Properties of Swedish Posttraumatic Stress Measures after a Disaster

Arnberg, Filip K.[a;] Michel, Per-Olof[a]; Johannesson, Kerstin Bergh[a]

[a]National Centre for Disaster Psychiatry, Department of Neuroscience, Uppsala University, Uppsala, Sweden. Postal address: National Centre for Disaster Psychiatry, Akademiska sjukhuset ing 10, SE-751 85, Uppsala, Sweden.

Per-Olof Michel, email: po.michel@gmail.com. Kerstin Bergh Johannesson, email: kerstin.bergh.johannesson@akademiska.se.

Correspondence concerning this paper should be addressed to Filip K. Arnberg, Akademiska sjukhuset ing 10 v3, SE-751 85, Uppsala, Sweden. Phone: +46-18 611 88 98. Fax: +46-18 611 21 53. Email: filip.arnberg@neuro.uu.se

Abstract

This study evaluated the properties of Swedish versions of self-report measures of posttraumatic stress disorder (PTSD), with emphasis on the Impact of Event Scale–Revised (IES-R). Survey data from adult survivors 1, 3, and 6 years after the 2004 Indian Ocean tsunami (*n*= 1,506) included the IES-R (from which the IES-6 was derived) and the 12-item General Health Questionnaire (GHQ-12). The PTSD Checklist (PCL) was included in one survey. A structured clinical interview was performed after 6 years (*n* = 142). Factor analyses of the IES-R and PCL indicated that a dysphoric-arousal model provided good fit invariant across assessments. Both measures were accurate in excluding PTSD while all measures provided poorer positive predictive values. The IES-R, but not the IES-6 and GHQ-12, evidenced stability across assessments. In conclusion, the Swedish IES-R and PCL are sound measures of chronic PTSD, and the findings illustrate important temporal aspects of PTSD assessment.

Keywords: Posttraumatic Stress Disorder; Longitudinal invariance; Confirmatory Factor Analysis; Sensitivity and Specificity; Mental Health; Stressful Events

1. Introduction

Disaster survivors are at risk of developing chronic posttraumatic stress disorder (PTSD; American Psychiatric Association [APA], 2013), which can persist for several years and is associated with significant comorbidity and disability (Arnberg, Johannesson, & Michel, 2013; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Taft, Stern, King, & King, 1999). Screening for PTSD thus serves an important purpose after disasters and evaluations of measures in appropriate contexts are needed. At the same time, there is no solid consensus about how to conceptualize the PTSD construct in terms of its symptom clusters and there are gaps in our knowledge about its longitudinal stability (Armour, Carragher, & Elhai, 2013; Elhai & Palmieri, 2011; McHugh & Treisman, 2007).

PTSD was devised as a three-dimensional construct that includes intrusions, avoidance, and hyperarousal reactions (APA, 1980). Since then, empirical studies have consistently found that models with four or five factors provide better fit than the DSM-IV model (Armour, Carragher, et al., 2013; Yufik & Simms, 2010). These models differ from DSM, and among themselves, in how they conceptualize the DSM avoidance and hyperarousal factors. King, Leskin, King, and Weathers (1998) found support for a four-factor numbing model, in which the DSM avoidance factor is split into two factors labelled effortful avoidance and emotional numbing. Simms, Watson, and Doebbeling (2002) argued that symptoms of emotional numbing were examples of general distress. In support of their claim, they found a good fit for a dysphoria model, in which emotional numbing is combined with the general symptoms in the hyperarousal factor (i.e., symptoms D1-D3 in DSM-IV: sleep difficulty, irritability, and concentration problems) into a factor labelled dysphoria.

More recently, a five-factor model (Elhai et al., 2011) has been proposed, in which avoidance is split into effortful avoidance and numbing, and hyperarousal into dysphoric arousal and anxious arousal. Elhai et al. (2011) noted that the general hyperarousal symptoms are conceptually different from both the hyperarousal and the numbing items: First, they differ from the remaining hyperarousal symptoms (i.e., startle and hypervigilance) that characterize anxious arousal prototypical for fear-based symptomatology. Second, symptoms D1-D3 represent agitation and restlessness, which corresponds poorly with the notion of dysphoria (Elhai et al., 2011). Note that nearly all confirmatory factor analyses (CFAs) have used measures mapping directly onto the DSM-IV symptom criteria. If a model can provide good fit also when applied to other PTSD measures it would strengthen the evidence and could prevent that the refinement process ends prematurely (King et al., 2009). Corroborations also across disaster types would strengthen the evidence, and of particular relevance to the present study, the dysphoric-arousal model seems to outperform the four-factor models in disaster contexts such as tsunamis (Armour, Carragher, et al., 2013) and earthquakes (L. Wang, Zhang, Shi, Zhou, Li, et al., 2011).

A factor model should also provide similar fit across several assessments (i.e., longitudinal invariance), as it is essential, perhaps particularly for PTSD that has a clear temporal aspect, to establish that we assess the same construct regardless of timing (Elhai & Palmieri, 2011). The few studies that exist suggest that at least configural invariance seems plausible, but more research is needed (King et al., 2009; Krause, Kaltman, Goodman, & Dutton, 2007; M. Wang, Elhai, Dai, & Yao, 2012).

With regard to screening, widely used self-report measures such as the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993) and the Impact of Event Scale–Revised (IES-R; Weiss, 2004) have been evaluated in clinical and military settings whereas less is known of their properties in the context of disasters (for reviews, see Brewin, 2005; McDonald & Calhoun, 2010; Wilkins, Lang, & Norman, 2011). In addition, several studies have used other self-report measures as the reference standard (e.g., Creamer, Bell, & Failla, 2003; Olde, Kleber, van der Hart, & Pop, 2006), which seem to inflate the screening accuracy as compared to studies that use a structured clinical interview as reference (Beck et al., 2008).

