = 0$, $p = 1.000$), the null hypothesis of no DIF could not be rejected. Similarly, the analyses for *b* parameter DIF also did not allow for rejection of the null hypothesis (item 1: $G^2(1) = .1$, $p = .752$; item 2: $G^2(1) = 0$, $p = 1.000$).

External Evidence of Equivalent Validity by Gender

Logistic Regression

In all logistic regression models, males were coded as 0 and females were coded as 1. Before summarizing the logistic regression results, the discussion begins with a review of the model evaluation analyses.

Model evaluation. According to Menard (2002), when assessing model fit a researcher should be concerned about unstandardized deviance residual values with an absolute value

Table 4

Item Parameters for Male and Female Subgroups on the PC-PTSD

Item	Content	Males		Females		Tests for DIF: G^2 (p)	
		a(*SE*)	*b*(*SE*)	*a*(*SE*)	*b*(*SE*)	*a*-DIF	*b*-DIF
	Suspected DIF items						
1	Reexperiencing	3.99 (0.23)	1.51 (0.03)	4.02 (0.48)	1.52 (0.06)	0 (1.000)	.1 (.752)
2	Avoidance	5.13 (0.37)	1.59 (0.03)	5.37 (0.78)	1.58 (0.05)	0 (1.000)	0 (1.000)
	Anchor items						
3	Arousal	3.03 (0.13)	1.45 (0.03)	3.03 (0.13)	1.45 (0.03)	--	--
4	Numbing	3.26 (0.17)	1.61 (0.03)	3.26 (0.17)	1.61 (0.03)	--	--

Note. For anchor items, each parameter constrained to equality across males and females. N = 9806 (Males = 8381, Females = 1425). Each value of G^2 given is at one degree of freedom. PC-PTSD = Primary Care Posttraumatic Stress Disorder screen; DIF = differential item functioning; a = the item difficulty or location parameter; b = the item discrimination or slope parameter; *SE* = standard error.

greater than 2 and, especially, 3. If Cook's *D* is larger than 1, an outlying case may be exerting undue influence on the regression coefficient (Orme & Combs-Orme, 2009). Finally, a couple of indicators of problematic multicollinearity among the IVs were used: tolerance and the variance inflation factor (VIF). Tolerance is indicative of possible multicollinearity problems at values lower than .1 (Norusis, 2006; Orme & Combs-Orme, 2009). The threshold indicating problematic multicollinearity for VIF, on the other hand, is for values greater than 10 (Orme & Combs-Orme, 2009).

Inspection of residuals. Among the five regression analyses using PTSD diagnosis as the DV, all five had some unstandardized deviance residuals that were just above 3. In each instance, the cases in question were those where the particular PC-PTSD score for that analysis equaled zero but the PTSD diagnosis variable equaled 1. This is, perhaps, not surprising given that just having the PTSD diagnosis variable equal to 1 was very unusual in this sample. Also, as will be discussed in the next section, the PC-PTSD score significantly predicted a PTSD diagnosis. This means that it would be even more unusual, in this sample, for a subject to be diagnosed with PTSD but not have that risk reflected in their PC-PTSD score.

Analysis 1 (total PC-PTSD score) and Analysis 4 (cutoff ≥ 1) both had cases with unstandardized residuals higher than 2 (but lower than 3). Again, in both instances the PTSD diagnosis variable equaled 1. The cases in this range for Analysis 1 were those where the total PC-PTSD score was either 1 or 2. For Analysis 4, these were the cases where cutoff ≥ 1 was scored 1. Again, since positive responses to both of these variables were relatively scarce in this

sample (the great majority of scores on both were zero), these would be unusual cases in the model.

The model that used MDD diagnosis as the DV, Analysis 6 (cutoff ≥ 3), also had unstandardized deviance residuals higher than the thresholds of 3 and 2. The pattern was similar to that found in the other analyses. Cases scored at 0 for cutoff ≥ 3 and 1 for MDD diagnosis had residuals above 3, whereas cases with both cutoff ≥ 3 and MDD diagnosis equal to 1 exhibited residuals in the range of 2 to 3.

Cook's D. No case in any of the analyses had a Cook's *D* that approached 1.0. In fact, the highest value among all analyses was .05. For all but one analysis, the highest strata of Cook's *D* values corresponded to the same cases that were associated with unstandardized deviance residuals greater than 3. The exception was Analysis 6, for which the highest Cook's *D* values were those associated with residuals between 2 and 3.

Multicollinearity. Inspection of the tolerance and VIF values revealed that there were no problems with excessive multicollinearity. Of course, multicollinearity diagnostics were only run for models that included more than one IV.

Results of analyses. Table 5 details the interaction models for each of the six analyses. Table 6 summarizes the main effect models. The final model in both tables is the divergent validity regression analysis using MDD rather than PTSD as the DV.

Interaction models. As is shown in Table 5, within each analysis the overall interaction model was statistically significant as indicated by the omnibus likelihood ratio test. However, none of the interaction terms in the initial models for each analysis were statistically significant.

Table 5

Interaction Analyses of the PC-PTSD Using Log-Log Regression

Independent Variables	Dependent Variable	Odds Ratio	*z*	Likelihood Ratio Test
Analysis 1	PTSD			381.97 @ 3 *d.f.****
Total PC-PTSD		1.54	15.26***	
Sex		1.09	1.09	
Total PC-PTSD X Sex		.93	-1.38	
Analysis 2	PTSD			245.38 @ 3 *d.f.****
Cutoff ≥ 3		3.31	12.19***	
Sex		1.05	.68	
Cutoff ≥ 3 X Sex		.97	-.15	
Analysis 3	PTSD			308.63 @ 3 *d.f.****
Cutoff ≥ 2		2.91	14.44***	
Sex		1.04	.51	
Cutoff ≥ 2 X Sex		.97	-.22	
Analysis 4	PTSD			307.80 @ 3 *d.f.****
Cutoff ≥ 1		2.65	14.58***	
Sex		1.17	1.67	
Cutoff ≥ 1 X Sex		.84	-1.23	
Analysis 5	PTSD			225.82 @ 3 *d.f.****
Avoidance item		2.88	12.46***	
Sex		1.09	1.23	
Avoidance X Sex		.88	-.74	
Analysis 6	MDD			37.26 @ 3 *d.f.****
Cutoff ≥ 3		1.72	4.11***	
Sex		1.32	3.50***	
Cutoff ≥ 3 X Sex		.96	-.16	

Note. Males coded as 0, females coded as 1. Odds ratio obtained from exp*(B)* in output for given model. *N* = 6999 (Males = 5785, Females = 1214). PC-PTSD = Primary Care Posttraumatic Stress Disorder screen; X = cross for crossproduct interaction term; *d.f.* = degrees of freedom.

****p*<.001

Table 6

Main Effect Analyses of the PC-PTSD Using Log-Log Regression

Scoring Method	Odds Ratio (99% *CI*)	*z*	Likelihood Ratio Test
DV: PTSD diagnosis			
Total PC-PTSD	1.52 (1.42-1.61)	17.14***	379.86 @ 1 *d.f.***
Cutoff ≥ 3	3.30 (2.66-4.10)	14.15***	244.92 @ 1 *d.f.***
Cutoff ≥ 2	2.90 (2.45-3.42)	16.46***	308.37 @ 1 *d.f.***
Cutoff ≥ 1	2.56 (2.20-2.97)	16.16***	305.14 @ 1 *d.f.***
Avoidance item	2.80 (2.32-3.38)	14.04***	224.29 @ 1 *d.f.***
DV: MDD diagnosis			
Cutoff ≥ 3	1.72 (1.31-2.26)	5.14***	24.29 @ 1 *d.f.***

Note. Males coded as 0, females coded as 1. Odds ratio obtained from value of exp(*B*) in statistical program output for given model. *N* = 6999 (Males = 5785, Females = 1214). PC-PTSD = Primary Care Posttraumatic Stress Disorder screen; *CI* = confidence interval; DV = dependent variable; *d.f.* = degrees of freedom.

***$p<.001$

This indicates that no interactions were detected between PC-PTSD score and gender. Thus, there was no evidence that the PC-PTSD was differentially predicting subsequent PTSD for males versus females. This was true for all configurations of PC-PTSD score.

Main effect models. Since no interactions were found, each analysis was modeled again using only the PC-PTSD score as a predictor variable. As can be seen in Table 6, the omnibus test for each main effect model was statistically significant.

Analysis 1. This analysis found that the total PC-PTSD score was significantly related in the predicted direction to whether or not a clinical PTSD diagnosis was given within 90 days of the screening. With each one point increase in the PC-PTSD score, the odds of being diagnosed with PTSD increase by 52% ($OR = 1.52$, $z(6999) = 17.14$, $p < .001$).

Analysis 2. Using the recommended cutoff of three or more positive responses was also predictive of a PTSD diagnosis. Returning USAF Iraq and Afghanistan deployers meeting this cutoff had more than three times the odds of receiving a PTSD diagnosis ($OR = 3.30$, $z(6999) = 14.15$, $p < .001$) compared to subjects not meeting the cutoff.

Analysis 3. An alternative cutoff of two or more positive PC-PTSD responses was likewise effective in predicting PTSD. The odds of a PTSD diagnosis nearly tripled ($OR = 2.90$, $z(6999) = 16.46$, $p < .001$) for those meeting a cutoff of 2 when compared to those who did not.

Analysis 4. The predictor for this analysis was a PC-PTSD cutoff of one or more positive responses. This is the "real world" scoring method used by the USAF to trigger a referral for further assessment. As with the previous scoring methods, this minimal cutoff was significantly related to a subsequent PTSD diagnosis. Answering "Yes" to at least one item on the PC-PTSD

resulted in more than two-and-a-half times the odds of being clinically diagnosed with PTSD ($OR = 2.56$, $z(6999) = 16.16$, $p < .001$), as opposed to answering "No" on all four items.

Analysis 5. The last analysis using the PC-PTSD score to predict clinical PTSD involved the use of the single-item screen suggested by Bliese et al. (2008). Once again, a positive response to the PC-PTSD avoidance question (item 3) was an effective predictor of later PTSD status. The odds of a PTSD diagnosis were nearly three times higher for Iraq and Afghanistan veterans answering "Yes" on this item ($OR = 2.80$, $z(6999) = 14.04$, $p < .001$) than for veterans selecting "No."

As expected from previous validation studies, among the dichotomous IVs, the cutoff value of 3 was the most predictive of a subsequent PTSD diagnosis. Also, the single-item screen was nearly as predictive of PTSD as the cutoff of 2, and was better at predicting PTSD than the cutoff of 1.

Divergent validity. Analysis 6 was included to provide an indication of whether the PC-PTSD was more effective at predicting PTSD than MDD, which, as delineated previously, is highly comorbid with PTSD. For this analysis, the recommended PC-PTSD cutoff of 3 was modeled with MDD status as the DV. The results were that, at this cutoff level, the PC-PTSD was able to significantly predict a subsequent clinical MDD diagnosis. Satisfying the cutoff of 3 was associated with 1.7 times higher odds of an MDD diagnosis ($OR = 1.72$, $z(6999) = 5.14$, $p < .001$) compared to not meeting the cutoff.

