Journal of Traumatic Stress
April 2013, *26,* 175–183

Gender Differences in Posttraumatic Stress Symptoms Among OEF/OIF Veterans: An Item Response Theory Analysis

Matthew W. King,[1] Amy E. Street,[1,2] Jaimie L. Gradus,[1,2,3] Dawne S. Vogt,[1,2] and Patricia A. Resick[1,2]
[1] Women's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts, USA
[2] Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts, USA
[3] Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

Establishing whether men and women tend to express different symptoms of posttraumatic stress in reaction to trauma is important for both etiological research and the design of assessment instruments. Use of item response theory (IRT) can reveal how symptom reporting varies by gender and help determine if estimates of symptom severity for men and women are equally reliable. We analyzed responses to the PTSD Checklist (PCL) from 2,341 U.S. military veterans (51% female) who completed deployments in support of operations in Afghanistan and Iraq (Operation Enduring Freedom/Operation Iraqi Freedom [OEF/OIF]), and tested for differential item functioning by gender with an IRT-based approach. Among men and women with the same overall posttraumatic stress severity, women tended to report more frequent concentration difficulties and distress from reminders whereas men tended to report more frequent nightmares, emotional numbing, and hypervigilance. These item-level gender differences were small (on average $d = 0.05$), however, and had little impact on PCL measurement precision or expected total scores. For practical purposes, men's and women's severity estimates had similar reliability. This provides evidence that men and women veterans demonstrate largely similar profiles of posttraumatic stress symptoms following exposure to military-related stressors, and some theoretical perspectives suggest this may hold in other traumatized populations.

It is well established that women on average are more likely to meet criteria for posttraumatic stress disorder (PTSD) following trauma exposure than men (e.g., Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Among potential explanations being explored, one possibility is a systematic gender difference in posttraumatic symptom expression. Tolin and Foa (2006) advance a hypothesis, for example, that women may be more likely to develop internalizing psychopathology after trauma exposure that is consistent with anxiety disorders like PTSD, whereas men may be predisposed towards externalized expressions of distress. It follows therefore that stress reactions typical of women are not more pathological than men's, only more likely to meet full diagnostic criteria according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association, 2000).

To the extent that such hypotheses are tenable, it logically implies that at least some individual *DSM-IV-TR* PTSD symptoms are a better descriptive match to women's stress reactions than to men's. Studies reporting gender differences in symptom prevalence tend to support the general expectation, but have not identified a consistent set of implicated symptoms. For example, among general hospital outpatients with PTSD, Zlotnick, Zimmerman, Wolfsdorf, and Mattia (2001) found that women were more likely to report reexperiencing symptoms. Implicating a different cluster, Breslau, Chilcoat, Kessler, Peterson, and Lucia (1999) found the largest gender disparity to be in avoidance/numbing symptoms among community members exposed to assaultive violence. Peters, Issakidis, Slade, and Andrews (2006) found in a trauma-exposed Australian sample that women reported more reexperiencing, insomnia, exaggerated startle, and avoidance of thoughts and feelings, but that men reported more feelings of emotional distance.

Some studies have, however, produced broadly disconfirmatory findings. Green (2003) compared symptom profiles of male and female patients in a PTSD specialty clinic and found only that men were more likely to report irritability. While testing a PTSD latent class model among young urban adults exposed to traumatic events, Chung and Breslau (2008) found that only trauma type, and not gender, was associated with differential symptom profiles. Armour et al. (2011) tested gender invariance

This research was supported by the National Center for PTSD, Office of Mental Health Services, Department of Veterans Affairs. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The funding agency was not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Correspondence concerning this article should be addressed to Matthew W. King, VA Boston Healthcare System (116B-3), 150 S. Huntington Avenue, Boston, MA 02130. E-mail: matthew.king2@va.gov

DOI: 10.1002/jts.21802

of PTSD latent structure using factor analysis among war-exposed Bosnian adolescents, and found that a generally well-supported factor model (i.e., King, Leskin, King, & Weathers, 1998) actually fit less well among girls, suggesting one customary conception of the structure of PTSD symptoms characterized their symptomatology less adequately, not more, than boys.

Identifying differences in women's and men's PTSD symptom expression is made difficult by sociocultural influences on posttraumatic symptomatology. Factors associated with vocation appear to be moderators; contrary to most population surveys, several studies conducting gender comparisons within military and police samples revealed no gender differences in risk for PTSD (Brewin, Andrews, & Valentine, 2000; Pole et al., 2001), and the effect of combat exposure on PTSD in post-deployed service members has been found to be no greater in women than in men among OEF/OIF veterans (Maguen, Luxton, Skopp, & Madden, 2012; Vogt et al., 2011). Culture too appears to play a role; a cross-cultural comparison of natural disaster survivors (Norris, Perilla, Ibañez, & Murphy, 2001) found that gender differences in PTSD symptom endorsement were generally smallest among the most egalitarian ethnic group and largest among the most patriarchal ethnic group.