Longitudinal stability is relevant also for diagnostic accuracy when screening for PTSD because of the long-standing notion of variability within and across symptom clusters as time passes from the event (Horowitz, Wilner, & Alvarez, 1979; Kessler et al., 1995). Longitudinal stability becomes particularly relevant to brief measures: A screening measure should be easy to administer, complete, and score (Brewin et al., 2002). Simple, brief measures would therefore be preferred over complex screening methods. Brevity is particularly important in contexts involving large groups or rapid assessments, whereas work with clinical populations may require comprehensive, detailed measures. Brief measures seem to perform as well as longer measures in cross-sectional comparisons (Brewin, 2005; Thoresen et al., 2010). However, abbreviated measures usually do not include all symptom clusters (Brewin et al., 2002; Ouimette, Wade, Prins, & Schohn, 2008), with the six-item abbreviation of the IES-R (IES-6) being an exception (Thoresen et al., 2010). The narrow coverage of symptoms in brief measures entails a greater risk of failing to detect symptoms prominent at the time of screening. To our knowledge, however, their stability has not been assessed. (Brewin et al., 2002; Thoresen et al., 2010)

In addition to brief measures, there are potential benefits of screening for PTSD with measures of general distress. Fewer screening tools could lower costs and, importantly, yield higher response rates to screening programs. A general measure could potentially screen for a range of anxiety and mood disorders whilst having similar diagnostic accuracy to detect PTSD, as would PTSD-specific measures. The 12-item General Health Questionnaire (GHQ-12; Goldberg, 1972) performed slightly worse than a 4-item PTSD measure when given to veteran soldiers in primary care (Ouimette et al., 2008), although the differences between the measures were small and warrant further investigation.

In summary, the issue of longitudinal stability is unclear both in terms of symptom clusters and with regard to screening. In a national perspective, several widely used measures have not been evaluated. The present study aims to shed light on the properties of the Swedish versions of self-report measures for posttraumatic stress in a disaster context. We describe convergent and construct validity, using CFA to evaluate different models with the PCL and IES-R. Longitudinal invariance was investigated with the best fitting model for the IES-R. The diagnostic accuracy were evaluated for these measures together with the IES-6 and a measure of general distress, the GHQ-12, and we present mean scores for survivors and a minimally exposed comparison group.

2. Materials and Methods

2.1 Procedure and participants

The present study is part of a national cohort study of Swedish survivors from the 2004 Southeast Asia tsunamis that followed from a massive earthquake in the Indian Ocean. Swedish authorities at the national airports registered all repatriated Swedish citizens from destinations in Southeast Asia, regardless of actual disaster exposure or not, during the first 3 weeks after the disaster. The study was approved by the Regional Ethical Review Board in Uppsala, Sweden.

2.1.1 Mail surveys

The assessments of the national cohort include a mail survey 1 year and 2 months after the disaster (T1, n = 4,932; Johannesson et al., 2009) and two mail surveys of the respondents from the first survey: at 3 years and 1 month (T2, n = 3,457; Johannesson, Lundin, Fröjd, Hultman, & Michel, 2011) and 6 years and 3 months after the event (T3, n = 2,643; Johannesson, Arnberg, & Michel, 2012). Disaster exposure was established at T1 based on previous analyses of this sample (Johannesson et al., 2009) with 30 multiple-choice items. The participants were included in the exposed group if they had been caught by the waves or been severely injured; experienced life threat, death or life threat to close ones; or witnessed horrifying events (i.e., deceased people, survivors with severe injuries, people searching for others among corpses). Participants who had not experienced any of the above were not included in the analyses but served as a non-exposed comparison group.

In order to facilitate direct comparisons across measures and assessments we excluded participants with two or more missing values at any measure at any assessment ($n = 941$). However, it was common to all measures and assessments that participants had one missing value ($n = 356$, 4-8% of participants across measures). In order to prevent unnecessary loss of data participants with one missing item at any assessment were included and missing values were imputed.

The final survey sample included 1,506 survivors and 541 comparisons. The survivors were 61% women, had a mean age of 48.8 years ($SD = 13.6$), 79% were mainly married or cohabiting, 48% had a university education, and 93% were either employed, students, or had retired from employment. The comparison group was similar to the survivors with respect to demographics (all ps > .05), included 56% women with a mean age of 50 years ($SD = 14.8$), 76% were cohabiting, and 94% were employed/students/retired. A majority (65%) of survivors indicated having been caught in or chased by the tsunamis, 28% were injured, 9% received hospital care, and 15% had been bereaved. In the immediate aftermath, 33% were unsure of the fate of their significant others, 39% were missing clothes/other belongings, and 62% witnessed several deceased bodies and injured people.

2.1.2 Clinical interviews

The participants at T2 were asked for consent to participate in a telephone interview and 62% agreed to participate. The interviews were conducted at 6 years post-disaster with a random sample of 200 individuals from the exposed group. An examination of the rates of psychopathology has been reported previously (Arnberg et al., 2013). Those who consented to an interview and those who did not were highly similar in age (both groups M = 45 years; t = 0.79, p = .4, partial $\eta^2 < 0.001$), gender (43% vs. 41% women, $\chi^2 = 3.1$, p = .22, φ = .03), and employment status (both 94% employed, $\chi^2 = 0.47$, p = .50, φ = .01). However, participants who declined to be interviewed had somewhat higher scores on the IES-R at T2 (M_{diff} = 3.65, t = 7.36, p < .001, partial $\eta^2 = 0.01$). Of the 200 participants approached for interviews, 27 declined and 31 could not be reached, yielding 142 completed interviews. Those completing the interviews and those who did not were similar with respect to gender (42% vs. 43% women, $\chi^2 = 0.4$, p = .88, φ = .01), work status (92% vs. 95% employed, $\chi^2 = 1.9$, p = .21, φ = .10), and posttraumatic stress as reported at T1 (M_{diff} = 1.3 t = 0.52, p = .61, partial $\eta^2 = 0.001$) or T2 (M_{diff} = 1.9, t = 0.82, p = .42, partial $\eta^2 = 0.003$), whereas completers were somewhat older (M_{diff} = 6.2 years, t = 2.89, p = .004, partial $\eta^2 = 0.04$).

2.2 Measures

The IES-R (Weiss, 2004) is a widely used measure for posttraumatic stress pertaining to a specific event. The IES originally included 15 items to assess intrusions and avoidance (Horowitz et al., 1979) and was revised by Weiss and Marmar (1997) to include hyperarousal items. The respondent is asked to rate how distressing the reactions pertaining to a specific event (here, the tsunami) have been in the past 7 days, *from not at all* (0) to *extremely* (4), yielding a total score between 0–88. The Swedish IES-R has been used since 1995 (Eriksson & Lundin, 1996). Evaluations of the IES-R version with frequency ratings have shown good reliability and validity (Sveen et al., 2010). The Swedish version with intensity rating has not been evaluated, although we believe it to be most widely circulated version today. The Cronbach's alphas are reported in Table 1.