Analysis 2 modeled the same PC-PTSD score configuration using the identical sample. The only difference between the main effect models in Analysis 2 and Analysis 6 was the DV:

Analysis 2 modeled PTSD and Analysis 6 modeled MDD. It appears that at the cutoff level of 3, the PC-PTSD was better at predicting PTSD (*OR* = 3.30) than MDD (*OR* = 1.72). However, it is unclear whether the difference between these *OR*s rises to the level of statistical significance. The CFA results later in this chapter will attempt to address the same divergent validity issue within a single model.

Diagnostic Utility

Differential prediction. The results of the diagnostic utility analyses are summarized in Table 7. Although the *DLR*+ was higher for males on each tested PC-PTSD scoring method, none of these differences were statistically significant. As Table 7 shows, none of the 99% confidence intervals overlapped between the male and female respondent groups. The pattern for whether men or women had the lowest *DLR*- was more variable than with the *DLR*+ gender differences. However, all differences were likewise statistically insignificant. This outcome was consistent with that of the logistic regression portion of the study in providing no evidence for differential prediction of PTSD by gender for the PC-PTSD.

Predictive value. Each tested PC-PTSD configuration was found to have some value in predicting later PTSD. This was demonstrated by the fact that none of the *CI*s for the *DLR*s were inclusive of the value of 1, and that all *DLR*+ values were greater than 1 and all *DLR*- values were less than 1. For example, a *DLR*+ equal to 1 indicates that those with and without the index disorder have an equal probability of testing positive on the instrument. If the *DLR*+ value is lower than 1, subjects with the disorder actually have a lower probability of testing positive on

Table 7

Comparison of PC-PTSD Diagnostic Utility by Gender

Cutoff	Sensitivity	Specificity	Efficiency	*PPV*	*NPV*	*DLR*+ (99% *CI*)	*DLR*- (99% *CI*)
≥ 1							
Male	.85	.89	.89	.10	1.00	7.56 (8.77-6.52)	.16 (.32-.08)
Female	.78	.84	.84	.10	.99	4.83 (6.56-3.56)	.26 (.65-.10)
≥ 2							
Male	.68	.94	.94	.16	.99	12.55 (15.86-9.94)	.33 (.50-.22)
Female	.75	.91	.91	.17	.99	8.47 (12.27-5.85)	.27 (.64-.12)
≥ 3							
Male	.47	.98	.97	.24	.99	19.76 (28.40-13.75)	.54 (.70-.42)
Female	.57	.96	.95	.24	.99	13.55 (23.56-7.80)	.45 (.78-.25)
4							
Male	.35	.99	.98	.43	.99	48.39 (83.77-27.95)	.66 (.80-.54)
Female	.28	.98	.96	.24	.98	13.55 (34.17-5.38)	.73 (.99-.54)
Avoidance item							
Male	.52	.96	.96	.18	.99	14.50 (19.94-10.54)	.50 (.66-.38)
Female	.54	.94	.93	.17	.99	8.82 (15.16-5.14)	.49 (.83-.29)

Note. N = 6999 (Males = 5785, Females = 1214). PC-PTSD = Primary Care Posttraumatic Stress Disorder screen; *PPV* = positive predictive value; *NPV* = negative predictive value; *DLR*+ = diagnostic likelihood ratio for a positive test; *DLR*- = diagnostic likelihood ratio for a negative test; *CI* = confidence interval.

the instrument in question than those without the disorder (Akobeng, 2006b). Since the PC-PTSD screen is designed to detect probable PTSD, either of these examples would be undesirable.

Cutoff ≥ 3 was the lowest cutoff level for which the DLR+ exceeded the threshold of 10 for both males and females (males: *DLR*+ = 19.76, 99% *CI* [28.40, 13.75]; females: *DLR*+ = 13.55, 99% *CI* [23.56, 7.80]), indicating that a respondent meeting this cutoff has a very strong probability of actually having the disorder. Specifically, men later diagnosed with PTSD were nearly 20 times more likely to answer three or more PC-PTSD items positively than men who did not receive the diagnosis. Women who, subsequent to the screening, received a PTSD diagnosis were over 13 times more likely to meet the three-item cutoff than women not so diagnosed. The diagnostic efficiency achieved at this cutoff level (.97 for men, .95 for women) was nearly as high as for the highest possible PC-PTSD scoring threshold of four positive items (.98 for men, .96 for women).

The USAF cutoff ≥ 1 had the lowest *DLR*+ values (males: *DLR*+ = 7.56, 99% *CI* [8.77, 6.52]; females: *DLR*+ = 4.83, 99% *CI* [6.56, 3.56]). This means that the probability of a male later identified with PTSD meeting the cutoff of 1 was about 7.5 times greater than for a male not identified with the disorder. Diagnosed females were about five times more likely of the same when compared to undiagnosed females. This cutoff exhibited the lowest specificity and diagnostic efficiency values for both genders (women = .84 and men = .89 for both statistics). However, this scoring configuration maximized sensitivity for both women (.78) and men (.85).

The single-item screen using only the avoidance question (males: *DLR*+ = 14.50, 99% *CI* [19.94, 10.54]; females: *DLR*+ = 8.82, 99% *CI* [15.16, 5.14]) had higher *DLR*+ values than either the cutoff of 1or the cutoff of 2 (males: *DLR*+ = 12.55, 99% *CI* [15.86, 9.94]; females: *DLR*+ = 8.47, 99% *CI* [12.27, 5.85]). While the *DLR*+ values for recommended cutoff (≥ 3) were higher than for the single-item screen, the sensitivity and specificity values were comparable. The specificity of the single-item screen (males = .96, females = .94) was only slightly lower than the specificity of the cutoff of 3 (males = .98, females = .96). Likewise, the disparities between the sensitivity values for the single-item screen (men = .52, women = .54) and those for the recommended cutoff (.47 and .57, respectively) were small. Efficiency for this scoring regime (.96 for males, .93 for females) was higher than for a cutoff of 2 (.94 for men, .91 for women), but a little lower than for the ≥ 3 cutoff (.97 and .95, respectively). This was similar to the result reported by Bliese and colleagues (2008).

Validity using CFA

The a priori model is illustrated in Figure 3. Two latent factors were included in the model and were allowed to correlate. Factor 1 corresponded to PTSD and had all four PC-PTSD items loading on it. Factor 2, depression, was measured by the 2-item PHQ. As mentioned in the chapter on methods, the variance of both factors was set to 1 and each path weight was freely estimated. The model fit results will be presented first, followed by the substantive results relating to validity.

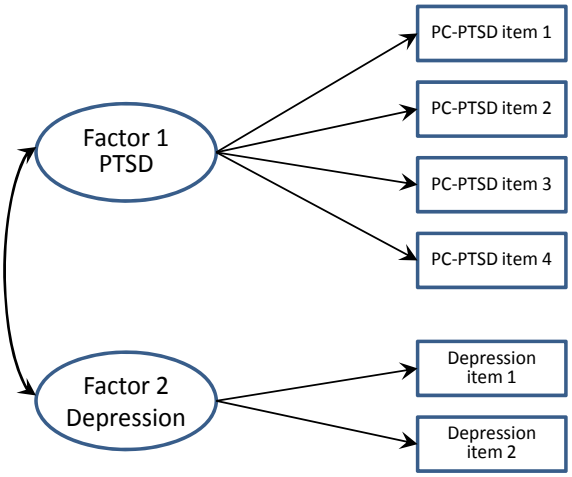

Figure 3. A priori CFA model for assessment of convergent and divergent validity in the PC-PTSD. The depression items are from the 2-item PHQ.

A Priori Model Fit

Nonzero paths. The first assessment of model fit involved testing whether each path weight was significantly different from zero. As shown in Table 8, for each path weight in the a priori model the ratio of the estimate divided by the estimate's standard error was much larger than 1.96—the threshold for significance with $\alpha = .05$. The lowest value was 57.47 for the arousal indicator of the PTSD factor. Also, none of the 95% *CI*s contained the value of zero.

Examination of model fit. The a priori model did not exhibit perfect fit with the data ($\chi^2 = 159.91$, 5 *d.f.*, $p < .001$). Close fit was indicated by the *CFI* (.989) exceeding .95. The value of the *RMSEA* (.067), on the other hand, was indicative of moderate fit.

The residual covariances were also inspected. All of the residuals were close to zero, indicating good fit. However, the three values showing the greatest deviation from zero were for PC-PTSD item 1 with PC-PTSD item 4 (-.095), PC-PTSD item 1 with the PHQ little interest item (-.080), and PC-PTSD item 4 with the PHQ little interest item (.085).

Three modification index (*MI*) values exceeded the chosen minimum of 10 and were therefore of concern. The largest (132.24) occurred for PC-PTSD item 4 (numbing) with Factor 2 (depression). The *MI*s for PC-PTSD item 1 (reexperiencing; *MI* = 30.99) and item 3 (arousal; *MI* = 30.44) also suggested that the model fit would improve by making these items indicators of Factor 2.

Specification of Model 2

Since the highest *MI* value suggested that allowing the PC-PTSD numbing indicator to load on the depression factor would result in closer fit for the model, item 4 was examined more

Table 8

A Priori Model: CFA of the PC-PTSD and 2-Item PHQ Screens

Scale	Estimates (*SE*)		Est./*SE*	95% *CI*
	PTSD	Depression		
PC-PTSD				
Reexperiencing	.88 (.01)	--	73.83	[.86, .91]
Avoidance	.92 (.01)	--	86.50	[.90, .94]
Arousal	.82 (.01)	--	57.47	[.79, .84]
Numbing	.94 (.01)	--	90.43	[.92, .96]
2-item PHQ				
Little interest	--	.93 (.01)	89.79	[.91, .95]
Feeling down	--	.93 (.01)	94.78	[.91, .95]
Correlated factors				
PTSD-Depression	.74 (.02)		50.44	[.71, .76]
χ^2	159.91 @ 5 *d.f.****			
CFI	.989			
RMSEA	.067			

Note. *N* = 6963 (Males = 5756, Females = 1207). PC-PTSD = Primary Care Posttraumatic Stress Disorder screen; PHQ = Patient Health Questionnaire; *SE* = standard error; Est. = estimate; *CI* = confidence interval; *d.f.* = degrees of freedom; *CFI* = comparative fit index; *RMSEA* = root mean square error of approximation.

****p* < .001.

closely in an attempt to provide some insight into this result. This item reads, "Felt numb or detached from others, activities, or your surroundings." One of the two-item PHQ screen questions reads, "Little interest or pleasure in doing things." In fact, the DSM-IV (APA, 1994) depression symptom from which this PHQ item was drawn elaborates to include little interest in "activities." The two queries appear to be covering—if not the same mood-related concept—at least similar aspects of a related concept. For instance, if a subject were experiencing PTSD symptoms and felt numb to people and/or activities to the degree that they endorsed item 4 on the PC-PTSD, it would logically follow that the same respondent might also be expected to endorse an item assessing whether they were having problems with lack of interest in "things," which might naturally include activities and social interaction. Thus, these items may have additionally been serving as indicators of anhedonia, a well-recognized symptom of depression.

To test this hypothesis, a second CFA model was specified in a manner that would allow a nested model comparison with the a priori model. For CFA model 2, item 4 was allowed to be an indicator variable for both PTSD (Factor 1) and depression (Factor 2). The structure of model 2 is illustrated in Figure 4.

