Journal of Anxiety Disorders 26 (2012) 263–270

Contents lists available at SciVerse ScienceDirect

Journal of Anxiety Disorders

Longitudinal invariance of posttraumatic stress disorder symptoms in adolescent earthquake survivors

Mengcheng Wang[a], Jon D. Elhai[b], Xiaoyang Dai[c], Shuqiao Yao[a,*]

[a] Medical Psychological Institute, Second Xiangya Hospital, Central South University, Changsha, China
[b] Department of Psychology, University of Toledo, Toledo, OH, USA
[c] Department of Psychology, Shenzhen University, Shenzhen, China

ARTICLE INFO

Article history:
Received 27 September 2011
Received in revised form
10 December 2011
Accepted 17 December 2011

Keywords:
Longitudinal invariance
Confirmatory factor analysis
Posttraumatic stress disorder
PTSD Checklist-Civilian version

ABSTRACT

The present study examined the invariance of four models of posttraumatic stress disorder symptoms (PTSD) across time through an analysis of symptom ratings from the PTSD Checklist-Civilian version. Participants included 403 junior middle school students recruited from Sichuan, China at 5 and 11 months after the 2008 Sichuan earthquake. Confirmatory factor analysis using measurement and structural invariance testing found that four tested models were non-invariant (i.e., different) over time on PTSD's variable intercepts (indicating symptom severity) and residual error variances, but were invariant on PTSD's factor loadings and structural parameters. The two 4-factor intercorrelated models and the newly proposed 5-factor model were superior to the model defined in *DSM-IV*; however, the best fitting model was the newest proposed 5-factor model. These findings extend our understanding of PTSD's factor structure that the factor structure may be quite stable over time.

1. Introduction

In recent years, researchers have debated the underlying symptom structure of PTSD, and no clear agreement has been reached (reviewed in Elhai & Palmieri, 2011). Various structural models ranging from 2- to 4-factor models, including first-order intercorrelated and hierarchical models, have been supported by empirical evidence in various studies. To date, in contrast to the *DSM-IV* three-factor model, two alternative four-factor, first-order models (King, Leskin, King, & Weathers, 1998; Simms, Watson, & Doebbeling, 2002) have received the most empirical support (reviewed in Elhai & Palmieri, 2011; Yufik & Simms, 2010). Unfortunately, few papers have tested the stability of the two alternative four-factor models over time (Krause, Kaltman, Goodman, & Dutton, 2007).

King, Leskin, King, and Weathers (1998) proposed a 4-factor, first-order model of PTSD that has 4 separate factors of intrusion (i.e., reexperiencing), avoidance, emotional numbing, and hyperarousal (PTSD-numbing, see Table 1), arguing that avoidance and numbing are statistically distinct constructs. This model has significant empirical support. It was first tested and supported in King et al.'s sample of 524 male veterans, in a series of nested models reflecting alternative representations of PTSD's dimensionality. Further support for the model was received from studies of various traumatic events among different populations including, for example, from military veterans (Elhai, Palmieri, Biehn, Frueh, & Magruder, 2010), college students (Elhai, Gray, Docherty, Kashdan, & Kose, 2007), cancer survivors (DuHamel et al., 2004) and refugees (Palmieri, Marshall, & Schell, 2007).

Simms, Watson, and Doebbeling (2002) proposed another 4-factor model for PTSD that comprises 4 intercorrelated factors: intrusion, avoidance, hyper-arousal, and dysphoria (PTSD-dysphoria, see Table 1). This model differs from that of King et al.'s (1998) model in that it combines all of the emotional numbing symptoms and 3 hyperarousal symptoms (trouble sleeping, irritability, and poor concentration) of PTSD to generate a new dysphoria factor, while it leaves 2 symptoms (being overly alert and easily startled) to comprise the hyperarousal factor. This model reflects that several PTSD symptoms (of dysphoria) represent items that are not unique to PTSD but involve anxiety and distress more generally. This model demonstrated superior fit in a group of CFA studies, such as those on veterans of the Gulf War (Simms et al., 2002), utility workers exposed to the World Trade Center Ground Zero site (Palmieri, Weathers, Difede, & King, 2007), women survivors of intimate partner violence (Krause et al., 2007), and adolescents from the general population (Elhai, Ford, Ruggiero, & Frueh, 2009).