Despite the inherent complexities, there is intrinsic scientific and practical value in exploring whether individual PTSD-related symptoms vary in their relevance to men and women's typical posttraumatic maladjustment. First, this research provides the foundation for identifying gender-specific mechanisms that contribute to the onset and course of PTSD. Second, this issue informs the design of PTSD assessment instruments. If PTSD as a diagnostic construct is not functioning as a unitary entity across men and women, the psychometric issue is one of measurement invariance: Are men and women's estimates of symptom severity equally reliable?

To advance understanding of these two issues, we investigated PTSD-related symptom expression in a large sample of U.S. veteran men and women who completed deployments in support of Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF). Gender differences were examined using item response theory (IRT; Lord, 1980) methodology, a framework for modeling the association between responses on an assessment instrument and an individual's standing on the latent construct the instrument measures (e.g., severity of posttraumatic stress pathology). Using an IRT model, the influence of gender can be separated from the effect of overall severity on responses to PTSD symptom items, and the items significantly influenced by gender can be identified.

Although IRT has been applied to examine the properties of PTSD symptom scales like the PTSD Checklist (PCL; Weathers, Huska, & Keane, 1991) in previous studies (e.g., Orlando & Marshall, 2002), only one known study has used IRT to assess specifically for the presence of differential PTSD symptom responding by gender. Palm, Strong, and MacPherson (2009) analyzed responses to the PTSD module of the National Comorbidity Study-Replication (Kessler & Merikangas, 2004) interview in community members who met *DSM-IV-TR* Criterion A and acknowledged some persistent emotional difficulties. They observed that among women and men with the same estimated severity level, women more frequently endorsed exaggerated startle and feelings of emotional distance, whereas men more frequently endorsed intrusive thoughts, nightmares, irritability, and foreshortened future. Also, flashback endorsement was a less reliable indicator of men's overall severity than women's. Unfortunately, the effects of these gender differences on the reliability of instrument measurement were not reported, and further, the data were not especially well-suited for IRT methods. This limits the conclusions that can be drawn and indicates the need for conceptual replication.

The current study extends this literature by testing for gender differences in the relevance of individual PTSD-related symptoms in a veteran population at high risk for the disorder. One expectation, consistent with hypotheses noted above, was that symptoms evidencing gender differences would on the whole be more salient to women (i.e., preferentially endorsed and more reliable indicators). In line with extant findings in several military samples, however, we also expected that any observed gender differences in symptom expression would be small and exert minimal effect on scale reliability.

Method

Participants and Procedure

Data were drawn from a cross-sectional mail survey of U.S. military veterans who completed deployments in support of OEF/OIF. The primary purpose of the larger survey was to examine deployment experiences and postdeployment adjustment. Complete details on the participant pool and sampling procedure were provided by Street, Gradus, Giasson, Vogt, and Resick (2012). In brief, 6,000 veterans were randomly selected from a roster of all OEF/OIF veterans separated from active duty service held by the Department of Veterans Affairs Environmental Epidemiology Service, with women deliberately oversampled to ensure comparable gender proportions; 1,139 men and 1,209 women returned valid surveys. This reflects a response rate of 48.6% after accounting for ineligible responders and estimated ineligibility among non-responders. All procedures were approved by the Institutional Review Board at the VA Boston Healthcare System.

Men in the sample were on average slightly older than women (men: $M = 37.0$ years, $SD = 10.0$; women: $M = 34.4$ years, $SD = 8.9$; Welch's $t = 6.53$, Cohen's $d = 0.27$) and served slightly longer total OEF/OIF deployments (men: $M = 12.2$ months, $SD = 8.5$; women: $M = 10.8$ months, $SD = 7.4$; Welch's $t = 4.01$, Cohen's $d = 0.17$). Women were more likely than men to be non-White (women: 68.4% White, 28.0% non-White; men: 80.8% White, 15.6% non-White; $\chi^2 = 52.6$, $OR = 2.13$). High proportions of men (81.7%) and women (73.4%) reported combat experiences during deployment, and women also reported a high rate of sexual harassment exposure (51.2%; Street et al., 2012).