Model 2 Fit

Nonzero paths. Table 9 summarizes the results for model 2. As with the a priori model, model 2 contained all nonzero paths. Evidence for this was found in the estimate to standard error ratios which were all larger than 1.96 (the smallest was 12.42) and in the 95% *CI*s—none of which were inclusive of zero.

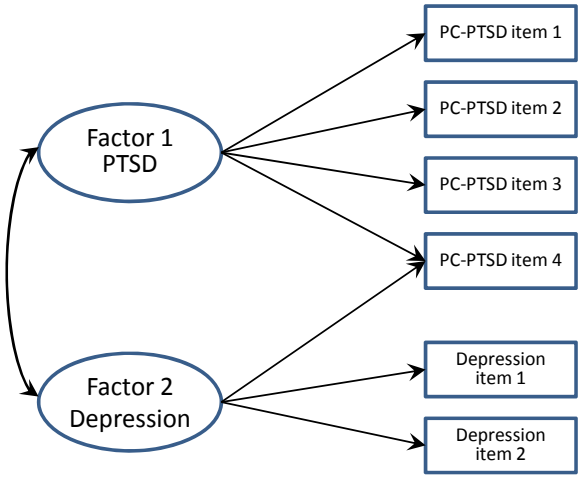

Figure 4. Model 2: CFA model for assessment of convergent and divergent validity in the PC-PTSD. PC-PTSD item 4 (numbing) is allowed to be an indicator for both the PTSD and depression factors. The depression items are from the 2-item PHQ.

Table 9

Model 2: CFA of PC-PTSD and 2-Item PHQ Screens

Scale	Estimates (*SE*)		Est./*SE*		95% *CI*	
	PTSD	Dep.	PTSD	Dep.	PTSD	Dep.
PC-PTSD						
Reexperiencing	.90 (.01)	--	75.04	--	[.87, .92]	--
Avoidance	.95 (.01)	--	88.60	--	[.93, .97]	--
Arousal	.84 (.02)	--	58.04	--	[.82, .87]	--
Numbing	.62 (.03)	.37 (.03)	21.73	12.42	[.56, .67]	[.31, .43]
2-item PHQ						
Little interest	--	.92 (.01)	--	90.95	--	[.90, .94]
Feeling down	--	.93 (.01)	--	95.99	--	[.91, .95]
Correlated factors						
PTSD-Depression	.66 (.02)		36.16		[.62, .70]	
χ^2	38.48 @ 5 *d.f.****					
CFI	.998					
RMSEA	.031					

Note. *N* = 6963 (Males = 5756, Females = 1207). Numbing is an indicator for both latent factors (PTSD, depression). PC-PTSD = Primary Care Postraumatic Stress Disorder screen; PHQ = Patient Health Questionnaire; Dep. = Depression; *SE* = standard error; Est. = estimate; *CI* = confidence interval; *d.f.* = degrees of freedom; *CFI* = comparative fit index; *RMSEA* = root mean square error of approximation.

****p* < .001.

Examination of model fit. Imperfect fit with the data was again indicated (χ^2 = 38.48, 5 *d.f.*, $p < .001$). However, the *CFI* (.998) exceeded the .95 threshold for close fit and the *RMSEA* (.031) suggested a good fit between the data and the specified model. The residuals for the covariances were generally lower than for the a priori model, with deviation the greatest for PC-PTSD item 3 with PC-PTSD item 2 (-.043) followed by PC-PTSD item 4 with PC-PTSD item 3 (.042) and the PHQ feeling down item with PC-PTSD item 2 (.039). One *MI*, for PC-PTSD item 2 (avoidance), exceeded the cutoff of 10 (10.70), suggesting that the model fit would improve if this item were made a manifest variable for the depression factor. Since this *MI* was only slightly above the threshold of 10 and there was no readily apparent theoretical justification for allowing depression to load on the PC-PTSD avoidance item, no further model modifications were specified.

Model Comparison

Due to estimator that Mplus uses with categorical variables in CFA, the model fit chi-square values reported previously for each model could not be used to directly calculate a chi-square difference. Instead, an extra analysis step was required which resulted in the output of a chi-square difference that is appropriate for the estimator. In this case, the resulting chi-square difference test (χ^2 = 91.28, 1 *d.f.*, $p < .001$) indicated that allowing both the PTSD and depression factors to load on the numbing item improved the model fit to a statistically significant degree when compared to the model that restricted numbing to being an exclusive indicator for PTSD.

Validity Results

Convergent validity. For the a priori model, the estimated path weights for each expected factor indicator were all very strong (≥ .60). In fact, the lowest estimate was for the arousal indicator (PC-PTSD item 3) of Factor 1 (PTSD) at .82. The others all approached or were above .90. Thus, it appears that the model demonstrated good convergent validity for the PC-PTSD.

Likewise, for model 2 most path weights were above .60 and, therefore, strong. While still judged as a "strong" path weight, numbing as an indicator of PTSD (.62) was the only PC-PTSD item whose path weight was not estimated at .80 or above. With the PTSD factor loading strongly on all four PC-PTSD items, model 2 provided further evidence of good convergent validity for scores on the PC-PTSD screen.

Divergent validity. In the a priori model, the correlation between the latent factors was somewhat high (.74). That the factors were highly correlated is, perhaps, not surprising given the high comorbidity of depression with PTSD. However, the correlation was below .80, indicating sufficient discriminant validity according to Brown (2006). This result appears to provide evidence of divergent validity for the PC-PTSD, at least at the scale level. However, due to the high *MI* for this model indicating better fit would be achieved by making numbing an indicator for the depression factor, attention was turned to model 2 and the items within the PC-PTSD scale.

As noted previously, allowing numbing to serve as an indicator for both latent factors in model 2 resulted in a statistically significant improvement in model fit. The inter-factor

correlation was further reduced to .66 in this model, again demonstrating scale-level divergent validity for the PC-PTSD scores. Upon closer inspection, the lowest path weight value in model 2 was the path linking depression with the numbing item (.37). This was considered weak (< .40). In contrast, not only were path weights linking PTSD to numbing (.62) and the other three PC-PTSD items strong (one greater than .80 and two at least .90), but the path weights associated with the two PHQ items as manifest variables for depression were also very high (both greater than .90). This marked disparity in path weights in model 2 pointed toward good divergent validity at the item-level for the PC-PTSD scores.

CHAPTER V: DISCUSSION

The purpose of this dissertation was to examine whether the scores from the PC-PTSD are equally valid for both male and female military veterans returning from combat. In this population, the results were generally consistent with the current practice of interpreting the PC-PTSD scores in the exact same manner for both genders. The large sample size and multiple methodologies used in the current study lend confidence that, if there were bias in the instrument, it very likely would have been detected. Several findings point to the lack of differential validity by gender for the PC-PTSD.

Summary of Findings on Research Questions

Research Question 1

This research question specifically addressed measurement non-equivalence in the PC-PTSD by gender. The IRT DIF results provided no evidence of differential functioning of the items on the PC-PTSD for males and females. While for each item there were differences in both the *a* and *b* parameters across the male and female respondent groups, none of these differences were statistically significant. The finding of no DIF was especially striking given the very large sample size which gave the analysis unusually high statistical power to detect DIF if it had existed.

This result means that males and females with similar levels of the PTSD trait (denoted by θ in the IRT model) will respond similarly to the PC-PTSD items. By implication, then, any gender differences in the likelihood to respond "Yes" to some or all of the PC-PTSD items were unlikely to be due to differential functioning within the items themselves. Rather, these results

support the idea that it is more probable that gender disparities in the manifestation of underlying PTSD are contributing to any observed response differences on this screening instrument. Some of the potential sources of such PTSD gender differences—trauma exposure, neurophysiology, and reactions to trauma—were discussed earlier in the literature review chapter.

Research Question 2

The second research question sought to determine whether the PC-PTSD showed differential prediction of PTSD among males and females using various scoring configurations. The cutoff levels of 1, 2, and 3 were assessed using both logistic regression interaction models and comparison of male and female diagnostic likelihood ratios (*DLR*s). Under both analysis regimes, there was no evidence to support differential prediction by gender. Likewise, no significant differences in PTSD prediction were found between males and females when the PC-PTSD total score was evaluated using logistic regression or when a cutoff of 4 was analyzed using DLRs. This confirmation of the PC-PTSD's validity was especially important because the scoring regime used by the USAF, cutoff ≥ 1, does not appear to be biased toward identifying higher rates of PTSD risk for either males or females.

Also, the results suggested that any gender differences in the rate at which the PC-PTSD identifies risk for PTSD are not related to any differential prediction by PC-PTSD scores. For example, although the PC-PTSD did identify a higher percentage of women than men as being at risk for PTSD, the magnitude of the difference was similar to the observed differential in PTSD prevalence by gender as judged by the clinical diagnosis criterion. Within the logistic regression/diagnostic utility sample ($N = 6999$), at the recommended cut score of 3 the PC-PTSD

identified 5.44% of women and 3.08% of men as being at risk of PTSD. The ratio of these rates shows that females were 1.77 times as likely to be screened positive for risk of PTSD. The sample prevalence of PTSD, as indicated by the clinical diagnosis criterion, was 2.31% for females versus 1.54% for males. PTSD was, therefore, 1.5 times more prevalent in this sample among females. Both gender differentials are comparable to those reported in the previously discussed PTSD prevalence literature and are reasonably similar to one another. Again, it is worth noting that, even with high levels of statistical power supplied by a relatively large sample size and using two distinct methods of analysis, no statistically significant differential prediction by gender was detected.

One *DLR*+ gender difference was, though not statistically significant, strikingly large. For the cutoff level of 4, the male *DLR*+ (48.39, 99% *CI* [83.77, 27.95]) was more than three times greater than the female *DLR*+ (13.55, 99% *CI* [34.17, 5.38]). There was also a great deal of variation between the *PPV* for men and women (.43 and .24, respectively). It appears, though, that relatively large 99% *CI*s prevented the *DLR*+ difference from being statistically significant. The wide *CI*s likely resulted from the extremely low number of respondents that met the cutoff of four (72/5785 men, 33/1214 women) and that were clinically diagnosed with PTSD (89/5785 men, 28/1214 women). The resulting cross tabulation cell for true positives had only 31 cases for males (43.1% of those positive on the cutoff of 4) and 8 cases for females (24.2% of those positive on the cutoff of 4). Since the small figures for both true and false positives are used in the denominator of the *DLR*+ *CI*, the calculated intervals were relatively large.

Research Question 3

This research question pertains to the predictive validity of the proposed one-item screen for PTSD using only the avoidance item (item 3) from the PC-PTSD. As with the various scoring configurations of the PC-PTSD in the discussion of the previous research question, neither the logistic regression interaction analysis nor the *DLR* analysis yielded evidence that the single-item screen differentially predicted PTSD between men and women. The implications were likewise similar for the avoidance screen: Differences in the rate of males and females identified as at risk for PTSD mirrored the observed differences between the genders in the sample prevalence of PTSD.