The difference between the four-factor PTSD models involves only which factor (dysphoria or hyperarousal) includes symptoms D1–D3 (i.e., sleep difficulty, irritability, and concentration problems). More recently, Elhai et al. (2011) proposed a new alternative

* Corresponding author at: Medical Psychological Institute, Second Xiangya Hospital, Central South University, #139 Ren-Min Zhong Road, Changsha 410011, China. Tel.: +86 731 85292130; fax: +86 731 85361328.
E-mail address: Shuqiaoyao@163.com (S. Yao).

0887-6185/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.janxdis.2011.12.009

Table 1
Item mapping for tested models.

DSM-IV PTSD symptoms	Model			
	DSM	King's	Simms's	Elhai's
B1. Intrusive thoughts	R	R	R	R
B2. Nightmares	R	R	R	R
B3. Reliving trauma/flashback	R	R	R	R
B4. Emotional cue reactivity	R	R	R	R
B5. Sudden physical reaction	R	R	R	R
C1. Avoidance of thoughts/feelings	A	A	A	A
C2. Avoidance of reminders	A	A	A	A
C3. Specific amnesia	A	N	D	N
C4. Loss of interest	A	N	D	N
C5. Feeling distant	A	N	D	N
C6. Feeling numb	A	N	D	N
C7. Hopelessness	A	N	D	N
D1. Trouble sleeping	H	H	D	DA
D2. Irritable/angry	H	H	D	DA
D3. Poor/difficulty concentration	H	H	D	DA
D4. Overly alert	H	H	H	AA
D5. Exaggerated startle response	H	H	H	AA

Note. DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; PTSD, posttraumatic stress disorder; Factors: P, posttraumatic stress; R, re-experiencing; A, avoidance; N, numbing; H, hyper-arousal; D, dysphoria; DA, dysphoric arousal; AA, anxious arousal.

model, which conceptualizes the three D1–D3 symptoms as a separate dysphoric arousal factor which differs conceptually from both PTSD's anxious arousal and dysphoria factors. Thus, the newly proposed five-factor model includes Re-experiencing (B1–B5), avoidance (C1–C2), numbing (C3–C7), dysphoric arousal (D1–D3) and anxious arousal (D4–D5). Elhai et al. (2011) found that the alternative five-factor model fit the data significantly better than both the four-factor models in a sample of 252 women victims of domestic violence. More recently, this model was identified to be the best fitting model in separate samples of Chinese adolescent earthquakes and riot victims (Wang, Long, Li, & Armour, 2011; Wang, Zhang, et al., 2011). Therefore, in our present investigation, we also took this newest model into account.

Although an increasing number of studies have confirmed the factor structure of PTSD symptoms through the use of a cross-sectional and/or single sample methodology, relatively few studies have examined the factorial invariance of PTSD symptoms across samples or time (discussed in Elhai & Palmieri, 2011). On the basis of the literature reviewed, we found several studies that investigated the factorial invariance of PTSD symptoms across samples (e.g., Armour et al., 2011; Engdahl, Elhai, Richardson, & Frueh, 2011; Hoyt & Yeater, 2010; Krause et al., 2007; Mansfield, Williams, Hourani, & Babeu, 2010; Marshal, 2004; Miles, Marshall, & Schell, 2008). For instance, in a recent investigation, Engdahl et al. (2011) revealed that model fit for the PTSD-numbing and PTSD-dysphoria models was not significantly different; however, measurement invariance testing demonstrated significant differences between deployed and nondeployed military veteran groups on all structural parameters, except for observed variable intercepts.

Until now, far fewer studies have examined the longitudinal invariance of PTSD symptoms. Baschnagel, O'Connor, Colder, and Hawk (2005) suggested that the PTSD-dysphoria model proposed by Simms et al. (2002) represents the best-fitting model compared with several alternative models examined in the literature at 1 and 3 months after the September 11th terrorist attacks. However, Baschnagel et al. (2005) did not strictly test longitudinal invariance, because they did not report fit statistics of the constrained model or which particular loadings were non-invariant. Another study conducted by Krause et al. (2007), which used samples of women who had suffered intimate partner violence, was a relatively strictly longitudinal invariance test because they only tested metric and phi invariance. They found that the PTSD-dysphoria model (Simms et al., 2002) with metric and phi invariance constrained fit the data very well. But Krause et al.'s research has two shortcomings: they did not test scalar and residual error invariance; and they did not conduct chi-square difference tests to assess if the constrained models were significantly different from the less constrained models. Thus we lack a comprehensive understanding of how PTSD's factor structure changes over time.

