Journal of Traumatic Stress, Vol. 24, No. 2, April 2011, pp. 208–212 (© 2011)

Posttraumatic Stress Disorder Instrument Wording Content Is Associated With Differences in Factor Structure

Jon D. Elhai and Tracey L. Biehn
University of Toledo

James A. Naifeh
Uniformed Services University of the Health Sciences

B. Christopher Frueh
University of Hilo

The authors examined posttraumatic stress disorder (PTSD) item wording differences on the factor structure of PTSD. Nonclinical, trauma-exposed participants were randomly assigned to complete a PTSD measure using item wording content from the PTSD Checklist (n = 182) or PTSD Symptom Scale (n = 203). Compared to the 4-factor emotional numbing PTSD model, the 4-factor dysphoria PTSD model fit best across groups based on smaller Bayesian information criterion (BIC) values. For PTSD Checklist participants, the numbing model's BIC was 6238.54 compared to the dysphoria model's BIC of 6156.03. For the PTSD Symptom Scale, the numbing model's BIC was 6161.38 compared to the dysphoria model's BIC of 6102.87. Groups differed on variable intercepts and residual variances. Instrument and construct implications are discussed.

The *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition, Text Revision (DSM-IV-TR;* American Psychiatric Association [APA], 2000) sets forth diagnostic criteria for posttraumatic stress disorder (PTSD), including 17 symptoms of reexperiencing (Criteria B1–B5), effortful avoidance/emotional numbing (C1–C7), and hyperarousal (D1–D5). However, 4-factor models by King, Leskin, King, and Weathers (1998) and Simms, Watson, and Doebbeling (2002) best represent the latent structure of PTSD instruments. Yet testing factor structure differences from different wording content in measures of PTSD has not been conducted.

King et al.'s (1998) emotional numbing model splits Criterion C of PTSD into separate avoidance (C1–C2) and numbing (C3–C7) factors, supported by empirical literature (reviewed in Asmundson, Stapleton, & Taylor, 2004). Simms et al.'s (2002) dysphoria model modifies King et al.'s (1998) model by combining three hyperarousal symptoms (D1–D3: sleep difficulty, irritability, concentration problems) with the numbing symptoms (C3–C7) to form a dysphoria factor (Table 1), consistent with theory and empirical findings (Watson, 2009).

Empirical evidence from various trauma-exposed samples supports the numbing model (most recently, Elhai, Palmieri, Biehn, Frueh, & Magruder, 2010; Mansfield, Williams, Hourani, & Babeu, 2010) and dysphoria model (most recently, Carragher, Mills, Slade, Teesson, & Silove, 2010; Elhai et al., 2009). Both models are consistently supported over competing PTSD models (reviewed in Miller, 2010; Palmieri, Weathers, Difede, & King, 2007; Simms, 2010; Yufik & Simms, 2010). Neither model has demonstrated superior fit in a clear majority of studies.

Palmieri and colleagues (2007) found the numbing model fit best using a clinician-administered PTSD instrument; the dysphoria model fit best with a self-report instrument. The current study extends this work, testing factor structure differences from differently worded self-report PTSD instruments— the PTSD Checklist (PCL) and PTSD Symptom Scale (PSS).

This study poses two research questions. First, does the difference in item wording content between PTSD self-report measures affect whether the numbing or dysphoria model fits best? Second, do wording content variations from the PCL and PSS impact other factor structure parameters? Results will improve knowledge into the underlying components or constructs of PTSD assessment instruments, with implications for essential symptom clusters for diagnosis, and targeting and monitoring symptom domains in treatment.

Jon D. Elhai, Tracey L. Biehn, Department of Psychology, University of Toledo; James A. Naifeh, Department of Psychiatry, Uniformed Services University of the Health Services; B. Chrisopher Frueh, Department of Psychology, University of Hilo.

Correspondence concerning this article should be addressed to: Jon Elhai, Department of Psychology, University of Toledo, Mail Stop #948, 2801 W. Bancroft St., Toledo, Ohio 43606-3390. Website: www.jon-elhai.com.