The study also confirmed some of the findings from the initial study proposing the single-item screen. Like Bliese et al. (2008), the diagnostic efficiencies of the avoidance-only screen (.96 for males, .93 for females) fell roughly between those of the cutoff of 2 (.94 for males, .91 for females) and the cutoff of 3 (.97 and .95, respectively). Also, while only the one-item screen male *DLR*+ rose above the threshold of 10, indicating a very high likelihood of diagnosis for those screening positive ($DLR+_{male} = 14.50$, $DLR+_{female} = 8.82$), both *DLR*+ gender values were higher than for both the ≥ 2 cutoff ($DLR+_{male} = 12.55$, $DLR+_{female} = 8.47$) and for the ≥ 1 cutoff used by the USAF ($DLR+_{male} = 7.56$, $DLR+_{female} = 4.83$). The specificity of this reduced screen (male = .96, female =.94) was almost as high as for the previously recommended ≥ 3 cutoff (male = .98, female =.96). Finally, the sensitivities were similar between the avoidance-only screen (.52 for men, .54 for women) and the ≥ 3 cutoff (.47 for men, .57 for women), but the

single-item screen appeared to have the advantage of less variability in the sensitivities for the male versus the female groups.

These results compared favorably to those of Bliese and colleagues (2008) on the avoidance-only PC-PTSD screen, but with a much larger sample size of females (1214 in the current study compared to 13 in the former). This analysis, therefore, extends support for the viability of this reduced screen to female as well as male veterans.

Research Question 4

As Prins and colleagues (2003) noted, in their study the recommended cutoff of 3 positive PC-PTSD items resulted in markedly lower sensitivity and specificity values for females when compared to males. The fourth research question was posed to determine whether similar differences in diagnostic utility would hold within a military population returning from deployment.

Generally, sensitivity exhibited greater variability than specificity across gender in this sample (see Table 7). Any variability in specificity may have been muted by the overwhelmingly large proportion of both negative PC-PTSD responses and negative PTSD clinical diagnosis status for both genders. Since specificity is given by the true negatives divided by all negative diagnosis results and the true negatives are those cases where both the test and diagnosis are negative, it is logical that the ratio of true negatives to total diagnostic negatives would be nearly one for both males and females in this study.

The scoring configuration with the largest sensitivity gender disparity occurred at the cutoff level of 3 (.57 for women, .47 for men). This finding was the opposite of that in Prins et

al. (2003) which reported lower sensitivity for women. The figures from this study are in a similar range to the sensitivity of cutoff ≥ 3 found with Army members returning from Iraq (.46 for males and females, combined) in Bliese et al. (2004).

The sensitivity of the PC-PTSD was also found to be higher for females in this sample using cutoff ≥ 2 (.75 for females, .68 for males) and avoidance-only (.54 for females, .52 for males). This pattern was reversed, however, for the cutoff ≥ 1 (.78 for females, .85 for males) and the cutoff of 4 (.28 for females, .35 for males). It is difficult to draw conclusions from the highly variable direction of differences in sensitivity between males and females. Certainly, one possible explanation, as was proposed in Prins et al. (2003) is that gender differences in comorbid conditions may be a factor.

Also, while sensitivity and specificity are theoretically independent of condition prevalence, these measures are influenced by factors such as the intensity of the index condition. Greater intensity in the tested condition generally leads to better diagnostic accuracy for the tested instrument (Zhou et al., 2002). So, gender differences in the intensity of PTSD or of particular symptoms could be contributing to this variability. Since none of the datasets used for this study contained information on the intensity of PTSD, it was impossible to control for this variable. It may be, then, that unmeasured "spectrum of disease" (Zhou et al., 2002, p. 21) gender differences, such as PTSD intensity, account for some of the variation between male and female sensitivity and specificity values.

Research Question 5

The final research question sought to examine evidence for convergent and divergent validity of the PC-PTSD scores in a sizable population of military males and females returning from combat deployments. Given that large-scale epidemiological research has shown that as many as 48% of those experiencing PTSD will also be diagnosed with major depression during their lifetime (Kessler et al., 1995), it was encouraging that the CFA analysis found evidence that the PC-PTSD items showed adequate convergent and divergent validity when modeled with a previously validated depression screen.

In both the a priori model and model 2, convergent validity was indicated by the strong path weights connecting the PTSD factor and the PC-PTSD items. Similarly, for each model the correlation between the latent factors (PTSD and depression) was sufficiently low to provide evidence of scale-level divergent validity for PC-PTSD scores. Item-level divergent validity was demonstrated in both models, as well: When the PC-PTSD numbing item was freed to indicate either PTSD or depression, the path weight associated with depression was weak while the PTSD path weight was strong (as were all other paths linking PTSD and PC-PTSD items). On balance, the findings from each of the planned analyses bolstered the construct validity of the PC-PTSD scores with a sample of returning military combat deployers.

Other Findings

Cutoff ≥ 1 appears to have provided the best balance between high sensitivity and high specificity in this sample of the PDHRA target population (*sens*. = .85 for males, .78 for females; *spec*. = .89 for males and .84 for females). As stated in a policy memorandum from the Office of

the USAF Surgeon General (2005, December 8), the purpose of the USAF PDHRA program is to identify potential deployment-related health problems "proactively" (Office of the USAF Surgeon General, 2005, December 8, p. 1) so as to facilitate early treatment. The emphasis is on identifying as many people as possible that are potentially suffering from PTSD symptoms in order to get them appropriate assistance sooner than later.

Therefore, it is important to maximize sensitivity in the PC-PTSD screen, because, as mentioned previously, higher sensitivity results in a lower rate of false negatives. In the present case, false negatives are those respondents who actually have a problem with PTSD but were not flagged by the screening method as such. Failure to refer an individual who, in reality, is at-risk for PTSD limits that individual's opportunity for early, appropriate care. A drawback, though, of choosing the cutoff of one is that it results in lower specificity than for the other PC-PTSD scoring configurations. However, it does not seem that a false positive rate (given by 1 – specificity) of 11-16% (for men and women, respectively) would overburden the military medical or mental health systems.

Limitations of the Study

Despite the various findings from this study that lend support to the validity of the PC-PTSD across gender groups, the conclusions drawn must be tempered by a number of limitations. First of all, while the main sample constituted the entire population of male and female USAF members who had returned from deployment to Iraq and Afghanistan and who had completed all PC-PTSD items during 2008 and 2009 within 90 to 180 days of their return, this sample did not constitute a random sample of any larger population (such as all returning

military deployers or military veterans-at-large) and so may not be generalizable beyond the study population. Also, as mentioned in the results section, excluding cases without a military medical visit 90 days after the screening resulted in the elimination of nearly all subjects who were either members of the national guard or reserve in both the logistic regression/diagnostic utility sample and the CFA sample. Consequently, the results of the predictive validity and CFA portions of the study may only be generalizable to active duty USAF military deployers and not to national guard members or reservists.

Another sample-related issue is that, with the available data, it is impossible to know why non-responders failed to respond. Since participation in the PDHRA is not required, some number of returning USAF deployers undoubtedly elected not to even log into the online screening instrument. However, no information is available regarding those who were offered the PDHRA but declined to participate. Others who did register some level of participation in the larger PDHRA did not respond to one or more PC-PTSD items. Such subjects were excluded from the analysis. In the case of either type of non-response, if there was a systematic reason for not responding the study sample would be less representative of the target population and the study results would be less generalizable due to the resulting bias. Such bias cannot be ruled out due to the lack of data regarding the reasons for any lack of response. However, as noted in the method section, the percentage of cases deleted because of missing data was very small and it is unlikely that any possible systematic differences between responders and non-responders would have much of an effect on the results given the substantial sample size. Also, there was no

statistically significant relationship between gender and case deletion due to missing response data in any of the three study samples.

The primary analysis in this study was to assess measurement equivalence across gender for the PC-PTSD items. However, use of the 2PL IRT model with these data resulted in lack of item-fit for two of the four items. M. O. Edelen stated in personal communication (April 26, 2010) that such lack of fit is not uncommon when an instrument yields very few positive responses upon use with the tested population. Indeed, very sparse positive responses were a characteristic of the data from this study sample. Dr. Edelen suggested that it might be helpful to identify a subsample of the original sample with a much higher proportion of positive responses on the PC-PTSD items by identifying a PTSD risk factor, such as the presence of depression.

However, it was unclear how to proceed with this recommendation within the current study. First of all, it would appear that higher rates of exposure to combat, which is assumed to be the case with this sample of returning Iraq and Afghanistan veterans (Smith et al., 2008), would already constitute a major risk factor for PTSD. Second, the number of cases identified clinically as having a depression diagnosis was so small that using this criterion to determine inclusion in a higher-PTSD risk subsample would result in at least one of the comparison group sample sizes falling markedly below the suggested minimum of 500 for meaningful IRT parameterization (Reise & Yu, 1990).

Another potential limitation related to the IRT DIF methods employed in this study is that the anchor items are assumed to be free of any DIF. While the preliminary anchor identification analyses using regression and graphical means found that items 3 and 4 appeared to exhibit

relatively less potential DIF, the potential DIF was not zero. Items 3 and 4 were chosen as anchors because they exhibited the lowest standardized residual values for the *b* and *a* parameters, respectively. However, the residuals for parameter *a* on the third item and for parameter *b* on the fourth item were relatively high. In fact, the item 3 residuals on the *a* parameter were the second highest out of all four items—higher than the *a* residuals on item 2, which was one of the suspected DIF items. Similarly, the *b* residuals associated with item 4 were nearly as high as those for item 1—the other suspected DIF item. This means that the model comparison method likely was only able to detect any DIF above and beyond that already present in the anchor criterion. The net result may have been attenuated ability of the likelihood ratio test to detect DIF in these items with this sample.

The need to identify DIF-free anchors would appear to constitute a weakness of using IRT DIF methods with an instrument constructed using so few items. It stands to reason that if there had been many more items, the likelihood of finding at least one anchor item with negligible, or at least very low, residuals on both parameters would have been greater. Also, the chosen anchoring method resulted in only two of the items (1 and 2) being analyzed for DIF using IRT modeling.

As was mentioned in the chapter on methods, it was less than ideal to use a clinical diagnosis as the criterion in the predictive validity analyses. Studies that had compared the performance of a clinical PTSD diagnosis to a structured interview (MacGruder et al., 2005; Prins et al., 2003) have revealed low sensitivity for the clinical assessment. If one takes the highest value for sensitivity between the studies, .61, that works out to a 39% false negative rate.

The sample for the present study was skewed strongly toward negative diagnosis status. A diagnostic criterion with better diagnostic utility may have reduced this skew somewhat and would certainly have provided a more accurate assessment of the predictive validity of the PC-PTSD scores.

Another limitation involved a form of the history effect. In this sense, history is defined as an event or circumstance, other than the IV(s), that is contemporary with the study and affects the DV (Schutt, 2006). Due to the time lapse (varying from 0 to 90 days for each respondent) from the screening to the clinical encounter, it is impossible to know whether the screen and the subsequent clinical evaluation were measuring PTSD symptoms related to the same potentially traumatizing event(s). Specifically, some other trauma-inducing event (sexual assault, MVA, etc.) could have occurred to a respondent after returning from deployment and participating in the PDHRA but prior to being medically evaluated. Likewise, these data provided no way to know whether such events occurred differently for men and women in some systematic way.

For example, men have been shown to more frequently experience potentially traumatizing events when compared to women (Kessler et al., 1995; Norris et al., 2003). If in the present study men more often experienced a trauma during the interim between screening and diagnosis, this might tend to increase any related PTSD symptoms and increase the likelihood that those males would receive a PTSD diagnosis. The current study was not able to account for this history effect and it is unclear what, if any, influence may have resulted on the observed predictive validity results for the PC-PTSD across males and females.