Table 1. *Internal Consistency and Correlations Across Measures and Assessments in 1,506 Disaster Survivors*

		IES-R			PCL	IES-6			GHQ-12		
Measure	**Time**	**T1**	**T2**	**T3**	**T2**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**
IES-R	T1	**.946**	.749	.675	.718	.956	.705	.639	.602	.422	.353
	T2		**.945**	.739	.873	.725	.958	.704	.498	.530	.394
	T3			**.943**	.705	.655	.700	.954	.467	.429	.508
PCL	T2				**.940**	.699	.835	.682	.572	.617	.431
IES-6	T1					**.842**	.702	.638	.580	.414	.341
	T2						**.848**	.681	.479	.503	.375
	T3							**.827**	.457	.421	.476
GHQ-12	T1								**.936**	.515	.415
	T2									**.905**	.459
	T3										**.899**

Note. The Cronbach's alpha is displayed in boldface in the diagonal. All correlations were significant at $p < .001$. T1 = one year after disaster. T2 = three years. T3 = six years after. GHQ-12 = 12-item General Health Questionnaire. IES-6 = 6-item Impact of Event Scale. IES-R = Impact of Event Scale–Revised. PCL = PTSD Checklist.

The IES-6, an abridged version of the IES-R (Thoresen et al., 2010), represents an attempt to devise a brief measure for use in epidemiologic surveys and as a screening tool. The IES-6 includes two IES-R intrusion items, two avoidance items, and two hyperarousal items. The validation study indicated that the IES-6 accounted for 91% of the variance in the full IES-R and possessed screening properties similar to the IES-R when compared to the PTSD Checklist (Thoresen et al., 2010). The Cronbach's alphas are reported in Table 1.

The PCL (Weathers et al., 1993) includes 17 items that map directly onto the DSM-IV PTSD symptoms and assesses reactions that pertain to a specific event (i.e., the PCL-Specific version was used here), specified as the tsunami. The respondent indicates how distressing the symptoms have been for the past month on a scale from *not at all* (1) to *extremely* (5), yielding a sum score of 17–85. Another screening method is based on the symptom cluster method where a dichotomous total score is achieved by assigning a score of 0 or 1 depending on if the respondent endorses (has a score of 3–5) at least one intrusion item, three avoidance items, and two hyperarousal items. The translation of the PCL was done in 2004 by a clinical psychologist, apparently by adapting the Swedish DSM-IV symptoms according to the original PCL, and is widely circulated today. The Swedish PCL has not been evaluated previously. The Cronbach's alpha is reported in Table 1.

The GHQ-12 (Goldberg, 1972) is widely used in primary care settings to screen for current anxiety and mood disorders and is reliable and valid in community samples in different cultural contexts (Furukawa & Goldberg, 1999; Goldberg et al., 1997), including Sweden (Sconfienza, 1998). The GHQ-12 comprises six items that focuses on inability to undertake normal functions and six items on the appearance of new and distressing phenomena, all subsumed under a general factor (Shevlin & Adamson, 2005). The respondents indicate whether they have experienced changes in these domains during the past few weeks. A dichotomous scoring method is recommended for screening purposes: The responses are scored as 0 for *better than,* or *as usual,* and 1 for *less* or *much less than usual*, providing a range of 0-12. The presence of three or more items (i.e., a GHQ-score of > 3) is commonly used as the cut-off for poor general mental health (Goldberg et al., 1997). The Cronbach's alphas are reported in Table 1.

The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997/1998) was used to assess PTSD. The interviewers were 10 M.Sc. students in psychology with at least 3 years of full-time studies. They received 8 hours of additional training. The principal author, blind to the results of the interviews (although not to the inclusion criterion regarding exposure), audited the PTSD module in a random sample of the interviews (n = 20). A Cohen's kappa of 1.0 indicated excellent inter-rater reliability. The time of onset and eventual remission of PTSD were ascertained during the interview by asking for (a) when the symptoms started/ceased and (b) when the symptoms started/ceased to cause functional impairment. Thus, the reference standard for T1 and T2 was retrospective in nature. The self-report data were not merged with diagnostic data until the data collection and reliability checks were completed.

2.3 Statistical analysis

Before the main analyses, the missing items were imputed by using the SPSS expectation-maximization algorithm with all other items as predictors. The IES-R and the PCL at T1 were subjected to CFAs with the original, DSM-IV model. Then numbing, dysphoria, and dysphoric arousal models were fitted. The IES-R was subjected to a longitudinal CFA to assess measurement invariance across all assessments. Configural invariance (i.e., equal form); weak, or metric, invariance (i.e., equal form and factor loadings); and strong/scalar invariance (i.e., weak invariance and equal factor intercepts) were evaluated (e.g., Meredith & Teresi, 2006). Tests of invariance evaluate whether the item-to-factor relationships are consistent across groups or assessments. Weak invariance concerns the strength of the factor loadings and postulate that, for all items, one unit change in the item score is scaled to a change in the factor score hat is equal across assessments. In this case it would speak to whether the factors of chronic PTSD mean the same thing (based on the observed items) both 1, 3, and 6 years after the disaster. Strong invariance postulates that also the intercepts are equal, because both intercept and slope (i.e., factor loading) need to be invariant in order to establish that the chronic PTSD factor scores are linked to the item scores in a consistent way over time. The mean-adjusted maximum likelihood estimator was used.

Model test statistics of fit included scaled χ^2-tests, the standardized root mean square residual (SRMR), root-mean-square error of approximation (RMSEA), Tucker Lewis Index (TLI), and Bentler comparative fit index (CFI). The χ^2-test is an absolute measure of fit but is overly conservative with larger samples. RMSEA values < 0.05 indicate good fit and values ranging between 0.08 and 0.10 moderate fit, while CFI values close to 0.95 indicate good fit and values > 0.90 acceptable fit (Byrne, 2009). Measurement invariance was investigated with the Satorra–Bentler scaled $\Delta\chi2$-test (Satorra & Bentler, 2001) and ΔCFI, where $\Delta CFI \leq -.01$ is regarded as a nonsignificant change (Cheung & Rensvold, 2002).

The screening properties of the measures were evaluated by using several indicators: sensitivity and specificity, positive predictive value (PPV) and negative predictive value (NPV), the positive likelihood ratio (LR+), and the discriminant ability (DA). Wilson score 95% confidence intervals (CIs) are reported except for LR+, for which profile likelihood CIs are reported. The DA represents the average of the sensitivity and specificity for a given threshold. The threshold scores were determined as the scores that yielded the optimal DA at T1 (Riegelman, 2005) with precedence to sensitivity if the optimal DA was found at more than one threshold. The optimal DA for T2 and T3 was also calculated in order to evaluate the reliability of the optimal threshold. Analyses were performed with SPSS v.20 for Mac (IBM, Chicago, IL), except for the CFAs that were performed in R v3.0 with the packages *lavaan* v0.5-15 (Rosseel, 2012) and *lavaan.survey* v1.0 to adjust for clustered data (see Bollen, Tueller, & Oberski, 2013).