It is worth noting that most of the above mentioned studies on the factor structure of PTSD have been conducted exclusively with adults; limited factor analytic studies were devoted to the structure of PTSD in children and adolescents (e.g., Armour et al., 2011; Elhai et al., 2009; Wang, Zhang, et al., 2011). To extend knowledge on the structure of PTSD symptoms, we examined four alternative models of PTSD symptoms developed with adults, specifically in a sample of adolescent earthquake survivors.

Moreover, if invariance does not hold over two or more longitudinal waves, differences in observed PTSD scores may not be directly comparable (Meade, Lautenschlager, & Hecht, 2005). The establishment of longitudinal measurement invariance is a natural prerequisite to modeling changes over time through random effects modeling techniques such as latent growth curve analysis (Vandenberg & Lance, 2000). Because mean levels of PTSD symptoms reportedly decline over time among trauma survivors (e.g., Blanchard, Jones-Alexander, Buckley, & Forneris, 1996), this result would be meaningless if this comparison of mean scores over time did not assume measurement invariance over time (Pitts, West, & Tein, 1996; Vandenberg & Lance, 2000).

In order to test the measurement and structural stability of PTSD symptoms across time and to choose the best fitting PTSD model, the present study use a longitudinal data set to examine the longitudinal invariance of PTSD symptoms in a sample of adolescent earthquake survivors not previously studied (Wang, Long, et al., 2011). The two most-supported alternative four-factor models, the PTSD-numbing (King et al., 1998) and PTSD-dysphoria (Simms et al., 2002) models, as well as the newly proposed five-factor model (Dysphoric Arousal model; Elhai et al., 2011) and the 3-factor model defined in the *DSM-IV*, were chosen for this examination.

2. Method

2.1. Participants and procedure

The sample consisted of 403 junior middle school students who were selected from Xiaojin, which ranked as the worst hit county among the 51 earthquake-afflicted counties in Sichuan, China in the 2008 Sichuan earthquake. Before administrating self-reported questionnaires to the participants, parents (or legal guardians) provided informed consent; children then provided assent for their own participation.

The first data wave was conducted on October 8th 2008, 150 days after the earthquake, 560 adolescents were recruited; the follow-up assessment was conducted 6 months after the time of the first assessment, 465 adolescents attended. We excluded participants who did not fill out the PTSD surveys at least one symptom item at two time point, leaving an effective sample size of 403. Of the final sample, 189 (33.3%) were males and 214 (53.1%) were females. 33.3% were in the equivalent of the United States' Grade 6, with 34% in Grade 7 and 32.8% in Grade 8. The participants' ages ranged 12–17 years ($M = 14$, $SD = 1.15$).

2.2. Measurement

PTSD symptoms were assessed with the PTSD Checklist-Civilian version (PCL-C), which is one of the most popular instruments used in previous PTSD studies (e.g., Palmieri, Weathers, et al., 2007). The

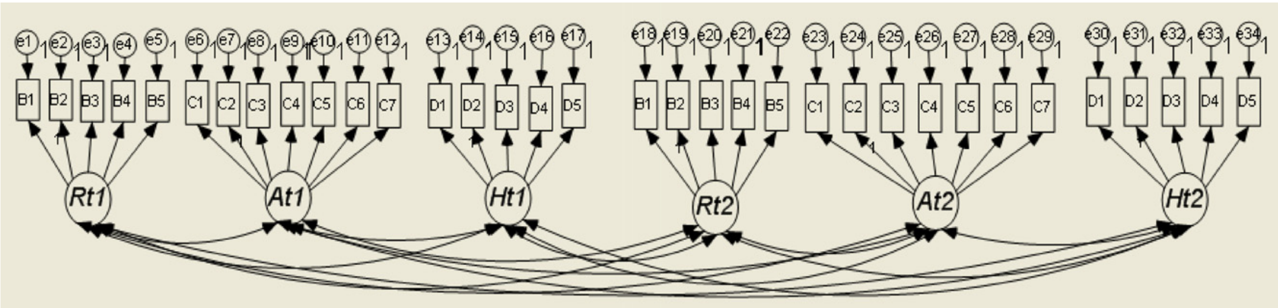

Fig. 1. Model for the longitudinal invariance analyses specified in *DSM-IV*. *Note*. R, re-experiencing; A, avoidance; H, hyper-arousal; *t*1, Time 1; *t*2, Time 2. For each symptom, it is Time 1 error variance was allowed to covary with its Time 2 error variance.