Finally, this study lacked any measures of PTSD intensity. The use of mostly dichotomous measures largely precluded this level of detail in condition analysis. Consequently, it is difficult to rule out this important disease spectrum factor as possibly contributing to gender differences in diagnostic accuracy values.

Conclusions and Recommendations

Despite the noted limitations, this dissertation makes significant contributions to knowledge about PTSD screening in a military population. First, the PC-PTSD exhibited no significant gender DIF when assessed using IRT methods. Second, the predictive validity of this instrument was shown to be consistent across both males and females. Third, previous results regarding the validity of a potential single-item PTSD screen were confirmed. Fourth, the PC-PTSD demonstrated good evidence supporting convergent and divergent validity when modeled with a highly comorbid condition. Finally, the USAF scoring regime for the PC-PTSD was also confirmed as valid. Importantly, each of these findings occurred within the context of good statistical power due to a large sample of returning military deployers that also contained a substantial number of women. This study extends the body of research confirming that the PC-PTSD shows robust characteristics that recommend it for continued use as a primary method for early identification of PTSD among military veterans—both male and female.

Future work remains to be done to ensure the measurement equivalence of the PC-PTSD across gender. Given the problems that arose in this study related to anchor items and model fit, CFA group difference methods could, perhaps, provide an alternative to IRT DIF methods that would not require anchoring (Byrne, 2001). Additionally, CFA model fit can be evaluated using

a wider range of indices than are available for IRT modeling (Reise et al., 1993). Future analyses should also focus on measurement equivalence across groups subdivided along characteristics other than gender, such as ethnicity and age.

The current study focused on the PC-PTSD, which has four items which approximate each of the PTSD symptom clusters as they are conceptualized in the four-cluster model (Asmundson et al., 2004). More insight could be gained into any gender bias within specific PTSD symptoms by conducting measurement equivalence studies across gender for instruments that contain a specific query for each of the seventeen PTSD symptoms. Such instruments include the PCL (Blanchard et al., 1996; Weathers et al., 1993) and the Modified PTSD Symptom Scale—Self-Report (MPSS; Falsetti, Resnick, Resick, & Kilpatrick, 1993).

While this study provided initial confidence for consistent prediction of PTSD by the scores on the PC-PTSD when used with male and female veterans, the research would have benefited from using a PTSD criterion that was more reliable than a clinical diagnosis. Future validity work related to PC-PTSD should, whenever feasible, use a structured clinical interview as the gold standard indicating presence or absence of PTSD. Both the CAPS (Blake et al., 1995) and the Composite International Diagnostic Interview (CIDI; World Health Organization, 1997) have demonstrated high reliability and, specifically, good diagnostic utility with PTSD. Although the CAPS was designed for use by clinical professionals, both the CAPS and the CIDI can be administered by paraprofessionals with adequate training (Cusack et al., 2002). Thus, it could be practical to use these structured interviews to assess relatively large numbers of men

and women to test for differential prediction across gender with greater reliability than was possible in the current design.

It will also be important to attempt to control for the intensity of the PTSD condition being tested. Chung and Breslau (2008) provided an example of how PTSD disturbance level might serve as a proxy for condition severity. By controlling for severity researchers may gain insight into any remaining variability in the diagnostic accuracy of instruments such as the PC-PTSD between males and females.

REFERENCES

Abercrombie, H. C., Kalin, N. H., Thurow, M. E., Rosenkranz, M. A., & Davidson, R. J. (2003). Cortisol variation in humans affects memory for emotionally laden and neutral information. *Behavioral Neuroscience, 117,* 505-516.

Akobeng, A. K. (2006a). Understanding diagnostic tests 1: Sensitivity, specificity, and predictive values. *Acta Paediatrica, 96,* 338-341.

Akobeng, A. K. (2006b). Understanding diagnostic tests 2: Likelihood ratios, pre- and post-test probabilities, and their use in clinical practice. *Acta Paediatrica, 96,* 487-491.

American Psychiatric Association (1952). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

American Psychiatric Association (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Amir, M. & Sol, O. (1999). Psychological impact and prevalence of traumatic events in a student sample in Israel: The effect of multiple traumatic events and physical injury. *Journal of Traumatic Stress, 12,* 139-154.

Andersen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for

Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine, 10,* 77-84.

Andreano, J. M., Arjomandi, H., & Cahill, L. (2008). Menstrual cycle modulation of the relationship between cortisol and long-term memory. *Psychoneuroendocrinology, 33*, 874-882.

Andreano, J. M. & Cahill, L. (2006). Glucocorticoid release and memory consolidation in men and women. *Psychological Science, 17*, 466-470.

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 96-116). Baltimore, MD: Johns Hopkins University Press.

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3-24). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Asmundson, G. J. G., Stapleton, J. A., & Taylor, S. (2004). Are avoidance and numbing distinct PTSD symptom clusters? *Journal of Traumatic Stress, 17,* 467-475.

Arroll, B., Goodyear-Smith, F., Kerse, N., Fishman, T., & Gunn, J. (2005). Effects of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *British Medical Journal, 331,* 884-887.

Baker, B. A., Caison, A. L., & Meade, A. W. (2007). Assessing gender-related differential item functioning and predictive validity with the Institutional Integration Scale. *Educational and Psychological Measurement, 67(3)*, 545-559.

Baker, D. G., West, S. A., Nicholson, W. E., Ekhator, N. N., Kasckow, J. W., Hill, K. K., Bruce, A. B., et al. (1999). Serial CSF corticotrophin-releasing hormone levels and adrenocortical activity in combat veterans with posttraumatic stress disorder. *American Journal of Psychiatry, 156*, 585-588.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40,* 569-573.

Beck, A. T. & Beck, R. W. (1972). Screening depressed patients in family practice: A rapid technic. *Postgraduate Medicine, 52,* 81-85.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1962). Reliability of psychiatric diagnoses: 2. A study of consistency of clinical judgments and ratings. *American Journal of Psychiatry, 119,* 351-357.

Bender, L. & Blau, A. (1937). The reaction of children to sexual relations with adults. *American Journal of Orthopsychiatry, 7,* 500-518.

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J. et al. (2003). Development and validation of a brief screening version of the Child Trauma Questionnaire. *Child Abuse and Neglect, 27,* 169-190.

Biello, S. M., Golombek, D. A., & Harrington, M. E. (1997). Neuropeptide Y and glutamate block each other's phase shifts in the suprachiasmatic nucleus in vitro. *Neuroscience, 77,* 1049-1057.

Bjorner, J. B., Smith, K. J., Edelen, M. O., Stone, C., Thissen, D., & Sun, X. (2007). *IRTFIT: A macro for item fit and local dependence tests under IRT models*. Lincoln, RI: QualityMetric Incorporated.

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

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD checklist (PCL). *Behavior Research and Therapy, 34,* 669-673.

Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., and Hoge, C. W. (2008). Validating the Primary Care Posttraumatic Stress Disorder screen and the Posttraumatic Stress Disorder Checklist with soldiers returning from combat. *Journal of Consulting and Clinical Psychology, 76*, 272-281.

Bliese, P. D., Wright, K. M., Adler, A. B., Thomas, J. L., & Hoge, C. W. (2007). Timing of postcombat mental health assessments. *Journal of Consulting and Clinical Psychology, 76*, 272-281.

Bliese, P. D., Wright, K. M., Adler, A. B., & Thomas, J. (2004). *Validation of the 90 to 120 day post-deployment psychological short screen* (Report No. 2004-002). Heidelberg, Germany: U.S. Army Medical Research Unit—Europe.

Bolton, E. E., Litz, B. T., Britt, T. W., Adler, A., & Roemer, L. (2001). Reports of prior exposure to potentially traumatic events and PTSD in troops poised for deployment. *Journal of Traumatic Stress, 14,* 249-256.

Boscarino, J. A. (1996). Posttraumatic stress disorder, exposure to combat, and lower plasma cortisol among Vietnam veterans: Findings and clinical implications. *Journal of Consulting and Clinical Psychology, 64*, 191-201.

Bremner, J. D., Innis, R. B., Ng, C. K., Staib, L., Salomon, R., Bronen, R. A., et al. (1997a). Positron emission tomography measurement of cerebral metabolic correlates of yohimbine administration in combat-related posttraumatic stress disorder. *Archives of General Psychiatry, 54,* 246-256.

Bremner, J. D., Randall, P., Vermetten, E., Staib, L., Bronen, R. A., Mazure, C., Capelli, S., et al. (1997b). Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biological Psychiatry, 41,* 23-32.

Bremner, J. D., Vermetten, E., & Kelley, M. E. (2007). Cortisol, dehydroepiandrosterone, and estradiol measured over 24 hours in women with childhood sexual abuse-related posttraumatic stress disorder. *Journal of Nervous and Mental Disease, 195,* 919-927.

Breslau, N., Chilcoat, H. D., Kessler, R. C., Peterson, E. L., & Lucia, V. C. (1999). Vulnerability to assaultive violence: further specification of the sex difference in post-traumatic stress disorder. *Psychological Medicine, 29*, 813-821.

Breslau, N., Davis, G. C., Andreski, P., Peterson, E., & Schultz, L. R. (1997). Sex differences in posttraumatic stress disorder. *Archives of General Psychiatry, 54*, 1044-1048.

Breslau, N. & Kessler, R. C. (2001). The stressor criterion in DSM-IV posttraumatic stress disorder: An empirical investigation. *Biological Psychiatry, 50*, 699-704.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry, 55*, 626-632.

Breslau, N., Reboussin, B. A., Anthony, J. C., & Storr, C. L. (2005). The structure of posttraumatic stress disorder: Latent class analysis in 2 community samples. *Archives of General Psychiatry, 62*, 1343-1351.

Brewin, C. R., Andrews, B., & Rose, S. (2000a). Fear, helplessness, and horror in posttraumatic stress disorder. *Journal of Traumatic Stress, 13*, 499-509.

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000b). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology, 68*, 748-766.

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

Bryant, R. A. & Harvey, A. G. (2003). Gender differences in the relationship between acute stress disorder and posttraumatic stress disorder following motor vehicle accidents. *Australian and New Zealand Journal of Psychiatry, 37*, 226-229.

Burgess, A. W. & Holmstrom, L. L. (1974). Rape trauma syndrome. *American Journal of Psychiatry, 131*, 981-986.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 221-256). Westport, CT: Praeger Publishers.

Camilli, G. & Shepherd, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications, Inc.

Ceresini, G., Freddi, M., Morganti, S., Rebecchi, I., Modena, A. B., Rinaldi, M., Manca, C., et al. (2000). The effects of transdermal estradiol on the response to mental stress in postmenopausal women: A randomized trial. *American Journal of Medicine, 109*, 463-468.

Charney, D. S. (2004). Psychobiological mechanisms of resilience and vulnerability: Implications for successful adaptation to extreme stress. *American Journal of Psychiatry, 161*, 195-216.

Chung, H. & Breslau, N. (2008). The latent structure of post-traumatic stress disorder: Tests of invariance by gender and trauma type. *Psychological Medicine, 38*, 563-573.

Costello, E. J., Erkanli, A., Fairbank, J. A., & Angold, A. (2002). The prevalence of potentially traumatic events in childhood and adolescence. *Journal of Traumatic Stress, 15,* 99-112.