3. Results

3.1 Survey Data

The analysis of construct validity included 1,506 directly and indirectly disaster-afflicted participants. All measures showed high internal consistency and high correlations among them (Table 1). In particular, the IES-6 scores were very highly associated with the scores of the full IES-R (rs > .94).

Table 2 displays the results of the CFA. For the IES-R, the original model provided poor fit. The residuals and modification indices strongly suggested that item 2 (“trouble sleeping”) was a poor indicator for intrusion. The two top modification indices concerned this item and suggested moving it to the hyperarousal factor and freeing one error covariance to

Table 2. *Fit Indices for the Impact of Event Scale–Revised (IES-R) and the PTSD Checklist (PCL) in 1,506 Disaster Survivors.*

Model	χ^2 ***(df)***	**CFI**	**TLI**	**BIC**	**RMSEA (90% CI)**	**SRMR**
IES-R (T1)						
DSM-IV	2144.3 (206)	.85	.83	87,027	0.079 (0.077–0.082)	0.064
DSM-IV[a, b]	1609.8 (205)	.89	.88	86,249	0.067 (0.065–0.070)	0.061
Dysphoria[b]	1719.8 (202)	.88	.87	86,442	0.071 (0.068–0.073)	0.063
Numbing[b]	1607.2 (202)	.89	.88	86,269	0.068 (0.065–0.071)	0.059
Dysphoric Arousal[b]	1463.3 (198)	.90	.89	86,088	0.065 (0.063–0.068)	0.057
Configural invariance	3599.2 (594)	.89	.88	235,342	0.058 (0.057–0.059)	0.055
Weak invariance	3819.6 (638)	.89	.88	235,320	0.058 (0.056–0.059)	0.060
Partial strong invariance[c]	4102.8 (670)	.88	.88	235,367	0.058 (0.057–0.060	0.062
Strong invariance	4244.9 (672)	.87	.87	235,550	0.059 (0.058–0.061)	0.062
PCL (T2)						
DSM-IV	708.3 (116)	.91	.90	55,548	0.058 (0.055–0.061)	0.051
Dysphoria	562.8 (113)	.94	.92	55,253	0.051 (0.048–0.054)	0.040
Numbing	503.2 (113)	.94	.93	55,142	0.048 (0.045–0.051)	0.042
Dysphoric Arousal	398.6 (109)	.96	.95	54,954	0.042 (0.039–0.045)	0.038

Note. The IES-R was administered approximately 1, 3, and 6 years post-disaster (T1, T2, and T3) and the PCL 6 years post-disaster (T2). The CFI = Comparative Fit Index. RMSEA = Root-Mean-Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. TLI = Tucker Lewis Index.
[a] Item 2 was moved from the intrusion to hyperarousal factor
[b] The error covariance between item 2 and 15 (staying and falling asleep) was set free.
[c] The intercept of item 3 was set free across assessments.

item 15 ("trouble falling asleep"). The indices were highly similar at three and six years. In fact, freeing the covariance between items 2 and 15 improved the fit for all models. The items are essentially two items based on the same symptom, and with this straightforward interpretation of this modification it was used in the analysis of IES-R. The dysphoric-arousal model provided best fit according to fit indices and a test against the four-factor models, scaled $\Delta\chi^2(4) = 72.3$ for the numbing and 81.6 for the dysphoria model; both *p*s < .001. For the PCL, the models generally provided acceptable or good fit. The original DSM-IV model provided worst fit and the dysphoric-arousal model again provided better fit compared to the four-factor models, scaled $\Delta\chi^2(4) = 42.4$ for numbing and 43.8 for dysphoria; both *p*s < .001.

The dysphoric-arousal model was used in the analysis of measurement invariance with the IES-R across T1–T3 (Table 2). Configural invariance was clearly indicated, and Table 3 lists standardized estimates for the dysphoric-arousal model across assessments.

Table 3. *Standardized Estimates for the Dysphoric-Arousal Model Fitted to the Impact of Event Scale–Revised across Three Assessments after a Disaster*

Factors	**Item**	**Abbreviated description**	**1 year**	**3 years**	**6 years**
				Factor loadings	
Intrusion	1	Reminders evoke feelings	.67	.67	.73
	3	Made to think about it	.78	.75	.73
	6	Did not mean to think about it	.81	.78	.76
	9	Pictures popped into my mind	.79	.79	.77
	14	Acting/feeling as at that time	.63	.67	.67
	16	Waves of feelings	.78	.81	.78
	19	Physical reactions	.70	.70	.69
	20	Dreams	.71	.67	.63
Avoidance	5	Avoid upset	.73	.72	.75
	8	Avoid reminders	.80	.81	.80
	11	Tried not think	.84	.83	.82
	12	Not deal with feelings	.66	.70	.67
	17	Remove from memory	.72	.74	.71
	22	Tried not talk	.69	.72	.68
Numbing	7	Hadn't happened, unreal	.52	.46	.46
	13	Numb feelings	.76	.78	.82
Dysphoric arousal	2	Trouble sleeping[a]	.76	.76	.75
	4	Irritable, angry	.77	.75	.73
	15	Trouble falling asleep[a]	.73	.72	.73
	18	Trouble concentrate	.83	.83	.84
Anxious arousal	10	Jumpy, startled	.83	.78	.76
	21	Watchful, on guard	.79	.79	.73
				Correlations	
Correlated errors	2–15		.51	.52	.49
Factor correlations	Intrusion–Avoidance		.70	.69	.72
	Intrusion–Numbing		.79	.76	.81
	Intrusion–Dysphoric arousal		.80	.81	.80
	Avoidance–Numbing		.84	.84	.85
	Avoidance–Dysphoric arousal		.83	.82	.81
	Avoidance–Anxious arousal		.63	.63	.67
	Numbing–Dysphoric arousal		.66	.67	.72
	Numbing–Anxious arousal		.76	.75	.75
	Dysphoric arousal–Anxious arousal		.78	.75	.80

Evidence for weak invariance across assessments was mixed: the absolute measure of fit indicated poor fit, scaled $\Delta\chi^2(44) = 218.5$, $p < .001$, whereas the relative fit measure indicated weak invariance, $\Delta CFI = -.006$. Strong invariance did not hold, scaled $\Delta\chi^2(24) = 633.6$, $p < .001$; $\Delta CFI = -.01$. Eight of the top ten modification indices involved intrusion items (mainly items 3, 6, and 9). Relaxing any of these items' intercepts yielded similar results, e.g., item 3 (other things kept making me think about it) provided strong partial invariance, scaled $\Delta\chi^2(32) = 360.9$, $p < .001$; $\Delta CFI = -.008$.