 DOI: 10.1002/jts.20628

Table 1. Item-Model Mappings, and Descriptive Statistics for 17 Posttraumatic Stress Disorder Items

	Item Mapping		Modified PCL (n = 182)				Modified PSS (n = 203)			
Item	Numbing model	Dysphoria model	*M*	*SD*	Skew	Kurtosis	*M*	*SD*	Skew	Kurtosis
Intrusive thoughts	R	R	0.57	0.67	1.09	1.41	0.59	0.71	1.20	1.48
Nightmares	R	R	0.37	0.64	1.92	3.87	0.28	0.52	1.72	2.10
Reliving trauma	R	R	0.29	0.58	2.42	6.89	0.37	0.58	1.45	1.88
Emotional cued reactivity	R	R	0.82	0.76	1.00	1.31	0.67	0.69	0.92	1.06
Physiological cued reactivity	R	R	0.42	0.67	1.56	1.91	0.21	0.48	2.84	10.45
Avoidance of thoughts	A	A	0.77	0.95	1.15	0.34	0.58	0.86	1.49	1.47
Avoidance of reminders	A	A	0.48	0.77	1.72	2.57	0.50	0.84	1.71	2.11
Trauma-related amnesia	N	D	0.35	0.66	2.02	3.79	0.32	0.65	2.26	5.16
Loss of interest	N	D	0.24	0.62	2.94	8.70	0.20	0.50	2.78	8.18
Feeling detached	N	D	0.51	0.85	1.51	1.19	0.34	0.64	2.02	3.96
Feeling numb	N	D	0.38	0.70	1.76	2.16	0.35	0.73	2.32	4.96
Hopelessness	N	D	0.31	0.65	2.11	3.75	0.38	0.70	2.07	4.21
Difficulty sleeping	H	D	0.69	0.93	1.24	0.45	0.51	0.83	1.52	1.34
Irritability or anger	H	D	0.52	0.75	1.30	0.89	0.36	0.66	2.03	4.17
Difficulty concentrating	H	D	0.64	0.89	1.25	0.58	0.44	0.69	1.48	1.47
Overly alert	H	H	0.74	1.00	1.15	0.06	0.66	0.97	1.32	0.56
Easily startled	H	H	0.58	0.85	1.42	1.16	0.45	0.80	1.86	2.79

Note. PCL = PTSD Checklist; PSS = PTSD Symptom Scale-Self Report; N = numbing; D = dysphoria; R = reexperiencing; A = avoidance; H = hyperarousal.

METHOD

Participants/Procedure

Undergraduates from an Ohio public university's psychology research pool volunteered between 2009–2010 using internal Web-based recruitment; the sample numbered 720. Potential participants were presented a Web-based consent statement describing the study as involving "stressful events and emotional functioning," with class credit offered, but no monetary compensation.

Instruments

We inquired about such demographics as, e.g., gender and age. Subsequently, we used the Stressful Life Events Screening Questionnaire (Goodman, Corcoran, Turner, Yuan, & Green, 1998) to measure previous self-reported psychological trauma exposure using 11 PTSD criterion A1 traumatic stressors and a 12th A1 item assessing "other" serious injury/life danger. Adequate validity and reliability have been found (Goodman et al., 1998). Trauma-exposed respondents were asked to nominate their most distressing trauma.

The PTSD Checklist-Specific Stressor Version (Weathers, Litz, Herman, Huska, & Keane, 1993) assesses responses to one's "stressful experience" for a list of past-month *DSM-IV* PTSD symptoms, using a Likert scale (1 = *not at all*, 2 = *a little bit*, 3 = *moderately*, 4 = *quite a bit*, and 5 = *extremely*). It has adequate reliability (α = .89 in this study) and validity across samples (reviewed in McDonald & Calhoun, 2010). We modified the PCL's response options to duplicate the PSS' 4-point scale. Thus we did not use a standard version of the PCL in this study.