Costello, E. J., Angold, A., Burns, B. J., Stangl, D. K., Tweed, D. L., Erkanli, A., et al. (1996). The Great Smoky Mountains Study of Youth: Goals, design, methods, and the prevalence of DSM-III-R disorders. *Archives of General Psychiatry, 53*, 1129-1136.

Cucinelli, F., Soranna, L., Barini, A., Perri, C., Leoni, F., Mancuso, S., & Lanzone, A. (2002). Estrogen treatment and body fat distribution are involved in corticotrophin and cortisol response to corticotrophin-releasing hormone in postmenopausal women. *Metabolism, 51*, 137-143.

Cusack, K., Falsetti, S., & de Arellano, M. (2002). Gender considerations in the psychometric assessment of PTSD. In R. Kimerling, P. Ouimette, & J. Wolfe (Eds.), *Gender and PTSD* (pp. 150-176). New York: Guilford Press.

De Bellis, M. D., Baum, A. S., Birmaher, B., & Ryan, N. D. (1997). Urinary catecholamine excretion in childhood overanxious and posttraumatic stress disorders. *Annals of the New York Academy of Sciences, 821*, 451-455.

De Leo, V., la Marca, A., Talluri, B., D'Antona, D., & Morgante, G. (1998). Hypothalamo-pituitary-adrenal axis and adrenal function before and after ovariectomy in premenopausal women. *European Journal of Endocrinology, 138*, 430-435.

Deane, R., Chummun, H., & Prashad, D. (2002). Differences in urinary stress hormones in male and female nurses at different ages. *Journal of Advanced Nursing, 37*, 304-310.

Defense Manpower Data Center (2009a). *Operation Enduring Freedom military deaths: October 7, 2001 through June 6, 2009*. Retrieved June 17, 2009, from http://siadapp.dmdc.osd.mil/personnel/CASUALTY/oefdeaths.pdf

Defense Manpower Data Center (2009b). *Operation Iraqi Freedom military deaths: March 19, 2003 through June 6, 2009*. Retrieved June 17, 2009, from http://siadapp.dmdc.osd.mil/personnel/CASUALTY/oif-deaths-total.pdf

Drasgow, F. & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*, 662-680.

du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.

Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach. *Medical Care 44(11) (Supplement 3)*, S134-S142.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Falsetti, S., Resnick, H., Resick, P., & Kilpatrick, D. (1993). The modified PTSD symptom scale: A brief self-report measure of posttraumatic stress disorder. *Behavior Therapist, 16*, 161-162.

Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement, 29(4)*, 278-295.

Flannery, W. P., Reise, S. P., & Widaman, K. F. (1995). An item response theory analysis of the General and Academic scales of the Self-Description Questionnaire II. *Journal of Research in Personality, 29*, 168-188.

Flett, R. A., Kazantzis, N., Long, N. R., McDonald, C., & Millar, M. (2004). Gender and ethnicity differences in the prevalence of traumatic events: Evidence from a New Zealand community sample. *Stress and Health, 20*, 149-157.

Flora, C. M. (2002). A short history of PTSD from the military perspective. In M. B. Williams & J. F. Sommer, Jr. (Eds.), *Simple and complex post-traumatic stress disorder: Strategies for comprehensive treatment in clinical practice* (pp. 3-8). Binghamton, NY: Haworth Maltreatment and Trauma Press.

Frankenhaeuser, M., von Wright, M. R., Collins, A., von Wright, J., Sedvall, G., & Swahn, C. G. (1978). Sex differences in psychoneuroendocrine reactions to examination stress. *Psychosomatic Medicine, 40*, 334-343.

Fullerton, C. S., Ursano, R. J., Epstein, R. S., Crowley, B., Vance, K., Kao, T., Dougall, A., et al. (2001). Gender differences in posttraumatic stress disorder after motor vehicle accidents. *American Journal of Psychiatry, 158*, 1486-1491.

Furukawa, T. & Goldberg, D. P. (1999). Cultural invariance of likelihood ratios for the General Health Questionnaire. *Lancet, 353*, 561-562.

Furukawa, T. A., Andrews, G., & Goldberg, D. P. (2002). Stratum-specific likelihood ratios of the general health questionnaire in the community: Help-seeking and physical co-morbidity affect the test characteristics. *Psychological Medicine, 32*, 743-748.

Furukawa, T. A., Goldberg, D. P., Rabe-Hesketh, S., & Ustun, T. B. (2001). Stratum-specific likelihood ratios of two versions of the general health questionnaire. *Psychological Medicine, 31*, 519-529.

Garakani, A., Mathew, S. J., & Charney, D. S. (2006). Neurobiology of anxiety disorders and implications for treatment. *Mount Sinai Journal of Medicine, 73*, 941-949.

Gill, J., Vythilingam, M., & Page, G. G. (2008). Low cortisol, high DHEA, and high levels of stimulated TNF-α, and IL-6 in women with PTSD. *Journal of Traumatic Stress, 21*, 530-539.

Gilmer, W. S. & McKinney, W. T. (2003). Early experience and depressive disorders: Human and non-human primate studies. *Journal of Affective Disorders, 75*, 97-113.

Girdler, S. S., Pedersen, C. A., Stern, R. A., & Light, K. C. (1993). Menstrual cycle and premenstrual syndrome: Modifiers of cardiovascular reactivity in women. *Health Psychology, 12*, 180-192.

Gnanadesikan, M., Novins, D. K., & Beal, J. (2005). The relationship of gender and trauma characteristics to posttraumatic stress disorder in a community sample of traumatized Northern Plains American Indian adolescents and young adults. *Journal of Clinical Psychiatry, 66*, 1176-1183.

Goldberg, L. R. & Freyd, J. J (2006). Self-reports of potentially traumatic experiences in an adult community sample: Gender differences and test-retest stabilities of the items in a brief betrayal-trauma survey. *Journal of Trauma & Dissociation, 7*, 39-63.

Goldstein, D. S., Levinson, P., & Keiser, H. R. (1983). Plasma and urinary catecholamines during the human ovulatory cycle. *American Journal of Obstetrics and Gynecology, 146*, 824-829.

Goldstein, J. M., Jerram, M., Poldrack, R., Ahern, T., Kennedy, D. N., Seidman, L. J., & Makris, N. (2005). Hormonal cycle modulates arousal circuitry in women using functional magnetic resonance imaging. *Journal of Neuroscience, 25*, 9309-9316.

Greber, S., Schwarzer, C., & Sperk, G. (1994). Neuropeptide Y inhibits potassium-stimulated glutamate release through Y_2 receptors in rat hippocampal slices *in vitro*. *British Journal of Pharmacology, 113*, 737-740.

Griffin, M. G., Resick, P. A., & Yehuda, R. (2005). Enhanced cortisol suppression following dexamethasone administration in domestic violence survivors. *American Journal of Psychiatry, 162*, 1192-1199.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Hardin, J. W. & Hilbe, J. M. (2007). *Generalized linear models and extensions* (2nd ed.). College Station, TX: Stata Press.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

Hays, R. D., Liu, H., Spritzer, K., & Cella, D. (2007). Item response theory analyses of physical functioning items in the Medical Outcomes Study. *Medical Care, 45(5) (Supplement 1)*, S32-S38.

Heim, C., Newport, D. J., Bonsall, R., Miller, A. H., & Nemeroff, C. B. (2001). Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *American Journal of Psychiatry, 158*, 575-581.

Heim, C., Plotsky, P. M., & Nemeroff, C. B. (2004). Importance of studying the contributions of early adverse experience to neurobiological findings in depression. *Neuropsychopharmacology, 29*, 641-648.

Helzer, J. E., Clayton, P. J., Pambakian, R., Reich, T., Woodruff, R. A., & Reveley, M. A. (1977). Reliability of psychiatric diagnosis: II. The test/retest reliability of diagnostic classification. *Archives of General Psychiatry, 34,* 136-141.

Helzer, J. E., Robins, L. N., & McEvoy, L. (1987). Post-traumatic stress disorder in the general population: Findings of the Epidemiological Catchment Area Survey. *The New England Journal of Medicine, 317*, 1630-1634.

Hepner, K. A., Morales, L. S., Hays, R. D., Edelen, M. O., & Miranda, J. (2008). Evaluating differential item functioning of the PRIME-MD mood module among impoverished black and white women in primary care. *Women's Health Issues, 18*, 53-61.

Hoge, C. W., Auchterlonie, J. L., & Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from

deployment to Iraq or Afghanistan. *Journal of the American Medical Association, 295*, 1023-1032.

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*, 209-218.

Horowitz, M., Wilner, N., & Kaltreider, N. (1980). Signs and symptoms of posttraumatic stress disorder. *Archives of General Psychiatry, 37*, 85-92.

Jackson, E. D., Payne, J. D., Nadel, L., & Jacobs, W. J. (2005). Stress differentially modulates fear conditioning in healthy men and women. *Biological Psychiatry, 59*, 516-522.

Jimenez-Vasquez, P. A., Mathe, A. A., Thomas, J. D., Riley, E. P., & Ehlers., C. L. (2001). Early maternal separation alters neuropeptide Y concentrations in selected brain regions in adult rats. Developmental Brain Research*, 34*, 405-412.

Jimenez-Vasquez, P. A., Overstreet, D. H., & Mathe, A. A. (2000). Neuropeptide Y in male and female brains of flinders sensitive line, a rat model of depression: Effects of electroconvulsive stimuli. *Journal of Psychiatric Research, 34*, 405-412.

Johnson, D. M., Delahanty, D. L., & Pinna, K. (2008). The cortisol awakening response as a function of PTSD severity and abuse chronicity in sheltered battered women. *Journal of Anxiety Disorders, 22*, 793-800.

Jordan, B. K., Marmar, C. R., Fairbank, J. A., Schlenger, W. E., Kulka, R. A., Hough, R. L., & Weiss, D. S. (1992). Problems in families of male Vietnam veterans with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 60*, 916-926.

Kanter, E. D., Wilkinson, C. W., Radant, A. D., Petrie, E. C., Dobie, D. J., McFall, M. E., Peskind, E. R., et al. (2001). Glucocorticoid feedback sensitivity and adrenocortical responsiveness in posttraumatic stress disorder. *Biological Psychiatry, 50*, 238-245.

Kardiner, A. (1941). *The traumatic neurosis of war*. New York: Hoeber.

Karl, T., Duffy, L., & Herzog, H. (2008). Behavioural profile of a new mouse model for NPY deficiency. *European Journal of Neuroscience, 28*, 173-180.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56*, 85-90.

Kessler, R. C, Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 52*, 1048-1060.

Kilpatrick, D., Resnick, H., Saunders, B., & Best, C. (1998). Rape, other violence against women, and posttraumatic stress disorder. In B. P. Dohrenwend (Ed.), *Adversity, stress, and psychopathology* (pp. 161-176). New York: Oxford University Press.

Kilpatrick, D. & Saunders, B. (1997). *Prevalence and consequences of child victimization: Results from the National Survey of Adolescents* (Final Report, Grant No. 93-IJ-CX-0023). Charleston, SC: Authors.

Kim, S. & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test on detection of differential item functioning. *Applied Measurement in Education, 8*, 291-312.

Kimerling, R., Ouimette, P., & Weitlauf, J. C. (2007). Gender issues in PTSD. In M. J. Friedman, T. M. Keane, & P. A. Resick (Eds.), *Handbook of PTSD: Science and practice* (pp. 207-228). New York: Guilford Press.