3.2 Interview Data: Screening Properties

The number of participants included in the analysis of screening properties was 140 for T1, 141 for T2, and 115 for T3. According to the SCID, there were 12 cases of PTSD at T1, 7 at T2, and 6 at T3. PTSD cases at T2 and T3 were cases also at the previous assessments. Although participants noted other potentially traumatic events when the interviewer probed for events that satisfied the A1 criterion, the disaster was the index event for all cases of PTSD. The lifetime prevalence was 21% for depression, 3% for alcohol abuse and 7% for specific phobia. PTSD was related to higher scores on all self-report measures, and also when compared to the minimally exposed comparison group (Table 4).

Table 4. *Descriptive Statistics Between Disaster Exposed PTSD-Cases and Non-Cases, and a Reference Sample on the IES-R, GHQ-12 and PCL*

		Total			**PTSD**			**No PTSD**			**Ref.**[a]	
Assessment		***N***	***M***	***SD***	***M***	***SD***	***Min***	***M***	***SD***	***Max***	***M***	***SD***
IES-R	T1	140	23.0	18.4	49.6	16.0	24	20.5	16.5	82	10.0	11.2
	T2	141	17.0	15.6	45.0	9.9	32	15.5	14.4	79	5.7	8.6
	T3	115	14.0	13.3	38.3	14.9	13	12.7	11.9	64	6.4	8.8
IES-6	T1	140	7.2	5.7	15.7	4.7	7	6.5	5.1	24	2.9	3.6
	T2	141	5.0	4.7	13.0	3.8	7	4.6	4.4	22	1.5	2.6
	T3	115	4.5	4.4	11.2	4.8	4	4.2	4.2	19	2.0	2.9
GHQ-12	T1	140	2.4	3.8	7.9	3.6	0	1.9	3.4	12	0.8	2.2
	T2	141	1.6	2.9	6.9	3.9	1	1.3	2.5	12	1.2	2.4
	T3	115	1.6	2.6	3.0	3.1	0	1.5	2.6	10	1.0	2.3
PCL	T2	141	28.2	12.6	56.3	10.3	42	26.7	10.9	17	19.7	6.3

Note. There were 12 PTSD cases at T1 (1 year), 7 cases at T2 (3 years), and 6 cases at T3 (6 years). PCL was only administered at T2. According to *t*-tests the PTSD group had higher scores than the other groups ($p < .001$) except for one comparison with the non-PTSD group scores on GHQ-12 at T3. GHQ-12 = 12-item General Health Questionnaire. IES-6 = 6-item Impact of Event Scale. IES-R = Impact of Event Scale–Revised. PCL = PTSD Checklist. [a]Comparison group ($n = 541$) with minimal exposure to the event.

For the IES-R the same cut-off score (IES-R > 30) provided optimal DA across all assessments (Table 5). The IES-R sensitivity was good or excellent (≥ .80) at all assessments whereas the specificity was low at T1. We explored if better PPV performance could be achieved. Increasing the cutoff score to IES-R > 40 yielded modest increases in the PPV (T1 = 0.35, T2 = 0.42, and T3 = 0.38) while the sensitivity was reduced to 0.75, 0.71, and 0.69 for T1, T2, and T3. For the PCL, a cut-off score of > 40 provided highest DA. The two coding schemes for the PCL yielded nearly identical screening properties, similar also to those of the IES-R at T2.

The optimal threshold for the IES-6 was > 14 at T1; however, the sensitivity was quite low at T1 and poor at T2. Compared with T1, the optimal DA for the IES-6 was found at different thresholds for T2 and T3 (> 10 and > 9, respectively). The properties of the GHQ-12 were similar to those of the IES-R and PCL although the sensitivity was somewhat lower. The optimal DA for GHQ-12 was found at a cut-off score of > 5 for both T1 and T2. The GHQ-12 performed poorly at T3, and a post-hoc analysis indicated that the sensitivity would increase only to 0.40 by using the optimal cut-off score for T3 of > 3, which provided a DA of .61.

Table 5. *PTSD Screening Properties for Measures of Posttraumatic Stress and General Health in 142Disaster Survivors Across Three Assessments*

Time	Measure	Sensitivity	Specificity	PPV	NPV	LR+	DA
T1: 14 mo.	IES-R	0.92 [0.65, 0.99]	0.75 [0.67, 0.81]	0.25 [0.15, 0.39]	0.99 [0.94, 1.00]	3.61 [2.57, 5.08]	0.83
	IES-6	0.75 [0.47, 0.91]	0.91 [0.85, 0.95]	0.45 [0.26, 0.66]	0.98 [0.93, 0.99]	8.73 [4.54, 16.76]	0.83
	GHQ-12	0.83 [0.55, 0.95]	0.86 [0.79, 0.91]	0.36 [0.21, 0.54]	0.98 [0.94, 1.00]	5.88 [3.58, 9.67]	0.85
T2: 3 yrs.	IES-R	1.00 [0.65, 1.00]	0.85 [0.78, 0.90]	0.26 [0.13, 0.45]	1.00 [0.97, 1.00]	6.70 [4.47, 1.04]	0.93
	IES-6	0.43 [0.16, 0.75]	0.98 [0.94, 0.99]	0.50 [0.19, 0.81]	0.97 [0.93, 0.99]	19.14 [4.68, 78.28]	0.70
	GHQ-12	0.71 [0.36, 0.92]	0.91 [0.85, 0.95]	0.29 [0.13, 0.53]	0.98 [0.94, 1.00]	8.04 [3.93, 16.43]	0.81
	PCL	1.00 [0.65, 1.00]	0.87 [0.81, 0.92]	0.29 [0.15, 0.49]	1.00 [0.97, 1.00]	7.88 [5.06, 12.29]	0.94
	PCL DSM[a]	1.00 [0.65, 1.00]	0.89 [0.82, 0.93]	0.32 [0.16, 0.53]	1.00 [0.97, 1.00]	8.93 [5.55, 14.39]	0.94
T3: 6 yrs.	IES-R	0.80 [0.38, 0.96]	0.93 [0.86, 0.96]	0.33 [0.14, 0.61]	0.99 [0.95, 1.00]	10.90 [4.91, 24.21]	0.86
	IES-6	0.80 [0.38, 0.96]	0.68 [0.59, 0.76]	0.10 [0.04, 0.24]	0.99 [0.93, 1.00]	2.51 [1.50, 4.22]	0.59
	GHQ-12	0.20 [0.04, 0.62]	0.89 [0.83, 0.93]	0.08 [0.01, 0.33]	0.96 [0.90, 0.98]	1.82 [0.29, 11.35]	0.54