PCL-C was developed to assess PTSD symptoms in civilian populations (Weathers, Litz, Herman, Huska, & Keane, 1993), and consists of 17 items that correspond directly to the PTSD symptom criteria listed in the *DSM-IV*. Participants were asked to rate the extent to which they had been bothered by each symptom in the past month on a 5-point Likert scale, ranging from 1 (not at all) to 5 (extremely). In present study, participants were specifically instructed to complete the PCL-C referring to the "Wen Chuan Da Di Zhen (Wen Chuan Earthquake)". Several prior studies have shown that the scale is temporally stable, reliable, and has strongly convergence with PTSD symptoms based on structured diagnostic interviews (e.g., Weathers et al., 1993; Blanchard et al., 1996; reviewed in Wilkins, Lang, & Norman, 2011).

The Chinese version of the PCL was adapted via a two-stage process of translation and back translation (Wu, Chan, & Yiu, 2008). Although the PCL was primarily validated in adults (age 18 and over), its applicability in adolescents has been demonstrated by previous studies with Western and Chinese samples (e.g., Barnes, Treiber, & Ludwig, 2005; Calderoni, Alderman, Silver, & Bauman, 2006; Li, Wang, Shi, Zhang, Wu, & Liu, 2010; Wang, Dai, & Wan, 2009; Wang, Long, et al., 2011; Wang, Zhang, et al., 2011). The internal consistency (Cronbach's Alpha) of the PCL-C in the current sample at the two time points was .879 at baseline and .880 at follow-up.

2.3. Model specification

The conventional way of testing invariance is through multiple-group confirmatory factor analysis. However, that method assumes categorization of multiple groups or samples of subjects. Testing longitudinal invariance of the factor structure requires analyzing data by fitting the two or more waves of data with two separate models simultaneously (Configural Invariance model specified in *DSM-IV*; see Fig. 1). Therefore, two separate factor models were constructed for Time 1 and Time 2 data with factor correlations estimated between Time 1 and Time 2 factors. In addition, correlations were estimated among all possible pairs of residual error variance between Time 1 and Time 2 because the same items were used across two time points (Pitts et al., 1996); in other words, a given symptom's residual error variance for Time 1 was allowed to covary with that symptom's error variance at Time 2.

If configural invariance (baseline model) is supported, further restrictive constraints can then be imposed on the model, as is conducted in the conventional multiple group CFA invariance test. First, factor loadings are constrained to be equal across time to test *metric* or *weak invariance* (Model 1). A chi-square difference test is conducted to assess if the baseline model is significantly different from the constrained model. A non-significant chi-square difference test would indicate that factor loadings are invariant across time, satisfying metric invariance. Further, based on the metric invariance model, intercepts are constrained to be equal across time to build Model 2, a test of *scalar* or *strong invariance*. Model 3 included the restrictions from Model 2 plus the additional constraint of equal item error variances across the two waves (*Invariant Error Variance or strict invariance*). Subsequent to the Model 3, residual error variances are not constrained to be equal across timepoints (Grouzet, Otis, & Pelletier, 2006); thus Model 4 would be compared to Model 2, in order to preserve nested model testing. Model 4 included the constraints from Model 2 plus the additional constraint of equal factor variances across the two waves (*invariant factor variances*). Model 5 included the restrictions from Model 4 plus the additional constraint of equal factor covariances across the two waves (*invariant factor covariances*). The first three invariance testing analyses are also called *measurement invariance*, while the next two invariance testing analyses are called *structural invariance*. The invariance routine was ordered in a similar manner to that of Grouzet et al. (2006).