Komesaroff, P. A., Esler, M. D., Clarke, I. J., Fullerton, M. J., & Funder, J. W. (1998). Effects of estrogen and estrous cycle on glucocorticoid and catecholamine responses to stress in sheep. *American Journal of Physiology, 275*, E671-E678.

Komesaroff, P. A., Esler, M. D., & Sudhir, K. (1999). Estrogen supplementation attenuates glucocorticoid and catecholamine responses to mental stress in perimenopausal women. *Journal of Clinical Endocrinology and Metabolism, 84*, 606-610.

Kosten, T. R., Mason, J. W., Giller, E. L., Ostroff, R. B., & Harkness, L. (1987). Sustained urinary norepinephrine and epinephrine elevation in post-traumatic stress disorder. *Psychoneuroendocrinology, 12*, 13-20.

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., et al. (1990). *Trauma and the Vietnam War generation: Report of findings from the National Vietnam Veterans Readjustment Study*. New York: Brunner/Mazel.

Kutz, I. & Dekel, R. (2006). Follow-up of victims of one terrorist attack in Israel: ASD, PTSD, and the perceived threat of Iraqi missile attacks. *Personality and Individual Differences, 40*(8), 1579-1589.

Lee, R., Geracioti, T. D., Kasckow, J. W., & Coccaro, E. F. (2005). Childhood trauma and personality disorder: Positive correlation with adult CSF corticotropin-releasing factor concentrations. *American Journal of Psychiatry, 162*, 995-997.

Lemieux, A. M. & Coe, C. L. (1995). Abuse-related posttraumatic stress disorder: Evidence for chronic neuroendocrine activation in women. *Psychosomatic Medicine, 57*, 105-115.

Lewandowski, J., Pruszczyk, P., Elaffi, M., Chodakowska, J., Wocial, B., Switalska, H., Januszewicz, W., et al. (1998). Blood pressure, plasma NPY and catecholamines during physical exercise in relation to menstrual cycle, ovariectomy, and estrogen replacement. *Regulatory Peptides, 75-76*, 239-245.

Liu, D., Caldji, C., Sharma, S., Plotsky, P. M., & Meaney, M. J. (2000). Influence of neonatal rearing conditions on stress-induced adrenocorticotropin responses and norepinephrine release in the hypothalamic paraventricular nucleus. *Journal Neuroendocrinology, 12*, 5-12.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing.

Luine, V. (2002). Sex differences in chronic stress effects on memory in rats. *Stress, 5*, 205-216.

MacGruder, K., Frueh, B. C., Knapp, R. G., Davis, L., Hamner, M. B., Martin, R. H., et al. (2005). Prevalence of posttraumatic stress disorder in Veterans Affairs primary care clinics. *General Hospital Psychiatry, 27*, 169-179.

McEwen, B. S. (2000). The neurobiology of stress: From serendipity to clinical relevance. *Brain Research, 886*, 172-189.

Mellman, T. A. (1997). Psychobiology of sleep disturbances in posttraumatic stress disorder. *Annals of the New York Academy of Sciences, 821*, 142-149.

Menard, S. W. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Meyer, J. S. & Quenzer, L. F. (2005). *Psychopharmacology: Drugs, the brain, and behavior*. Sunderland, MA: Sinauer Associates.

Milad, M. R., Goldstein, J. M., Orr, S. P., Wedig, M. M., Klibanski, A., Pitman, R. K., & Rauch, S. L. (2006). Fear conditioning and extinction: Influence of sex and menstrual cycle in healthy humans. *Behavioral Neuroscience, 120*, 1196-1203.

Morgan, C. A., Wang, S., Southwick, S. M., Rasmusson, A., Hazlett, G., Hauger, R. L., & Charney, D. S. (2000). Plasma neuropeptide-Y concentrations in humans exposed to military survival training. *Biological Psychiatry, 47*, 902-909.

Moore, B. A. & Reger, G. M. (2007). Historical and contemporary perspectives of combat stress and the Army Combat Stress Control Team. In C. R. Figley & W. P. Nash (Eds.), *Combat stress injury: Theory, research, and management* (pp.161-181). New York: Routledge.

MULTILOG (Version 7.0.2327.3) [Computer software]. Lincolnwood, IL: Scientific Software International.

Muthen, L. K. & Muthen, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Muthen & Muthen.

Muthen, L. K. & Muthen, B. O. (2006). Mplus (Version 4.0) [Computer software]. Los Angeles: Muthen & Muthen.

National Archives and Records Administration (2007). *Statistical information about casualties of the Vietnam War*. Retrieved June 18, 2009, from http://www.archives.gov/research/vietnam-war/casualty-statistics.html

Nelson-Goff, B. S., Crow, J. R., Reisbig, A. M., & Hamilton, S. (2007). The impact of individual trauma symptoms of deployed soldiers on relationship satisfaction. *Journal of Family Psychology, 21*, 344-353.

Nemeroff, C. B., Bremner, J. D., Foa, E. B., Mayberg, H. S., North, C. S., & Stein, M. B. (2006). Posttraumatic stress disorder: A state-of-the-science review. *Journal of Psychiatric Research, 40*, 1-21.

Norris, F. H. (1992). Epidemiology of trauma: Frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60*, 409-418.

Norris, F. H., Murphy, A. D., Baker, C. K., Perilla, J. L., Rodriguez, F. G., & Rodriguez, J. G. (2003). Epidemiology of trauma and posttraumatic stress disorder in Mexico. *Journal of Abnormal Psychology, 112*(4), 646-656.

Norris, F. H. & Perilla, J. (1996). The Revised Civilian Mississippi Scale for PTSD. *Journal of Traumatic Stress, 9,* 285-298.

Norris, F. H., Perilla, J. L., Ibanez, G. E., & Murphy, A. D. (2001). Sex differences in symptoms of posttraumatic stress: Does culture play a role? *Journal of Traumatic Stress, 14,* 7-28.

Office of the USAF Surgeon General (2005, December 8). *Post-deployment health reassessment (PDHRA) procedures for active duty airmen.* [Memorandum to all USAF major commands].

Office of the USAF Surgeon General (2008). *Post-deployment health reassessment (PDHRA): Application user's guide*. Retrieved January 5, 2010, from https://kx.afms.mil

Orlando, M. & Marshall, G. N. (2002). Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. *Psychological Assessment, 14,* 50-59.

Orlando, M. & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24,* 50-64.

Orme, J. G. & Combs-Orme, T. (2009). *Multiple regression with discrete independent variables*. New York: Oxford University Press.

Ouimette, P., Wade, M., Prins, A., & Schohn, M. (2008). Identifying PTSD in primary care: Comparison of the Primary Care-PTSD screen (PC-PTSD) and the General Health Questionnaire (GHQ). *Anxiety Disorders, 22*, 337-343.

Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychological Bulletin, 129,* 52-73.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.

Perkonigg, A. & Wittchen, H. (1999). Prevalence and comorbidity of traumatic events and posttraumatic stress disorder in adolescents and young adults. In A. Maercker, M.

Schutzwohl, & Z. Solomon (Eds.), *Post-traumatic stress disorder: A lifespan developmental perspective* (pp.113-133). Kirkland, WA: Hogrefe and Huber.

Peters, L., Issakidis, C., Slade, T., & Andrews, G. (2006). Gender differences in the prevalence of DSM-IV and ICD-10 PTSD. *Psychological Medicine, 36*. 81-89.

Pitman, R. K. & Orr, S. P. (1990). Twenty-four hour urinary cortisol and catecholamine excretion in combat-related posttraumatic stress disorder. *Biological Psychiatry, 27*, 245-247.

Plotsky, P. M., & Meaney, M. J. (1993). Early, postnatal experience alters hypothalamic corticotrophin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. *Molecular Brain Research, 18*, 195-200.

Prins, A., Ouimette, P., Kimerling, R., Cameron, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., et al. (2003). The primary care PTSD screen (PC-PTSD): Development and operating characteristics. *Primary Care Psychiatry, 9*, 9-14.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495-502.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 53*, 495-502.

Rasmusson, A. M., & Friedman, M. J. (2002). Gender issues in the neurobiology of PTSD. In R. Kimerling, P. Ouimette, & J. Wolfe (Eds.), *Gender and PTSD* (pp. 43-75). New York: Guilford Press.

Rasmusson, A. M., Hauger, R. L., Morgan, C. A., Bremner, J. D., Charney, D. S., & Southwick, S. M. (2000). Low baseline and Yohimbine-stimulated plasma neuropeptide Y (NPY) levels in combat-related PTSD. *Biological Psychiatry, 47*, 526-539.

Rasmusson, A. M., Lipschitz, D. S., Wang, S., Hu., S., Vojvoda, D., Bremner, D., et al. (2001). Increased pituitary and adrenal reactivity in premenopausal women with posttraumatic stress disorder. *Biological Psychiatry, 50*, 965-977.

Rasmusson, A. M., Vythilingam, M., & Morgan, C. A. (2003). The neuroendocrinology of posttraumatic stress disorder: New directions. *CNS Spectrums, 8*, 651-656; 665-667.

Reeve, B. B. (2000). *Item- and scale-level analysis of clinical and non-clinical sample responses to the MMPI-2 depression scales employing item response theory* (Doctoral dissertation). Retrieved from Dissertation Abstracts International. (Accession number: 2000-95020-218).

Reise, S. P. & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45-58.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.

Reise, S. P. & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27*, 133-144.

Resnick, H. S., Kilpatrick, D. G., Dansky, B. S, Saunders, B. E., & Best, C. L. (1993). Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. *Journal of Consulting and Clinical Psychology, 61*, 984-991.

Rilling, J. K., Winslow, J. T., O'Brien, D., Gutman, D. A., Hoffman, J. M., & Kilts, C. D. (2001). Neural correlates of maternal separation in rhesus monkeys. *Biological Psychiatry, 49*, 146-157.

Rinne, T., de Kloet, E. R., Wouters, L., Goekoop, J. G., DeRijk, R. H., van den Brink, W. (2002). Hyperresponsiveness of hypothalamic-pituitary-adrenal axis to combined dexamethasone/corticotropin-releasing hormone challenge in female Borderline Personality Disorder subjects with a history of sustained childhood abuse. *Biological Psychiatry, 52*, 1102-1112.

Rosen, G. (1975). Nostalgia: A 'forgotten' psychological disorder. *Psychological Medicine, 5*. 340-354.

Sackett, D. L. & Straus, S. (1998). On some clinically useful measures of the accuracy of diagnostic tests. *Advanced Clinical Practice Journal Club, 129*, A17-A19.

Saigh, P. A. & Bremner, J. D. (1999). The history of posttraumatic stress disorder. In P. A. Saigh & J. D. Bremner (Eds.), *Posttraumatic stress disorder: A comprehensive text* (pp.1-17). Boston: Allyn and Bacon.

Saunders, J. A., Morrow-Howell, N., Spitznagel, E., Dore, P., Proctor, E. K., & Pescarino, R. (2006). Imputing missing data: A comparison of methods for social work researchers. *Social Work Research, 30*, 19-31.

Saxe, G. & Wolfe, J. (1999). Gender and posttraumatic stress disorder. In P. A. Saigh & J. D. Bremner (Eds.), *Posttraumatic stress disorder: A comprehensive text* (pp.160-179). Boston: Allyn and Bacon.