Note. The cutoff was set according to optimal DA at T1. PCL was administered only at T2. 95% confidence intervals are shown in brackets. DA = Discriminant ability. GHQ-12 = 12-item General Health Questionnaire (cutoff > 5). IES-6 = 6-item version of the Impact of Event Scale (cutoff > 14). IES-R = Impact of Event Scale–Revised (cutoff > 30). LR+ = Positive likelihoood ratio. NPV = Negative predictive value. PCL = PTSD Checklist (cutoff > 40). PPV = Positive predictive value.

[a]Symptom cluster scoring method.

4. Discussion

The present study investigated the psychometric properties of PTSD measures in a Swedish disaster context with an emphasis on the IES-R. In summary, the findings provide support for the use of the IES-R and PCL as sound measures of chronic PTSD. In addition, in the context of a low prevalence of PTSD and the uncertainty that arises with few PTSD cases, the findings provide preliminary evidence that the IES-R proved reliable across assessments to detect chronic forms of PTSD. The IES-6 and the GHQ-12 were more inconsistent over time in both thresholds and screening properties.

This study provides evidence that chronic posttraumatic stress, as measured by the IES-R and outlined by the dysphoric-arousal model, show configural invariance, clear indications of weak/metric invariance, and nearly strong/scalar invariance. Available studies of longitudinal invariance have found support for strong invariance when assessing chronic PTSD, while the overall pattern suggests weaker forms of invariance when assessments were less distant from the event (King et al., 2009; Krause et al., 2007; M. Wang et al., 2012). The discrepancy may indicate a greater stability of the PTSD construct when used to assess chronic rather than acute and semi-acute symptoms. The construct of intrusion was a likely source for the worse fit when fixing the intercepts, suggesting that the point at which intrusive thoughts are perceived as distressing changes over time in the long term, although replications of these data clearly are needed.

In the cross-sectional CFAs, the original DSM model of the IES-R provided poor fit. In their revision of the IES, Weiss and Marmar (1997) split the sleep item into two items, whereas the findings herein clearly supported previous findings that items 2 and 15 (trouble with “falling asleep” and “staying asleep”) are highly related items. Previous researchers have interpreted the poor fit of these items as indicating a separate sleep-disturbance factor (Morina et al., 2010; L. Wang, Zhang, Shi, Zhou, Huang, et al., 2011). However, we chose to free an error covariance between these items, as it was a straightforward and the least intrusive modification. It is reasonable to believe that very similar measures such as the IES-R and should tap similar underlying constructs. Nonetheless, the models tested herein were developed with the DSM-IV symptoms in mind and so they may fit somewhat worse when applied to the IES-R as compared to the PCL. As the intention of Weiss and Marmar was to make the IES-R conform to the DSM, these data may suggest that this particular revision is not culturally invariant. However, the dysphoric-arousal model was able to provide acceptable fit for the IES-R data, which supports the validity of this Swedish version and also strengthens the empirical base for this model. In addition, there now seems to be several studies pointing to the dysphoric-arousal model as superior in the context of natural disasters (Armour, Raudzah Ghazali, & Elklit, 2013; L. Wang, Zhang, Shi, Zhou, Li, et al., 2011; M. Wang et al., 2012)

The optimal cutoff for the IES-R was the same for all three assessments, providing evidence for the stability of the measure. Overall, the efficiency of the IES-R was good and compares favorably to previous findings (Beck et al., 2008; Brewin, 2005). The sensitivity was good for all assessments, which suggests that the IES-R may work well to determine non-cases. However, the low PPVs underscore the risk of error when interpreting positive screening results as equivalent to a PTSD diagnosis in samples with a low prevalence of PTSD. The post-hoc analysis with a higher cutoff score did not improve PPV to any greater extent, likely due to overlap in IES-R scores also among the cases and noncases with the highest scores.

In the original study of the IES-6 the optimal threshold according to the DA was found to be at a total score of 8, yielding a DA of 0.88 when compared to the PCL (Thoresen et al., 2010). In the present study, however, the optimal DA was found at different thresholds for T1, T2, and T3 (i.e., at 14, 10, and 9, respectively) and the diagnostic accuracy varied across the assessments. The variability seen in the IES-6 but not the IES-R underscores the risk of a variable performance of brief versions as compared to lengthier variants. Further refinements may be needed to improve reliability of the IES-6 in this setting.

Interestingly, the GHQ-12 performed well as a screening tool at T1 but its accuracy decreased from T1 to T3, indicating that the GHQ-12 might not detect long-lasting PTSD. Further evaluations are clearly needed. In particular, the differences between PTSD-cases and noncases with regard to comorbidity should be taken into account in future studies: The screening properties of the GHQ-12 may rapidly change to the worse with a higher prevalence of other disorders than PTSD.

The PCL evidenced a discriminant validity that was highly similar to the IES-R and seems to provide good operating characteristics for detecting chronic PTSD in this setting. Moreover, the two available scoring methods for the PCL provided nearly identical results, which concurs with the conclusions from a review of the PCL (McDonald & Calhoun, 2010). Previous studies have found high specificity but variable sensitivity (ranging from 0.60 to 0.94; Andrykowski, Cordova, Studts, & Miller, 1998; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Bliese et al., 2008). None of these studies used the same cutoff score, however, and so the generalizability of the screening properties of the PCL is as of yet not fully explored. Note also that the interchangeability of the three versions of the PCL has not been evaluated (McDonald & Calhoun, 2010). The findings herein may thus not translate to the civilian and military versions.

There are limitations to the present study that must be recognized. With regard to the interview data, there are two main limitations. First, lifetime prevalence was assessed retrospectively and thus recall bias may have affected the diagnostic status at T1 and T2 (Moffitt et al., 2010). Second, the low number of PTSD cases may bias the data on stability and hamper the generalizability of the results. The data on screening performance should therefore be interpreted with caution. These concerns are compounded by the survey nonresponse of interviewees.