2.4. Data analyses

Descriptive statistical analysis was performed with the Statistical Package of Social Science (SPSS) 15.0 software. The longitudinal invariance analysis within the framework of CFA was conducted with the Mplus 5.1 program (Muthén & Muthén, 1998–2007). The skewness for each item ranged from .25 (D5 at Time 2) to 1.68 (C7 at Time 2), kurtosis ranged from .01 (C3 at Time 1) to 2.08 (C7 at Time 2) (Table 2). Thus, we used maximum likelihood estimation with a mean-adjusted chi-square (MLM) which is robust to non-normality. The MLM chi-square test statistic is also referred to as the Satorra–Bentler chi-square (Muthén & Muthén, 1998–2007; Satorra & Bentler, 2001).

Following generally accepted practice, we evaluated the fit of each model by examining multiple fit indices (Kline, 2010). We used the Satorra–Bentler chi-square statistic, root-mean-square error of approximation (RMSEA), standardized root-mean-square residual (SRMR), Tucker–Lewis index (TLI), and comparative fit index (CFI). On the basis of extensive simulation studies conducted by Hu and Bentler (1999), it appears that good-fitting models have CFI and TLI values greater than .95, RMSEA values less than .06, and less than .08.

The corrected scaled chi-square difference test developed by Satorra and Bentler (2001; Muthén & Muthén, 2006) was used to compare nested models. However, tests of the change in CFI are superior to chi-square difference tests of invariance, because they are not affected by the sample size (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008). Thus, the corrected scaled chi-square difference test and change in CFI were used to compare nested models. When both results contradict each other, however, we primarily depended on results of CFI differences. Bayesian information criterion (BIC) was used to compare non-nested models. As suggested by Raftery (1995), a BIC difference of

Table 2
Item-level descriptive statistics for the PTSD Checklist-Civilian version (PCL-C) at two time points.

PCL-C item	Time 1				Time 2			
	M	*SD*	Skew	Kurt	*M*	*SD*	Skew	Kurt
B1. Intrusive thoughts	2.34	1.13	.73	−.09	2.34	1.06	.77	.25
B2. Nightmares	2.00	1.18	1.06	.17	1.74	.94	1.34	1.49
B3. Reliving trauma/flashback	2.24	1.19	.76	−.32	2.10	1.10	.82	−.06
B4. Emotional cue reactivity	2.73	1.25	.34	−.84	2.66	1.19	.35	−.75
B5. Sudden physical reaction	2.21	1.25	.81	−.37	2.02	1.07	.88	.10
C1. Avoidance of thoughts/feelings	2.11	1.10	.92	.21	2.13	1.04	.82	.20
C2. Avoidance of reminders	2.12	1.13	.85	−.03	2.17	1.08	.73	−.11
C3. Specific amnesia	2.14	1.11	.85	−.01	2.11	1.09	.86	.14
C4. Loss of interest	1.91	1.16	1.17	.36	1.93	1.08	1.09	.45
C5. Feeling distant	1.88	1.12	1.30	.92	1.99	1.08	.99	.26
C6. Feeling numb	1.89	1.14	1.08	.08	2.08	1.16	.90	−.07
C7. Hopelessness	1.82	1.21	1.34	.66	1.65	1.04	1.68	2.08
D1. Trouble sleeping	2.02	1.17	1.06	.22	2.10	1.13	.98	.32
D2. Irritable/angry	2.12	1.16	.99	.26	2.36	1.16	.07	−.21
D3. Poor/difficulty concentration	2.34	1.17	.78	−.17	2.58	1.20	.50	−.60
D4. Overly alert	2.10	1.23	.93	−.18	2.09	1.08	.87	.16
D5. Exaggerated startle response	2.82	1.27	.29	−.98	2.71	1.20	.25	−.86

Note. Items are listed the order as they appear in the *DSM-IV* (the Diagnostic and Statistic Manual of Mental Disorder, Fourth Edition).

6–10 points indicates strong support, and a difference of >10 points indicates very strong support for the model with lower BIC value.

According to the suggestion of Cheung and Rensvold (2002), the change in CFI was chosen to evaluate the measurement invariance. A value of CFI difference <.01 indicates that the invariance hypothesis should not be rejected; mean differences exist when CFI differences are .01 to .02; and definite differences exist when CFI differences are >.02 (Cheung & Rensvold, 1999, 2002).

3. Results

We sought to evaluate more systematically the stability of the factor structures of three alternative models well identified in prior studies (e.g., Elhai & Palmieri, 2011; Krause et al., 2007; Palmieri, Weathers, et al., 2007; Yufik & Simms, 2010), as well as the 3-factor model defined in *DSM-IV* (American Psychiatric Association, 2000). Thus, 4 separate sets of factorial invariance analyses were performed. Table 3 presents the results of the model fits and comparisons.