Table 1
PCL-M Item Statistics by Gender

Item (*DSM-IV-TR* criterion)	Women			Men			Cohen's *d*
	M	*SD*	$r_{i\text{-}t}$	*M*	*SD*	$r_{i\text{-}t}$	
Intrusive thoughts (B1)	1.93	1.2	.84	2.04	1.2	.85	0.09
Nightmares (B2)	1.76	1.2	.84	1.88	1.2	.81	0.10
Flashbacks (B3)	1.57	1.1	.83	1.65	1.1	.82	0.07
Distressed by reminders (B4)	1.95	1.3	.86	1.93	1.3	.85	0.02
Physiological reactivity (B5)	1.74	1.2	.84	1.76	1.2	.83	0.01
Avoids thoughts (C1)	1.99	1.3	.82	1.99	1.3	.80	0.00
Avoids places (C2)	1.77	1.3	.84	1.76	1.2	.82	0.01
Memory lapses (C3)	1.62	1.1	.70	1.63	1.1	.66	0.01
Anhedonia (C4)	1.97	1.4	.83	2.02	1.3	.81	0.04
Feelings of detachment (C5)	2.14	1.4	.84	2.21	1.4	.83	0.05
Emotionally numb (C6)	1.88	1.3	.80	2.07	1.4	.78	0.14
Foreshortened future (C7)	1.70	1.3	.72	1.79	1.3	.76	0.07
Difficulty sleeping (D1)	2.49	1.5	.71	2.47	1.5	.76	0.01
Irritability (D2)	2.37	1.5	.82	2.44	1.5	.83	0.05
Difficulty concentrating (D3)	2.28	1.4	.82	2.18	1.3	.83	0.07
Hypervigilance (D4)	2.11	1.4	.81	2.34	1.4	.79	0.16
Exaggerated startle (D5)	2.07	1.4	.82	2.09	1.4	.83	0.01

Note. $N = 2{,}341$. PCL-M = PTSD Checklist-Military version; i-t = item total. Item-total correlation is corrected for item overlap and scale reliability.

Measures

Included in the survey was the 17-item PCL-Military version (Weathers et al., 1991). Instructions prompted respondents to rate how much they were bothered in the past month by their reactions to "stressful deployment experiences." Items correspond to the 17 *DSM-IV-TR* PTSD Criterion B, C, and D symptoms and are rated on a Likert scale from 1 = *not at all* to 5 = *extremely*. The PCL is the most frequently used self-report measure of PTSD symptoms and has demonstrated good sensitivity and specificity against structured interview PTSD diagnostic standards in service member and veteran samples (McDonald & Calhoun, 2010).

Of the 2,348 respondents, seven completed no PCL-M items and were excluded. Of those remaining, 73 (37 women) had missing data for an average of 1.6 items ($SD = 1.5$, range: 1–11). Missing responses (constituting 0.3% of possible observations) were filled in using a single-imputation method based on random forest recursive partitioning (Breiman, 2001) and subsequently treated as observed.

Cronbach's α reliability was .97 and did not differ between genders. Item statistics by gender are provided in Table 1. Item-total correlations were high and comparable across groups. Cohen's *d* effect sizes for the gender difference in item means were very small, ranging from < 0.005 to 0.16. Similarly the difference in average PCL-M total score between women ($M = 33.35$, $SD = 18.2$) and men ($M = 34.23$, $SD = 17.9$) was also small ($d = 0.05$). Accordingly, proportions of probable PTSD were not significantly different (men: 23.4%; women: 21.0%), $OR = 0.87$, 95% CI = [0.7, 1.1], using a commonly recommended cutoff score of 50 (Tanielian & Jaycox, 2008).

Data Analysis

Given the ordinal response format of the PCL items, the most appropriate IRT model was the graded response model (GR; Samejima, 1969). The GR model characterizes the likelihood of a particular item response (i.e., choice of 1–5) as a function of both latent person and item characteristics. Both persons and items are conceptualized as varying in degree of severity on a single posttraumatic stress (PTS) dimension. Each person's level of PTS severity is estimated as a latent score falling on this dimension. Each item is associated with four sequential difficulty parameters, one for each response cutpoint (i.e., > 1, > 2, > 3, > 4). These mark the locations on the PTS dimension where a person falling at that point is equally likely to respond above the cutpoint as below it. Between items, comparatively higher difficulty indicates a symptom of more severe PTS (i.e., persons must have higher severity to exceed a given cutpoint). Each item is also associated with a discrimination parameter, which denotes the extent to which responses on the item reliably indicate differences between persons' overall severity scores. Comparatively higher item discrimination suggests the symptom has a stronger association with the underlying PTS dimension.

In theory, all individuals with a given PTS severity level should have the same likelihood of a given item response.

If men and women with the same PTS severity systematically respond to an item in different ways, however, the item is said to demonstrate differential item functioning (DIF; Holland & Wainer, 1993). Gender-based DIF may affect difficulty (i.e., the symptom represents more severe PTS in one gender than in the other), discrimination (i.e., item responses are more strongly associated with the PTS dimension in one gender than in the other), or both.