Schutt, R. K. (2006). *Investigating the social world: The process and practice of social research*. Thousand Oaks, CA: Sage Publications.

Shea, A. K., Streiner, D. L., Fleming, A., Kamath, M. V., Broad, K., & Steiner, M. (2007). The effect of depression, anxiety, and early life trauma on the cortisol awakening response during pregnancy: Preliminary results. *Psychoneuroendocrinology, 32*, 1013-1020.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., et al. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*, 22-33.

Simmons, C. A. (2007). Speculation as to why women "get" PTSD more often than men. *Women & Therapy, 30*, 85-98.

Simonoff, J. S. (2003). *Analyzing categorical data*. New York: Springer-Verlag.

Sita, A. & Miller, S. B. (1996). Estradiol, progesterone, and cardiovascular response to stress. *Psychoneuroendocrinology, 21*, 339-346.

Smith, T. C., Ryan, M. A., Wingard, D. L., Slymen, D. J., Sallis, J. F., & Kritz-Silverstein, D. (2008). New onset and persistent symptoms of post-traumatic stress disorder self reported after deployment and combat exposures: Prospective population based US military cohort study. *British Medical Journal, 336*, 371-375.

Sofowara, G. S., Singh, I., He, H. B., Wood, A. J. J., & Stein, C. M. (2005). Effect of acute transdermal estrogen administration on basal, mental stress, and cold pressor-induced sympathetic responses in postmenopausal women. *Clinical Autonomic Research, 15*, 193-199.

Southwick, S. M., Bremner, J. D., Rasmusson, A., Morgan, C. A., Arnsten, A., & Charney, D. S. (1999). Role of norepinephrine in the pathophysiology and treatment of posttraumatic stress disorder. *Biological Psychiatry, 46*, 1192-1204.

Southwick, S. M., Krystal, J. H., Morgan, C. A., Johnson, D., Nagy, L. M., Nicolaou, A., et al. (1993). Abnormal noradrenergic function in posttraumatic stress disorder. *Archives of General Psychiatry, 50*, 266-274.

Southwick, S. M., Rasmusson, A., Barron, J., & Arnsten, A. (2005). Neurobiological and neurocognitive alterations in PTSD. In J. J. Vasterling & C. R. Brewin (Eds.), *Neuropsychology of PTSD: Biological, cognitive, and clinical perspectives* (pp. 27-58). New York: Guilford Press.

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of the PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association, 282*, 1737-1734.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1995). *Structured Clinical Interview for DSM-IV—Patient Version*. New York: Biometric Research Department, New York State Psychiatric Institute.

Stein, M. B., Walker, J. R., & Forde, D. R. (2000). Gender differences in susceptibility to posttraumatic stress disorder. *Behaviour Research and Therapy, 38*, 619-628.

Stein, M. B., Walker, J. R., Hazen, A. L. & Forde, D. R. (1997). Full and partial posttraumatic stress disorder: Findings from a community survey. *American Journal of Psychiatry, 154*, 1114-1119.

Steinberg, L. & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods, 11*(4), 402-415.

Stone, C. A. (2000). Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models. *Journal of Educational Measurement, 37*, 58-75.

Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in Medicine, 19*, 1651-1683.

Tersman, Z., Collins, A., & Eneroth, P. (1991). Cardiovascular responses to psychological and physiological stressors during the menstrual cycle. *Psychosomatic Medicine, 53,* 185-197.

Thissen, D. (2001). *IRTLRDIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. Retrieved from http://www.unc.edu/~dthissen/dl.html

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Tolin, D. F. & Foa, E. B. (2006). Sex differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin, 132*(6), 959-992.

Turner, J. B., Turse, N. A., & Dohrenwend, B. P. (2007). Circumstances of service and gender differences in war-related PTSD: Findings from the National Vietnam Veteran Readjustment Study. *Journal of Traumatic Stress, 20,* 643-649.

van der Kolk, B. A., van der Hart, O., & Burbridge, J. (2002). Approaches to the treatment of PTSD. In M. B. Williams & J. F. Sommer, Jr. (Eds.), *Simple and complex post-traumatic stress disorder: Strategies for comprehensive treatment in clinical practice* (pp. 23-45). Binghamton, NY: Haworth Maltreatment and Trauma Press.

van Stegeren, A. H., Goekoop, R., Everaerd, W., Scheltens, P., Barkhof, F., Kuijer, J. P., & Rombouts, S. A. (2005). Noradrenaline mediates amygdala activation in men and women during encoding of emotional material. *Neuroimage, 24,* 898-909.

Vaiva, G., Brunet, A., Lebigot, F., Boss, V., Ducrocq, F., Devos, P., et al. (2003). Fright (effroi) and other peritraumatic responses after a serious motor vehicle accident: Prospective influence on acute PTSD development. *Canadian Journal of Psychiatry, 48*, 395-401.

Vrana, S. R. & Lauterbach, D. (1994). Prevalence of traumatic events and posttraumatic stress. *Journal of Traumatic Stress, 7,* 289-302.

Vythilingam, M., Heim, C., Newport, J., Miller, A. H., Anderson, E., Bronen, R., Brummer, M. et al. (2002). Childhood trauma associated with smaller hippocampal volume in women with major depression. *American Journal of Psychiatry, 159,* 2072-2080.

Walker, J. L, Carey, P. D., Mohr, N., Stein, D. J., & Seedat S. (2004). Gender differences in the prevalence of childhood sexual abuse and in the development of pediatric PTSD. *Archives of Women's Mental Health, 7*, 111-121.

Wang, W. C. & Yeh, Y. L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement, 27*, 479-498.

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993). *The PTSD checklist-civilian version (PCL-C)*. Boston, MA: National Center for PTSD.

Whooley, M. A., Avins, A. L., Miranda, J., & Browner, W. S. (1997). Case-finding instruments for depression: Two questions are as good as many. *Journal of General Internal Medicine, 12,* 439-445.

Women's Research and Education Institute (2007). *Active duty service personnel by branch of service, officer/enlisted status and sex as of 30 September 2007*. Retrieved June 17, 2009, from http://www.wrei.org

Wooley, C. F. (2002). *The irritable heart of soldiers and the origins of Anglo-American cardiology*. Burlington, VT: Ashgate Publishing.

World Health Organization (1997). *Composite International Diagnostic Interview (CIDI), Version 2.1*. Geneva, Switzerland: Author.

Wright, K. M., Bliese, P. D., Thomas, J. L., Adler, A. B., Eckford, R. D., & Hoge, C. W. (2007). Contrasting approaches to psychological screening with U.S. combat soldiers. *Journal of Traumatic Stress, 20*, 965-975.

Yehuda, R. (2001). Biology of posttraumatic stress disorder. *Journal of Clinical Psychiatry, 62* (Supplement 17), 41-46.

Yehuda, R. (2009). Stress hormones and PTSD. In P. J. Shiromani, T. M. Keane, & J. E. LeDoux (Eds.), *Post-traumatic stress disorder: Basic science and clinical practice* (pp. 257-275). Totowa, NJ: Humana Press.

Yehuda, R., Boisoneau, D., Lowy, M. T., & Giller, E. L. (1995). Dose-response changes in plasma cortisol and lymphocyte glucocorticoid receptors following dexamethasone administration in combat veterans with and without posttraumatic stress disorder. *Archives of General Psychiatry, 52*, 583-593.

Yehuda, R., Boisoneau, D., Mason, J. W., & Giller, E. L. (1993). Relationship between lymphocyte glucocorticoid receptor number and urinary-free cortisol excretion in mood, anxiety, and psychotic disorder. *Biological Psychiatry, 34*, 18-25.

Yehuda, R., Siever, L. J., Teicher, M. H., Levengood, R. A., Gerber, D. K., Schmeidler, J. & Yang, R. (1998). Plasma norepinephrine and 3-methoxy-4-hydroxiphenylglycol concentrations and severity of depression in combat posttraumatic stress disorder and major depressive disorder. *Biological Psychiatry, 44*, 56-63.

Yehuda, R., Southwick, S. M., Giller, E. L., Xiaowan, M. A., & Mason, J. W. (1992). Urinary catecholamine excretion and severity of PTSD symptoms in Vietnam combat veterans. *Journal of Nervous and Mental Disease, 180*, 321-325.

Yehuda, R., Teicher, M., Trestman, R., Levengood, R., & Siever, L. (1996). Cortisol regulation in posttraumatic stress disorder and major depression: A chronobiological analysis. *Biological Psychiatry, 40*, 79-88.

Yen, W. M. & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 111-153). Westport, CT: Praeger Publishers.

Young, E. A., Altemus, M., Parkison, V., & Shastry, S. (2001). Effects of estrogen antagonists and agonists on the ACTH response to restraint stress in female rats. *Neuropsychopharmacology, 25*, 881-891.

Young, E. A. & Breslau, N. (2004a). Cortisol and catecholamines in posttraumatic stress disorder: An epidemiological study. *Archives of General Psychiatry, 61*, 394-401.

Young, E. A. & Breslau, N. (2004b). Saliva cortisol in posttraumatic stress disorder: A community epidemiological study. *Biological Psychiatry, 56*, 205-209.

Zhou, X. H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. New York: John Wiley & Sons.

Zukowska-Grojec, Z. (1995). Neuropeptide Y: A novel sympathetic stress hormone and more. *Annals of the New York Academy of Sciences, 771*, 219-233.

VITA

Mark A. Oliver completed his Bachelors of Science degree in Psychology from Virginia Tech in 1993. After working in various human service positions primarily with children, in 1996 he was accepted to the School of Social Work at the University of North Carolina at Chapel Hill. Mark graduated with his Masters degree in Social Work (MSW) in 1998 with a focus on clinical practice. Soon after graduation he accepted a position as Lead Clinical Staff at Southeastern Children's Home in South Carolina. In this position, Mark conducted individual and group therapy, provided case management, and supervised all aspects of treatment with the adolescent residents in two moderate management group homes. In February of 2001, Mark was commissioned as an officer in the Biomedical Science Corps of the United States Air Force, completing Commissioned Officer Training in May of that year. His first assignment was to Sheppard Air Force Base in Wichita Falls, Texas, where he began as a staff clinical social worker in the mental health clinic of the base hospital. Mark was promoted to element chief of the Alcohol and Drug Abuse Prevention and Treatment Program in December 2002. Then, from March to May of 2003 he was deployed to Kuwait in support of Operation IRAQI FREEDOM. There, Mark served as the element chief of the lone mental health clinic serving more than 8000 active duty, reserve, and National Guard members of the Air Force, Army, Navy, and Marines. In 2004, he began his next assignment as the element chief of the Family Advocacy Program at Eielson Air Force Base, Alaska. In this remote location Mark held several additional duties such as staff clinical social worker in the mental health clinic, Behavioral Health Optimization Program clinician in the Family Practice Center, Special Needs Coordinator, base Suicide

Prevention Coordinator, and alternate flight commander. He was selected by the Air Force in 2007 to pursue doctoral education in social work at a civilian institution. After being accepted into the College of Social Work at the University of Tennessee and beginning his Ph.D. studies, Mark was the co-author of a publication on motivational interviewing and interpersonal violence. Mark is married with three wonderful children. In July 2010, he will begin his next assignment on the academic faculty at the United States Air Force Academy in Colorado.