The PTSD Symptom Scale-Self Report (Foa, Riggs, Dancu, & Rothbaum, 1993) assesses PTSD symptoms over the past 2 weeks using a Likert scale with frequency and intensity anchor points (0 = *Not at all*, 1 = O*nce per week or less/a little bit/once in a while*, 2 = *2 to 4 times per week/somewhat/half the time*, and 3 = *5 or more times per week/very much/almost always*). The PSS uses a question format, including ending stems not present in the PCL of "since the trauma" for several items, and several added item descriptors ("intense," "continuously" or "persistently"). Reliability is adequate (α = .91 in this study), as is validity (Foa et al., 1993). We modified the PSS's rating timeframe to be consistent with the PCL, making it also not comparable to standard measures in this aspect.

Using conditional branch logic that automatically routes subjects through the survey based on their particular responses, participants endorsing a Criterion A1 trauma were routed to rate PTSD symptoms from their worst trauma, randomly assigned wording content from the PCL or PSS (17-item self-report PTSD measures, mirroring *DSM-IV*'s PTSD symptom criteria). Verbatim instructions were "Please indicate how much/often you have been bothered by each problem/complaint in the past month. Note:

Please answer these questions based on your personal reactions to the one event that causes you the most distress (that you indicated above)." After survey completion, participants were presented contact information for the principal investigator and treatment resources.

Assigned Groups/Exclusions/Missing Data

Of 720 study volunteers, 21 skipped the stressor questionnaire and 284 denied previous trauma exposure, and were therefore excluded from analyses. Of 415 remaining trauma-exposed participants, 30 skipped the PTSD assessment, leaving 385 remaining subjects. Nearly half ($n = 182$) of participants (47.3% of trauma-exposed subjects) were randomly administered the modified PCL; 203 participants (52.7%) were administered the modified PSS. Forty-six subjects had nominal amounts of missing PTSD item-level data, estimated using multiple imputation generated across five imputed datasets. Confirmatory factor analysis (CFA) parameter estimates were averaged across imputed datasets with Mplus 5.2 software.

Analyses

The PTSD items were nonnormally distributed (Mardia's multivariate skewness and kurtosis tests $p < .001$). Several items had skewness or kurtosis statistics >2.0 (Table 1). Therefore, CFA included maximum likelihood estimation with Satorra and Bentler's (2001) chi-square statistic. We tested the numbing and dysphoria models separately for the modified PCL and modified PSS, with error covariances fixed to zero.

For our first research question, numbing and dysphoria models were examined separately in each group. Robust goodness of fit indices were examined: the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Models fitting very well (or adequately) are indicated by CFI and TLI $\geq .95$ (.90–.94), RMSEA $< .06$ (to .08), and SRMR $< .08$ (to .10; Hu & Bentler, 1999). We used the Bayesian information criterion (BIC) to compare the numbing and dysphoria models for both modified measures; a 10-point difference represents a 150:1 likelihood in favor of the smaller BIC value ($p < .05$; Raftery, 1995).

Next, measurement invariance testing used multigroup CFAs (separately for numbing and dysphoria models), testing invariance across groups on factor loadings, observed variable intercepts, observed variable residual variances, factor variances and covariances, and factor means, following established procedures (e.g., Meredith & Teresi, 2006). First, Model A included both groups, allowed to vary with each other on all parameters. Subsequent models tested progressively more conservative restrictions, constraining particular parameter estimates to be equal across groups, tested against the model at the prior step (except when otherwise noted in Table 2).

Table 2. Pairwise Comparisons of Modified PTSD Checklist and Modified PTSD Symptom Scale Conditions Using Measurement Invariance Procedures

	Numbing		Dysphoria	
Models compared	SB diff	*df*	SB diff	*df*
A vs. B	18.12	13	18.60	13
B vs. C	44.45***	13	48.25***	13
C vs. D	41.46***	17	46.84***	17
C vs. E	16.20	10	19.67*	10
E vs. F	7.30	4	4.50	4

Note. Table presents Satorra-Bentler chi-square difference test results for the numbing and dysphoria posttraumatic stress disorder (PTSD) models across groups. SB_{Diff} = Satorra-Bentler chi-square difference; Model A = no parameters constrained equal across groups; Model B = factor loadings constrained to be equal; Model C = observed variable intercepts and factor loadings constrained to be equal; Model D = residual variances, observed variable intercepts and factor loadings constrained to be equal; Model E = factor variances and covariances, observed variable intercepts and factor loadings constrained to be equal; Model F = factor means, factor variances and covariances, observed variable intercepts, and factor loadings constrained to be equal.
$^{*}p < .05$. $^{***}p < .001$.