In the current study, DIF was investigated using an ordinal logistic regression framework (OLR; Crane, Gibbons, Jolley, & van Belle, 2006; Zumbo, 1999). First a single GR model estimated latent severity scores for all persons. For each item, these scores were used as the initial regression predictor such that higher severity naturally predicts greater odds of exceeding any given cutpoint. To assess for gender-based DIF, persons' gender and the interaction of gender and persons' severity were then entered stepwise into the model while accounting for the main effect of severity. The effect of each new term was tested with a model likelihood ratio test χ^2 against the previous (reduced) model. A significant main effect of gender indicates that men and women with the same severity have different odds of exceeding any given cutpoint. If the interaction is also significant, this indicates that the magnitude of the gender difference in odds varies across the different cutpoints. Measures of effect size customary with logistic regression (e.g., McFadden pseudo R^2) were used to quantify the magnitude of DIF.

With DIF items identified, GR models were recalibrated such that DIF item parameters were estimated separately between genders, using non-DIF items as anchors. Several aspects of model results were then examined to assess impact on PCL-M measurement. First, DIF items' difficulty and discrimination were examined to determine the direction of gender differences. Second, for each DIF item we calculated expected item scores, which are GR model predictions scaled on the familiar PCL metric (i.e., 1–5) for persons at a given severity level. Comparing expected item scores between genders allowed us to estimate the likely magnitude of observed score differences among men and women with the same PTS severity. Finally, for each DIF item we compared item information values, which quantify the contribution of the particular item to the total measurement precision of the PCL-M. Gender comparisons on item information reveal whether a particular item reduces uncertainty (i.e., estimation error) in persons' latent severity scores in one gender more than the other.

All analyses were conducted using the *R* statistical programming software (R Development Core Team, 2011), including several contributed packages (Choi, Gibbons, & Crane, 2011; Revelle, 2012; Rizopoulos, 2006).

Results

Unidimensionality Assessment

One assumption when using a GR model is that PCL items load predominantly on a single latent dimension. A size ratio above 4:1 between the first and second common factors and no evidence of major secondary factors from parallel analysis of simulated samples provide strong evidence for essential unidimensionality, given the large sample and high communalities (Slocum-Gori & Zumbo, 2011). Weighted least squares factor analysis was conducted on the polychoric interitem correlation matrix separately by gender. For men, the first two PCL-M eigenvalues were 12.57 and 0.65 (a ratio above 19:1), and for women, the eigenvalues were 12.71 and 0.74 (a ratio above 17:1). Parallel analysis revealed a ratio of the first PCL-M versus first simulated eigenvalue over 26:1 in men and over 42:1 in women, compared to ratios of 3:1 and 4:1, respectively, between the second PCL-M eigenvalue and its simulated counterpart. We determined this combination of findings sufficient to justify proceeding with a GR model.

Identification of DIF Items

The OLR analysis identified five PCL-M items as demonstrating differential functioning between genders: nightmares (*DSM-IV-TR* Criterion B2), distressed by reminders (*DSM-IV-TR* Criterion B4), emotionally numb (*DSM-IV-TR* Criterion C6), difficulty concentrating (*DSM-IV-TR* Criterion D3), and hypervigilance (*DSM-IV-TR* Criterion D4). For these items, Table 2 provides effect sizes for significant terms (at $\alpha = .01$). All five DIF items were characterized by a main effect of gender, indicating that men and women with the same severity had different odds for any given response option. Specifically, men were more likely to endorse a higher level of nightmares, emotional numbing, and hypervigilance, whereas women were more likely to endorse a higher level of reminder distress and concentration difficulties. Additionally, for hypervigilance the interaction of gender and person severity was significant, indicating that the magnitude of the gender difference varied across PTS severity levels. All effect sizes associated with DIF findings were very small by accepted guidelines (Zumbo, 1999). Table 2 provides estimated difficulty and discrimination parameters by gender for the five DIF items. For comparison, among the 12 items that did not demonstrate DIF, the average difficulty estimates were 0.17 ($SD = 0.39$), 0.84 ($SD = 0.32$), 1.36 ($SD = 0.36$), and 1.98 ($SD = 0.37$), and the average discrimination estimate was 2.51 ($SD = 0.42$). Thus DIF items were not systematically different in difficulty or discrimination from other items. In the final GR models (i.e., adjusted for DIF), the marginal reliability of the latent severity scores (expected a posteriori; Wright & Masters, 1982) was .93 in men and .92 in women.

Expected Item Scores

Gender differences in expected scores are most easily examined using graphic plots of the item characteristic curves (ICCs), which reveal the predicted item responses for each point along the PTS severity dimension. The first five panels of Figure 1 show the ICC plots by gender for the five DIF items, with detectable but slight gender differences.