As usual with large-scale disasters, the survivors have heterogeneous exposure in content and severity. It is therefore unclear what exact event characteristics induced posttraumatic stress, leading to concern particularly in the small interview sample. Furthermore, the versions of IES-R and PCL used here are widely used in Sweden, and so we opted to evaluate these current versions while we are cognizant of the fact that the gold standard procedure with back-translation by independent professional translators was not used.

Finally, the use of telephone interviews may have compromised the validity or reliability of the diagnoses. However, phone interviews seem to be a reliable method to assess PTSD (Aziz & Kenford, 2004; Rohde, Lewinsohn, & Seeley, 1997). Moreover, the inter-rater reliability was excellent although perhaps somewhat inflated, as both the interviewers and the auditor knew the exposure criterion.

Looking into the strengths of this study, the longitudinal design provided an opportunity to explore the stability of operating characteristics over time. The random sampling of individuals with established disaster exposure increases the generalizability of the findings, and the sample size for the CFA was more than sufficient to detect differences among models. Also worth noting is that the sample size of screening sample is actually larger than for the majority of diagnostic accuracy studies for PTSD (Brewin, 2005; McDonald & Calhoun, 2010). The use of a standardized interview as the reference standard performed by raters who were blind to the screening results reduced the risk for potential bias.

4.1 Conclusions

The current findings support the IES-R and the PCL as sound measures in the assessment of chronic PTSD. The findings for the IES-R are clearly encouraging with respect to the temporal stability of both the construct of chronic PTSD as outlined by the dysphoric-arousal model and the screening properties of the IES-R. The data on screening performance should be interpreted with caution, yet there is tentative evidence that these measures provide acceptable screening ability albeit with many false positives in this context. The performance of the IES-6 and GHQ-12 is encouraging, although at the same time pointing to areas of concern with regard to brief or general screening measures. These results notwithstanding, the development in this field with respect to diagnostic criteria, and advancements in the understanding of the dimensionality, will require further updates of the validity of PTSD measures.

Acknowledgements

This work was funded by the Swedish Board of Health and Welfare and by the Uppsala University Hospital. The funding agencies had no involvement in the conduct of this study.

References

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Andrykowski, M. A., Cordova, M. J., Studts, J. L., & Miller, T. W. (1998). Posttraumatic stress disorder after treatment for breast cancer: Prevalence of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C) as a screening instrument. *Journal of Consulting and Clinical Psychology, 66*(3), 586-590. doi:10.1037/0022-006X.66.3.586

Armour, C., Carragher, N., & Elhai, J. D. (2013). Assessing the fit of the Dysphoric Arousal model across two nationally representative epidemiological surveys: The Australian NSMHWB and the United States NESARC. *Journal of Anxiety Disorders, 27*(1), 109-115. doi:10.1016/j.janxdis.2012.10.006

Armour, C., Raudzah Ghazali, S., & Elklit, A. (2013). PTSD's latent structure in Malaysian tsunami victims: Assessing the newly proposed Dysphoric Arousal model. *Psychiatry Research, 206*(1), 26-32. doi:10.1016/j.psychres.2012.09.012

Arnberg, F. K., Johannesson, K. B., & Michel, P.-O. (2013). Prevalence and duration of PTSD in survivors six years after a natural disaster. *Journal of Anxiety Disorders, 27*(3), 347-352.

Aziz, M. A., & Kenford, S. (2004). Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. *Journal of Psychiatric Practice, 10*(5), 307-313. doi:10.1097/00131746-200409000-00004

Beck, J. G., Grant, D. M., Read, J. P., Clapp, J. D., Coffey, S. F., Miller, L. M., & Palyo, S. A. (2008). The Impact of Event Scale-Revised: Psychometric properties in a sample of motor vehicle accident survivors. *Journal of Anxiety Disorders, 22*(2), 187-198. doi:10.1016/j.janxdis.2007.02.007

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34*(8), 669-673. doi:10.1016/0005-7967(96)00033-2

Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., & 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*(2), 272-281. doi:10.1037/0022-006X.76.2.272

Bollen, K., Tueller, S., & Oberski, D. (2013). Issues in the Structural Equation Modeling of Complex Survey Data. *Proceedings of the 59th World Statistics Congress 2013*. International Statistical Institute: Hong Kong. Retrieved from http://daob.nl/publications/

Brewin, C. R. (2005). Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress, 18*(1), 53-62. doi:10.1002/jts.20007

Brewin, C. R., Rose, S., Andrews, B., Green, J., Tata, P., McEvedy, C., . . . Foa, E. B. (2002). Brief screening instrument for post-traumatic stress disorder. *British Journal of Psychiatry, 181*(2), 158-162. doi:10.1192/bjp.181.2.158

Byrne, B. M. (2009). *Structural equation modeling with AMOS: basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. doi:10.1207/s15328007sem0902_5

Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale-Revised. *Behaviour Research and Therapy, 41*(12), 1489-1496. doi:10.1016/j.brat.2003.07.010

Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders, 25*(3), 340-345. doi:http://dx.doi.org/10.1016/j.janxdis.2010.10.007

Elhai, J. D., & Palmieri, P. A. (2011). The factor structure of posttraumatic stress disorder: A literature update, critique of methodology, and agenda for future research. *Journal of Anxiety Disorders, 25*(6), 849-854. doi:10.1016/j.janxdis.2011.04.007

Eriksson, N.-G., & Lundin, T. (1996). Early traumatic stress reactions among Swedish survivors of the m/s Estonia disaster. *British Journal of Psychiatry, 169*(6), 713-716. doi:10.1192/bjp.169.6.713

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1998). *Structured Clinical Interview for DSM-IV Axis I disorders (SCID I)* (J. Herlofson, Trans.). Stockholm, Sweden: Pilgrim Press (Original publication: 1997).

Furukawa, T., & Goldberg, D. P. (1999). Cultural invariance of likelihood ratios for the General Health Questionnaire. *Lancet, 353*, 561–562. doi:10.1016/S0140-6736(98)05470-1

Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire: Maudsley monograph No.21*. London, England: Oxford University Press.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*(1), 191-197. doi:10.1017/S0033291796004242

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*(3), 209-218.

Johannesson, K. B., Arnberg, F. K., & Michel, P.-O. (2012). Svenskarna som överlevde tsunamin mår relativt bra: Uppföljning sex år efter katastrofen 2004. [Swedish survivors from the tsunami are recovering: A follow-up six years after the disaster in 2004]. *Lakartidningen, 109*(37), 1607-1609.