We progressively constrained factor loadings (Model B), adding intercepts (Model C), residual variances (Model D), factor variances and covariances (but not residual variances, Model E), and factor means (Model F). Statistical significance between models was assessed with chi-square difference tests, using a correction factor (Satorra & Bentler, 2001).

RESULTS

Of 385 qualified participants, 70% were women ($n = 268$). Age averaged 19.96 years ($SD = 4.44$). Almost one fifth were Hispanic/Latino ($n = 69$, 17.9%). Most were Caucasian ($n = 281$, 73.0%) or African American ($n = 83$, 21.6%). The mean years of schooling was 12.90 ($SD = 1.24$).

A slight majority reported one ($n = 120$, 32.5%) or two ($n = 77$, 20.9%) traumas. The most common traumatic event endorsed was unexpected loss of family member or close friend from homicide, suicide, or accident ($n = 168$, 43.6%), which was also the most nominated worst event ($n = 163$, 42.3%).

Results from CFA demonstrate that for the modified PCL, the numbing model fit inadequately, Satorra-Bentler (SB) $\chi^2(113, N = 182) = 243.56$, $p < .001$, CFI = .84, TLI = .81, RMSEA = .08, SRMR = .07, BIC = 6238.54; the dysphoria model fit adequately, SB $\chi^2(113, N = 182) = 184.68$, $p < .001$, CFI = .91, TLI = .90, RMSEA = .06, SRMR = .06, BIC = 6156.03. For the modified PSS, the numbing model fit inadequately, SB $\chi^2(113, N = 203) = 223.48$, $p < .001$, CFI = .89, TLI = .86, RMSEA = .07, SRMR = .06, BIC = 6161.38; the

dysphoria model fit adequately, SB $\chi^2(113, N = 203) = 185.98$, $p < .001$, CFI = .91, TLI = .91, RMSEA = .06, SRMR = .06, BIC = 6102.87. The BIC values favored the dysphoria model in both groups. The BIC values should be trusted more than relative differences in other goodness of fit indices, for which comparing across models is inappropriate (Fan & Sivo, 2009).

Measurement invariance results (Table 2) demonstrated that groups differed on intercepts (with the intercepts of the modified PCL higher across both models; results not displayed); the groups differed most on flashbacks, physiological reactivity and irritability, but least on anhedonia and traumatic amnesia. The groups also differed on error variances (estimates were not consistently higher for a particular group), and (for the dysphoria model only) variances and covariances, which were consistently higher for the modified PSS. Finally, our use of a consistent response option scaling across instruments was not solely responsible for the noninvariance between groups on intercepts. Adding a constant to the modified PSS item-level data resulted in even more pronounced between-group differences.

DISCUSSION

Our results demonstrated that with nonclinical subjects the dysphoria model fit best (albeit, not very well) when using modified PCL and modified PSS item wording content, instructions, and response formats. These variations resulted in different intercepts and error variances (and variances/covariances for the dysphoria model). Interestingly, there are several obvious differences in format and content between the modified versions of the PCL and PSS, such as question format, reference to the trauma, and wording stems (described above). Any of these format differences could account for obtained empirical differences found; for example, the higher intercepts for several arousal and reactivity items on the modified PCL could be due to the requirement on the PSS that these symptoms are experienced continuously or intensely.

Limitations apply to this exploratory study. We sampled college students with a skewed representation of indirect traumatic stressors, and with a minimal level of severity of PTSD symptoms. Second, we changed the PCL's 5-point response scale of item intensity to the 4-point scale of frequency and intensity used in the PSS to standardize that feature; we do not know how this change may affect the factor structure. Also for purposes of standardization, we changed the PSS's timeframe for symptom reporting to the past month. These changes were necessary to maintain standardization across group administrations, preventing the comparison of apples to oranges. Because of these changes, however, we can only draw conclusions based on differences between two modified measures, compromising generalizability to an unknown degree.