Table 2
Effect Sizes and Graded Response Model Parameters by Gender for PCL-M Items With Differential Item Functioning

Item (*DSM-IV-TR* criterion)	Gender R^2	G × S R^2	Item difficulty >1	>2	>3	>4	Item discrimination	Item information
Nightmares (B2)	.002	.000						
Men			0.19	0.93	1.45	2.13	2.64	6.75
Women			0.39	0.97	1.55	2.18	2.81	7.12
Distressed by reminders (B4)	.002	.000						
Men			0.17	0.80	1.38	2.03	3.23	8.93
Women			0.08	0.83	1.30	1.86	2.87	7.29
Emotionally numb (C6)	.004	.000						
Men			0.08	0.72	1.20	1.89	2.14	4.76
Women			0.34	0.93	1.33	1.97	2.23	4.80
Difficulty concentrating (D3)	.003	.000						
Men			-0.20	0.57	1.20	1.76	2.58	6.56
Women			-0.29	0.45	0.98	1.61	2.50	6.13
Hypervigilance (D4)	.005	.001						
Men			-0.28	0.39	1.05	1.66	2.33	5.61
Women			0.00	0.67	1.12	1.68	2.42	5.50

Note. PCL-M = Posttraumatic Stress Disorder Checklist-Military version. All effect sizes refer to ordinal logistic regression (OLR) likelihood ratio test χ^2s and are McFadden pseudo R^2s. Gender R^2 = main effect of gender; G × S R^2 = interaction of Gender × Person Severity; item difficulty = level of posttraumatic stress (PTS) associated with 50% chance of an item response above the given cutoff (above 1, above 2, etc.). Between genders, a comparatively higher number means that gender reaches a given cutoff at a higher severity on average, so the given item response is indicative of more severe PTS. Item discrimination = magnitude of association between item responses and the underlying PTS dimension; item information = magnitude of the item's contribution to the total measurement precision of the PCL.

In the three left-hand panels of Figure 1, the ICCs for men are shifted slightly left relative to women. This indicates that men are expected to endorse higher levels of nightmares and emotional numbing than women at the same severity level. The ICCs for hypervigilance are similar, but reflecting the significant interaction, converge at higher severity levels, indicating that the gender difference holds only at low to moderate PTS severity; at higher levels, men's and women's expected responses are nearly identical.

In the top two right-hand panels of Figure 1, the ICCs for women are shifted slightly left relative to men. This indicates that women are expected to endorse higher levels of reminder distress and concentration difficulties than men at the same severity level.

Expected scores for each item can be summed to produce expected scores at the instrument level. The final panel of Figure 1 shows the ICC by gender for the entire 17-item PCL-M. The plots for men and women are virtually indistinguishable, suggesting that despite the presence of statistically significant DIF, the impact of gender on observed PCL total scores is negligible.

Item Information

Table 2 provides total item information values by gender; these represent single-number estimates of information summed across the entire PTS dimension. Differences are slight, reflecting the modest effect sizes of the DIF findings. The largest gender difference in information was observed for distressed by reminders, indicating that, although women were more likely to endorse this item than men at a similar PTS severity, women's responses were somewhat less reliable indicators of their overall severity level. This contributed to a very small gender difference in total instrument information across all 17 items (105.80 for men vs. 104.03 for women).

To assess the effects on overall PCL-M reliability, the standard error of persons' severity estimates was calculated separately by gender at each point on the PTS dimension, and plotted in Figure 2. Between groups, a comparatively larger standard error indicates that severity levels in that group are being estimated with greater uncertainty (i.e., larger confidence intervals). Figure 2 reveals that throughout the range of primary interest (moderate to high severity), men's and women's severity estimates have similar reliability. Therefore, the presence of DIF in five items did not appreciably impact measurement precision for either group.

Discussion

We evaluated the PCL-M responses of postdeployed OEF/OIF veterans to identify whether gender influenced item responses above and beyond the effects of posttraumatic stress severity level. Five items demonstrated statistically significant DIF. Among men and women at the same symptom severity level, men generally endorsed higher responses than women for items assessing nightmares (*DSM-IV-TR* Criterion B2), emotional