3.1. Configural invariance

The 3-factor model defined in *DSM-IV* as the baseline model for configural invariance resulted in an adequate but not excellent fit; its fit indices only reached recommended marginal values (TLI = .899, CFI = .911, RMSEA = .39, SRMR = .049). Moreover, BIC values were the largest (and thus worst) ones among four tested models. Especially, compared to other alternative models, the 3-factor model was inferior; even so, we retained this model in the next step of the analysis. Moreover, the three alternative models were excellent as baseline models for configural invariance, because those models' fit indices indicated excellent model fit. The RMSEA and SRMR for the 3 factor model were distinctly higher than the RMSEA and SRMR of the three alternative models. Solely from the point of model fitting, however, the five-factor Dysphoric Arousal model has the lowest chi-square (594.336, df = 465) and best fit indices (TLI = .953, CFI = .961, RMSEA = .26, SRMR = .42) compared to other models. Furthermore, chi-square difference tests indicated that the Dysphoric Arousal model fit significantly better than the Dysphoria model (Table 4), S–Bχ^2_{diff} (17, N = 403) = 35.92, p < .01, but not significantly different than the Emotional Numbing model, S–Bχ^2_{diff} (17, N = 403) = 25.56, p > .05. BIC value differences between the Emotional Numbing model and Dysphoria model was 12.67 (>10), suggesting the Emotional Numbing model fit the data significantly better than the Dysphoria model.

3.2. Metric invariance

Based on the results of configural invariance testing, factor loadings are constrained to be equal across two times for four models. The chi-square difference tests indicated that all of the models were rejected at .05 significant level, but two models, *DSM-IV*'s model and PTSD-dysphoria, were rejected at the .01 significant level. Meanwhile, the fit indices supported those four models with metric invariance because all CFI differences were less than .01 (Cheung & Rensvold, 1999, 2002; ΔTLI = .001 and ΔCFI = .004 for *DSM-IV*'s model; ΔTLI = .004 and ΔCFI = .003 for the Dysphoria model; ΔTLI = .004 and ΔCFIs = .005 for the Numbing model; ΔTLI = .004 and ΔCFI = .005 for the Dysphoric Arousal model). Thus, we further constrained observed variable intercepts to be equal (testing strong or scalar invariance).

Chi-square difference test indicated that the Dysphoric Arousal model fit significantly better than the Dysphoria model, S–Bχ^2_{diff} (18, N = 403) = 35.12, p < .01, but no significant difference was found compared with Emotional Numbing model, S–Bχ^2_{diff} (18, N = 403) = 26.29, p > .05. BIC value differences between the Emotional Numbing model and Dysphoria model was 10.11 (>10), suggesting Emotional Numbing model fit the data significantly better than the Dysphoria model at the metric invariance level.

3.3. Scalar invariance

Observed variable intercepts were constrained to be equal and resulted in adequate model fit for the three alternative models rather than for *DSM-IV*'s model. The chi-square difference tests indicated that the model constraining observed variable intercepts to be equal fit the data significantly worse than the less constrained model, p < .001. The fit index differences also indicated mean differences existing between the more constrained models and less constrained models, >.01 (ΔTLI = .015 and ΔCFI = .017 for *DSM-IV*'s model; ΔTLI = .015 and ΔCFI = .015 for the Dysphoria model; ΔTLI = .017 and ΔCFI = .017 for the Numbing model; ΔTLI = .017 and ΔCFI = .016 for the Dysphoric Arousal model).

The Dysphoric Arousal model fit the data significantly better than the Dysphoria model, S–Bχ^2_{diff} (18, N = 403) = 35.12, p < .05, as well as the Emotional Numbing model, S–Bχ^2_{diff} (18, N = 403) = 30.00, p < .05. However, there were no difference

Table 3
Goodness-of-fit indices and model comparisons for tested models.