Despite these limitations, this study's results provide additional evidence to the existing literature that the methodology used for assessment affects the factor structure of the PTSD symptoms (Elhai et al., 2009; Palmieri et al., 2007)—in particular, item wording differences. Because the majority of CFAs have used the PCL (Table 1 in Palmieri et al., 2007), knowledge about the structure of the 17 symptom criteria measured by questionnaires may be slightly skewed. The selection of a PTSD measure by researchers and clinicians should consider that the relationship among items, and the constructs they attempt to measure, may differ from instrument to instrument. Future research should further test various assessment and methodological manipulations to investigate their effect on the factor structures extracted.

REFERENCES

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

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. doi:10.1007/s10960-004-5795-7

Carragher, N., Mills, K., Slade, T., Teesson, M., & Silove, D. (2010). Factor structure of posttraumatic stress disorder symptoms in the Australian general population. *Journal of Anxiety Disorders, 24*, 520–527. doi:10.1016/j.janxdis.2010.03.009

Elhai, J. D., Engdahl, R. M., Palmieri, P. A., Naifeh, J. A., Schweinle, A., & Jacobs, G. A. (2009). Assessing posttraumatic stress disorder with or without reference to a single, worst traumatic event: Examining differences in factor structure. *Psychological Assessment, 21*, 629–634. doi:10.1037/a0016677

Elhai, J. D., Palmieri, P. A., Biehn, T. L., Frueh, B. C., & Magruder, K. M. (2010). Posttraumatic stress disorder's frequency and intensity ratings are associated with factor structure differences in military veterans. *Psychological Assessment, 22*, 723–728. doi:10.1037/a0020643

Fan, X., & Sivo, S. A. (2009). Using Δgoodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*, 54–67. doi:10.1080/10705510802561311

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6*, 459–473. doi:10.1007/BF00974317

Goodman, L., Corcoran, C., Turner, K., Yuan, N., & Green, B. L. (1998). Assessing traumatic event exposure: General issues and preliminary findings for the Stressful Life Events Screening Questionnaire. *Journal of Traumatic Stress, 11*, 521–542. doi:10.1023/A:1024456713321

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/10705519909540118

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

Mansfield, A. J., Williams, J., Hourani, L. L., & Babeu, L. A. (2010). Measurement invariance of posttraumatic stress disorder symptoms among U.S. military personnel. *Journal of Traumatic Stress, 23*, 91–99. doi:10.1002/jts.20492

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

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, 11, 69–77. doi:10.1097/01.mlr.0000245438.73837.89

Miller, M. W. (2010). On comparing competing models of PTSD: Response to Simms. *Journal of Traumatic Stress, 23*, 642–644. doi:10.1002/jts.20561

Palmieri, P. A., Weathers, F. W., Difede, J., & King, D. W. (2007). Confirmatory factor analysis of the PTSD Checklist and the Clinician-Administered PTSD Scale in disaster workers exposed to the World Trade Center Ground Zero. *Journal of Abnormal Psychology, 116*, 329–341. doi:10.1037/0021-843X.116.2.329

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163. doi:10.2307/271063

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514. doi:10.1007/BF02296192

Simms, L. J. (2010). On comparing competing structural models of PTSD: Comments on Miller et al. *Journal of Traumatic Stress, 23*, 639–641. doi:10.1002/jts.20567

Simms, L. J., Watson, D., & Doebbeling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology, 111*, 637–647. doi:10.1037/0021-843X.111.4.637

Watson, D. (2009). Differentiating the mood and anxiety disorders: A quadripartite model. *Annual Review of Clinical Psychology, 5*, 221–247. doi:10.1146/annurev.clinpsy.032408.153510

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD checklist: Reliability, validity, & diagnostic utility.* Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, Texas.

Yufik, T., & Simms, L. J. (2010). A meta-analytic investigation of the structure of posttraumatic stress disorder symptoms. *Journal of Abnormal Psychology, 119*, 764–776. doi:10.1037/a0020981