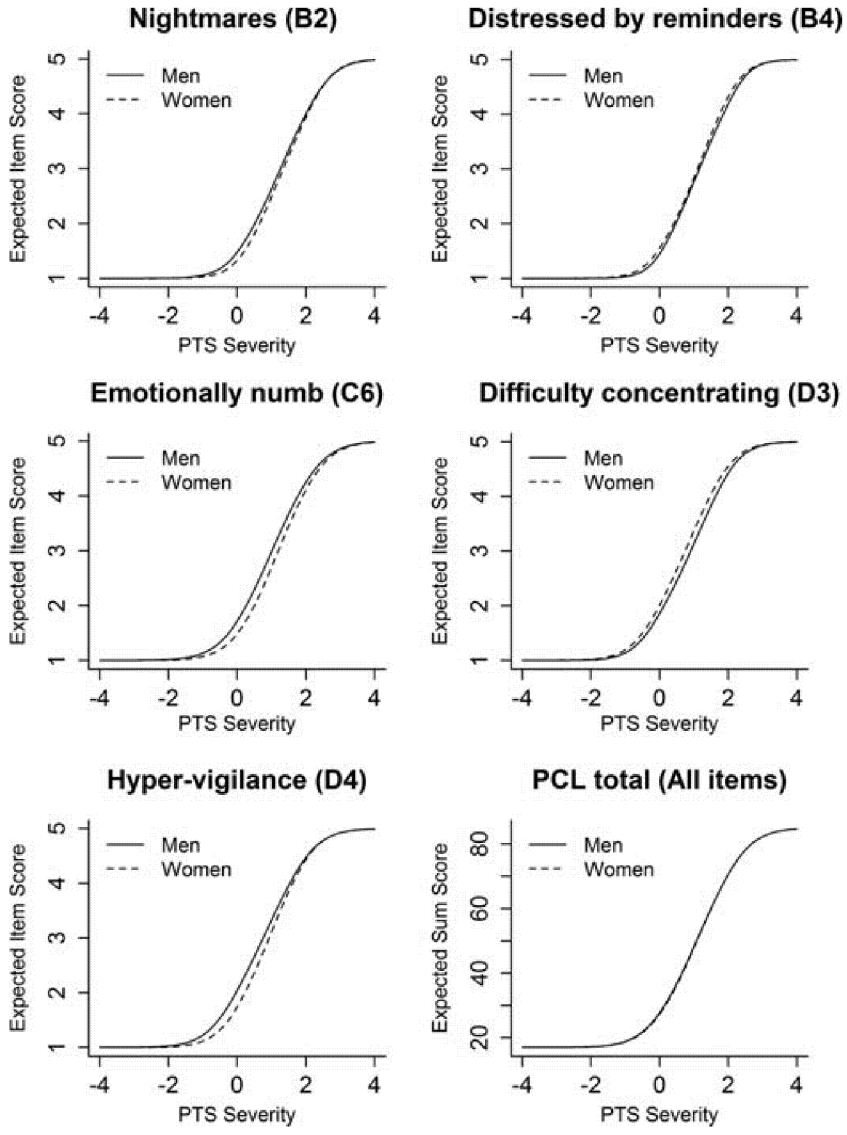

Figure 1. Item characteristic curves by gender for the five PCL-M items with differential item functioning. The ICCs show the expected score difference between men and women at all points along the posttraumatic stress (PTS) severity dimension. The final panel shows the characteristic curve for the overall instrument (PCL-M total score).

numbing (*DSM-IV-TR* Criterion C6), and hypervigilance (*DSM-IV-TR* Criterion D4), the latter only at low to moderate severity levels. Women generally endorsed higher responses on items assessing reminder distress (*DSM-IV-TR* Criterion B4) and concentration difficulties (*DSM-IV-TR* Criterion D3).

The effects of DIF on PCL-M measurement properties were negligible, however. At both the item and instrument levels, gender differences in expected scores were very small. Thus gender comparisons (on individual items or PCL-M total scores) at the level of group means should reflect genuine severity differences with relatively little gender-related bias. Men and women also varied little in reliability of item responses, with one possible exception being the distressed by reminders item, where women's responses were slightly less reliable indicators of their posttraumatic stress severity level than men's. This exerted an extremely small effect at the instrument level, however; for practical purposes, the PCL-M appears to be measuring posttraumatic stress severity with similar reliability for both male and female OEF/OIF veterans.

This investigation explored the hypothesis that *DSM-IV-TR* PTSD symptom criteria are, on the whole, more relevant to women's posttraumatic reactions than men's. This prediction was not supported; although modest gender differences were observed, differentiating symptom reports were not appreciably biased towards women's experiences. Rather, on three of five symptoms with differences, it was men who appeared predisposed to endorse greater symptom severity. The relatively small observed disparities may reflect genuine differences in men's and women's expression of posttraumatic stress. Alternatively, these findings may indicate that gender roles