Johannesson, K. B., Lundin, T., Fröjd, T., Hultman, C. M., & Michel, P.-O. (2011). Tsunami-exposed tourist survivors: signs of recovery in a 3-year perspective. *Journal of Nervous and Mental Disease, 199*(3), 162-169. doi:10.1097/NMD.0b013e31820c73d1

Johannesson, K. B., Michel, P.-O., Hultman, C. M., Lindam, A., Arnberg, F., & Lundin, T. (2009). Impact of exposure to trauma on posttraumatic stress disorder symptomatology in Swedish tourist tsunami survivors. *Journal of Nervous and Mental Disease, 197*(5), 316-323. doi:10.1097/NMD.0b013e3181a206f7

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*(12), 1048-1060. doi:10.1001/archpsyc.1995.03950240066012

King, D. W., Leskin, G., King, L., & Weathers, F. (1998). Confirmatory factor analysis of the clinician-administered PTSD scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment, 10*, 90 - 96.

King, D. W., Orazem, R. J., Lauterbach, D., King, L. A., Hebenstreit, C. L., & Shalev, A. Y. (2009). Factor structure of posttraumatic stress disorder as measured by the Impact of Event Scale-Revised: Stability across cultures and time. *Psychological Trauma: Theory, Research, Practice, and Policy, 1*(3), 173-187. doi:10.1037/a0016990

Krause, E. D., Kaltman, S., Goodman, L. A., & Dutton, M. A. (2007). Longitudinal factor structure of posttraumatic stress symptoms related to intimate partner violence. *Psychological Assessment, 19*(2), 165-175. doi:http://dx.doi.org/10.1037/1040-3590.19.2.165

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review, 30*(8), 976-987. doi:10.1016/j.cpr.2010.06.012

McHugh, P. R., & Treisman, G. (2007). PTSD: a problematic diagnostic category. *Journal of Anxiety Disorders, 21*(2), 211-222. doi:S0887-6185(06)00134-4

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11, Suppl. 3), S69–S77.

Moffitt, T. E., Caspi, A., Taylor, A., Kokaua, J., Milne, B. J., Polanczyk, G., & Poulton, R. (2010). How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine, 40*(06), 899-909. doi:10.1017/S0033291709991036

Morina, N., Böhme, H. F., Ajdukovic, D., Bogic, M., Franciskovic, T., Galeazzi, G. M., . . . Priebe, S. (2010). The structure of post-traumatic stress symptoms in survivors of war: Confirmatory factor analyses of the Impact of Event Scale—Revised. *Journal of Anxiety Disorders, 24*(6), 606-611. doi:10.1016/j.janxdis.2010.04.001

Olde, E., Kleber, R. J., van der Hart, O., & Pop, V. J. (2006). Childbirth and posttraumatic stress responses: A validation study of the Dutch Impact of Event Scale-Revised. *European Journal of Psychological Assessment, 22*(4), 259-267. doi:10.1027/1015-5759.22.4.259

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-12 (GHQ). *Journal of Anxiety Disorders, 22*(2), 337-343. doi:10.1016/j.janxdis.2007.02.010

Riegelman, R. K. (2005). *Studying a study and testing a test: How to read the medical evidence* (5th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1997). Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *American Journal of Psychiatry, 154*(11), 1593-1598.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514. doi:10.1007/bf02296192

Sconfienza, C. (1998). Mätning av psykiskt välbefinnande bland ungdomar i Sverige: användning av GHQ-12 [Measuring mental well-being among young people in Sweden: application of GHQ 12]. In *Arbete och hälsa 1998:22* National Institute for Working Life.

Shevlin, M., & Adamson, G. (2005). Alternative Factor Models and Factorial Invariance of the GHQ-12: A Large Sample Analysis Using Confirmatory Factor Analysis. *Psychological Assessment, 17*(2), 231-236. doi:10.1037/1040-3590.17.2.231

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.

Sveen, J., Low, A., Dyster-Aas, J., Ekselius, L., Willebrand, M., & Gerdin, B. (2010). Validation of a Swedish version of the Impact of Event Scale-Revised (IES-R) in patients with burns. *Journal of Anxiety Disorders, 24*(6), 618-622. doi:10.1016/j.janxdis.2010.03.021

Taft, C. T., Stern, A. S., King, L. A., & King, D. W. (1999). Modeling physical health and functional health status: The role of combat exposure, posttraumatic stress disorder, and personal resource attributes. *Journal of Traumatic Stress, 12*(1), 3-23. doi:10.1023/A:1024786030358

Thoresen, S., Tambs, K., Hussain, A., Heir, T., Johansen, V. A., & Bisson, J. I. (2010). Brief measure of posttraumatic stress reactions: impact of Event Scale-6. *Social Psychiatry and Psychiatric Epidemiology, 45*(3), 405-412. doi:10.1007/s00127-009-0073-x

Wang, L., Zhang, J., Shi, Z., Zhou, M., Huang, D., & Liu, P. (2011). Confirmatory factor analysis of posttraumatic stress symptoms assessed by the Impact of Event Scale-Revised in Chinese earthquake victims: Examining factor structure and its stability across sex. *Journal of Anxiety Disorders, 25*(3), 369-375. doi:10.1016/j.janxdis.2010.10.011

Wang, L., Zhang, J., Shi, Z., Zhou, M., Li, Z., Zhang, K., . . . Elhai, J. D. (2011). Comparing alternative factor models of PTSD symptoms across earthquake victims and violent riot witnesses in China: Evidence for a five-factor model proposed by Elhai et al. (2011). *Journal of Anxiety Disorders, 25*(6), 771-776. doi:10.1016/j.janxdis.2011.03.011

Wang, M., Elhai, J. D., Dai, X., & Yao, S. (2012). Longitudinal invariance of posttraumatic stress disorder symptoms in adolescent earthquake survivors. *Journal of Anxiety Disorders, 26*(2), 263-270. doi:http://dx.doi.org/10.1016/j.janxdis.2011.12.009

Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (1993). *The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility*. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX.

Weiss, D. S. (2004). The Impact of Event Scale-Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (2nd ed., pp. 168-189). New York, NY: Guilford Press.

Weiss, D. S., & Marmar, C. R. (1997). The Impact of Event Scale-Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 399-411). New York, NY: Guilford Press.

Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. *Depression and Anxiety, 28*(7), 596-606. doi:10.1002/da.20837

Yufik, T., & Simms, L. J. (2010). A meta-analytic investigation of the structure of posttraumatic stress disorder symptoms. *Journal of Abnormal Psychology, 119*(4), 764-776. doi:10.1016/j.paid.2005.11.034