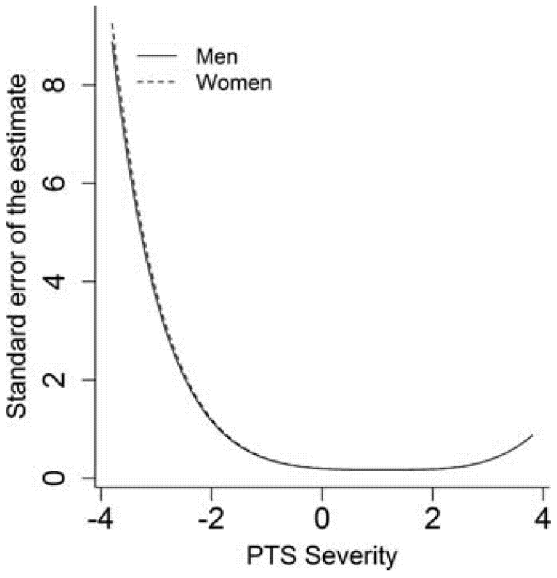

Figure 2. Standard errors of posttraumatic stress (PTS) severity level estimates by gender. A comparatively larger standard error between men and women at any given point indicates that severity estimates for the group at that point are made with greater uncertainty.

differentially influenced the symptoms men and women considered more acceptable to report (cf. McLean & Anderson, 2009). In either case, because gender comparisons in PTSD research have not converged on a consistent set of symptoms, it is difficult to make strong inferences at present.

Further, it is important not to overinterpret observed differences in light of the small effects. In fact, an absence of substantial gender differences aligns with our second prediction. Results generally cohere with extant studies finding few PTSD-related gender differences among military cohorts (Brewin et al., 2000; Tolin & Foa, 2006). This is consistent with the perspective that sociocultural differences across different populations may be a dominant influence on posttraumatic adjustment. Within the military, the work environment is becoming increasingly egalitarian with respect to job training, duties, and exposure to stressful deployment experiences (Hoge, Clark, & Castro, 2007; Street, Vogt, & Dutra, 2009). It is also plausible that women who choose a military profession constitute a unique sociocultural group (e.g., with respect to egalitarian values and motivations) relative to women who choose civilian professions. Factors associated with trauma type or environment of trauma exposure likely also serve as an equalizer of symptom expression between genders, a perspective consistent with previous research (e.g., Chung & Breslau, 2008). In all cases, these factors may be more potent determinants of symptom expression than dispositions associated with customary gender socialization in the civilian setting. As modern gender roles, however, become increasingly egalitarian in all segments of society (Brewster & Padavic, 2000), results such as these may become the generalized expectation across traumatized populations.

It is interesting to note that the one known study of PTSD gender-based DIF found a larger and mostly nonoverlapping set of symptoms to show differential responding in a civilian sample (Palm et al., 2009). Although sociocultural similarities between women's and men's experiences in the military may explain the ostensibly contrasting results, it is difficult to draw substantive inferences given the limited interpretability of the earlier investigation. Clearly, more research is warranted to clarify these differences across investigations. Ideally, future studies could take advantage of IRT-based equating methods in a combined civilian and military sample to draw reliable comparisons on the effects of gender across populations.

With respect to limitations, only recently have evolutions in military practice permitted research with men and women who performed similar duties in a similar deployment environment, so additional studies will be helpful in determining the extent to which findings generalize to veterans of eras prior to OEF/OIF. Similarly, results also pertain most directly to the current assessment context (i.e., confidential self-report) and may not hold for other contexts with stronger demand characteristics. Additional research is also needed to establish whether current results extend to PTSD assessments other than the PCL-M. Finally, it should be noted that there are a variety of accepted analytic methods for assessing DIF in psychometric instruments (Teresi, 2006), so it is possible that a different method may have identified a slightly different set of DIF items, although it is unlikely that results pertaining to impact on PCL-M measurement would substantially differ.

In conclusion, current findings have both theoretical and practical implications. Most directly, they suggest that the overt expression of stress reactions following deployment experiences (as captured by the PTSD diagnosis) may be fairly unitary between men and women veterans. Customary symptom conceptualizations appeared to be equally relevant across genders to estimations of PTS severity in this important population. From a practical standpoint, current results should assure users of the PCL-M with OEF/OIF veterans that men's and women's scores have comparable reliability, a key psychometric finding that informs its use in both research and practice. More broadly, this lends circumstantial support to the notion that sociocultural factors (perhaps especially related to gender roles) and trauma type do strongly moderate the relationship between gender and susceptibility to PTSD, suggesting a need to continue incorporating these perspectives into future gender comparison research.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Armour, C., Elhai, J. D., Layne, C. M., Shevlin, M., Durakovic-Belko, E., Djapo, N., & Pynoos, R. S. (2011). Gender differences in the factor structure of posttraumatic stress disorder symptoms in war-exposed adolescents. *Journal of Anxiety Disorders*, *25*, 604–611. doi:10.1016/j.janxdis.2011.01.010

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. doi:10.1023/A:1010933404324

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. doi:10.1017/S0033291799008612

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*, 626–631. doi:10.1001/archpsyc.55.7.626

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology, 68*, 748–766. doi:10.1037/0022-006X.68.5.748

Brewster, K. L., & Padavic, I. (2000). Change in gender-ideology, 1977–1996: The contributions of intracohort change and population turnover. *Journal of Marriage and the Family, 62*, 477–487. doi:10.1111/j.1741-3737.2000.00477.x

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*, 1–30.

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. doi:10.1017/S0033291707002589

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care, 44*(11, Suppl 3), S115–S123. doi:10.1097/01.mlr.0000245183.28384.ed

Green, B. (2003). Post-traumatic stress disorder: Symptom profiles in men and women. *Current Medical Research and Opinion, 19*, 200–204. doi:10.1185/030079903125001604

Hoge, C. W., Clark, J. C., & Castro, C. A. (2007). Commentary: Women in combat and the risk of post-traumatic stress disorder and depression. *International Journal of Epidemiology, 36*, 327–329. doi:10.1093/ije/dym013

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Kessler, R. C., & Merikangas, K. R. (2004). The National Comorbidity Survey Replication (NCS-R): Background and aims. *International Journal of Methods in Psychiatric Research, 13*, 60–68. doi:10.1002/mpr.166

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. doi:10.1001/archpsyc.1995.03950240066012

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment, 10*, 90–96. doi:10.1037/1040-3590.10.2.90

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

Maguen, S., Luxton, D. D., Skopp, N. A., & Madden, E. (2012). Gender differences in traumatic experiences and mental health in active duty soldiers redeployed from Iraq and Afghanistan. *Journal of Psychiatric Research, 46*, 311–316. doi:10.1016/j.jpsychires.2011.11.007

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD Checklist: A critical review. *Clinical Psychology Review, 30*, 976–987. doi:10.1016/j.cpr.2010.06.012

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review, 29*, 496–505. doi:10.1016/j.cpr.2009.05.003

Norris, F. H., Perilla, J. L., Ibañez, G. E., & Murphy, A. E. (2001). Sex differences in symptoms of posttraumatic stress disorder: Does culture play a role? *Journal of Traumatic Stress, 14*, 7–28. doi:10.1023/A:1007851413867

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. doi:10.1037/1040-3590.14.1.50

Palm, K. M., Strong, D. R., & MacPherson, L. (2009). Evaluating symptom expression as a function of a posttraumatic stress disorder severity. *Journal of Anxiety Disorders, 23*, 27–37. doi:10.1016/j.janxdis.2008.03.012

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. doi:10.1017/S003329170500591X

Pole, N., Best, S. R., Weiss, D. S., Metzler, T. J., Liberman, A. M., Fagan, J. A., & Marmar, C. R. (2001). Effects of gender and ethnicity on duty-related posttraumatic stress symptoms among urban police officers. *Journal of Nervous and Mental Disease, 189*, 442–448. doi:10.1097/00005053-200107000-00005

R Development Core Team. (2011). *R: A language and environment for statistical computing (Version 2.13.2)*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2012). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, IL: Northwestern University.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores (Psychometrika Monographs Whole No. 17). Richmond, VA: Psychometrica Society.

Slocum-Gori, S., & Zumbo, B. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research, 102*, 443–461. doi:10.1007/s11205-010-9682-8

Street, A. E., Gradus, J. L., Giasson, H. L., Vogt, D., & Resick, P. A. (2012). Gender differences among veterans deployed in support of the wars in Afghanistan and Iraq. *Journal of General Internal Medicine*.

Street, A. E., Vogt, D. S., & Dutra, L. (2009). A new generation of women veterans: Stressors faced by women deployed to Iraq and Afghanistan. *Clinical Psychology Review, 29*, 685–694. doi:10.1016/j.cpr.2009.08.007

Tanielian, T. L., & Jaycox, L. H. (2008). *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery*. Santa Monica, CA: RAND Corp.

Teresi, J. A. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care, 44*, S152–S170. doi:10.1097/01.mlr.0000245142.74628.ab

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*, 959–992. doi:10.1037/0033-2909.132.6.959

Vogt, D. S., Vaughn, R., Glickman, M. E., Schultz, M., Drainoni, M.-L., Elwy, A. R., & Eisen, S. V. (2011). Gender differences in combat-related stressors and their association with postdeployment mental health in a nationally representative sample of U.S. OEF/OIF veterans. *Journal of Abnormal Psychology, 120*, 797–806. doi:10.1037/a0023452

Weathers, F. W., Huska, J. A., & Keane, T. M. (1991). *The PTSD Checklist – Military Version*. Boston, MA: National Center for PTSD.

Wright, B. D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.

Zlotnick, C., Zimmerman, M., Wolfsdorf, B. A., & Mattia, J. I. (2001). Gender differences in patients with posttraumatic stress disorder in a general psychiatric practice. *American Journal of Psychiatry, 158*, 1923–1925. doi:10.1176/appi.ajp.158.11.1923

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.