Model	S–Bχ^2	*df*	$\Delta\chi^2$	Δdf	*p*	TLI	CFI	SRMR	BIC	RMSEA (90% CI)
DSM's model										
Configural invariance	792.233	495	–	–	–	.899	.911	.049	39342.838	.039 [.034, .044]
Metric invariance	818.22	509	25.984	14	.026	.898	.907	.054	39287.538	.039 [.034, .044]
Scalar invariance	891.852	526	80.766	17	<.001	.883	.890	.056	39269.652	.042 [.037, .046]
Error variance invariance	952.882	543	64.740	17	<.001	.873	.877	.058	39239.207	.043 [.039, .048]
Invariant factor variances	894.829	529	2.090	3	.554	.883	.889	.058	39253.903	.041 [.037, .046]
Invariant factor covariances	896.885	532	2.049	3	.562	.884	.890	.058	39238.543	.041 [.037, .046]
Simms's model/PTSD-dysphoria										
Configural invariance	630.582	482	–	–	–	.948	.955	.044	39217.775	.028 [.021, .035]
Metric invariance	656.880	495	26.784	13	.014	.945	.951	.049	39170.145	.028 [.022, .034]
Scalar invariance	726.030	512	78.472	17	<.001	.930	.936	.051	39147.625	.032 [.027, .037]
Error variance invariance	781.979	529	59.440	17	<.001	.919	.924	.053	39111.348	.034 [.029, .039]
Invariant factor variances	734.619	516	9.362	4	.053	.929	.934	.057	39132.887	.032 [.027, .038]
Invariant factor covariances	743.643	522	8.961	6	.176	.928	.933	.054	39109.454	.032 [.027, .038]
King's model/PTSD-numbing										
Configural invariance	619.979	482	–	–	–	.952	.959	.043	39205.105	.027 [.020, .033]
Metric invariance	648.300	495	29.467	13	.006	.948	.954	.048	39160.036	.028 [.021, .033]
Scalar invariance	722.119	512	85.965	17	<.001	.931	.937	.051	39143.356	.032 [.026, .037]
Error variance invariance	787.285	529	68.466	17	<.001	.918	.922	.053	39118.419	.035 [.030, .040]
Invariant factor variances	730.461	516	8.655	4	.070	.930	.936	.056	39128.295	.032 [.027, .037]
Invariant factor covariances	738.335	522	7.920	6	.244	.930	.935	.056	39103.140	.032 [.027, .037]
Elhai's model/dysphoric arousal										
Configural invariance	594.336	465	–	–	–	.953	.961	.042	39274.897	.026 [.019, .032]
Metric invariance	622.011	477	28.453	12	.005	.949	.956	.047	39235.582	.027 [.021, .033]
Scalar invariance	692.121	494	80.746	17	<.001	.932	.940	.050	39214.344	.032 [.026, .037]
Error variance invariance	748.412	511	127.234	17	<.001	.922	.929	.052	39178.702	.034 [.029, .039]
Invariant factor variances	701.052	499	9.301	5	.098	.932	.939	.055	39193.904	.032 [.026, .037]
Invariant factor covariances	713.101	509	12.201	10	.272	.932	.939	.052	39150.397	.032 [.026, .037]

Note. S–Bχ^2, scaled Satorra–Bentler χ^2; CFI, comparative fit index; TLI, Tucker–Lewis index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval. Configural invariance, no parameters constrained to be equal over time, correlations were estimated among all possible pairs of residual error variance between Time 1 and Time 2; metric invariance, configural invariance plus factor loadings are constrained to be equal across time; scalar invariance, metric invariance plus item intercept are constrained to be equal across time; error variance invariance, scalar invariance plus the additional constraint of equal item error variances across the two waves; invariant factor variances, scalar invariance plus the additional constraint of equal factor variances across the two waves; invariant factor covariances, invariant factor variances plus factor covariances are constrained to be equal across time.

between the Emotional Numbing model and Dysphoria model; the BIC value difference was 4.269 (<6).

3.4. Residual error invariance

Error Variance was constrained to be equal and yielded adequate model fit for three alternative models rather than for *DSM-IV*'s model. The chi-square difference tests indicated that the more constrained models fit the data significantly worse than the less constrained models, $p < .001$. The fit index differences also indicated mean differences existing between the more constrained models and less constrained models (ΔTLI = .010 and ΔCFI = .013 for *DSM-IV*'s model; ΔTLI = .011 and ΔCFI = .012 for the Dysphoria model; ΔTLI = .013 and ΔCFI = .015 for the Numbing model; ΔTLI = .010 and ΔCFI = .011 for the Dysphoric Arousal model).

The Dysphoric Arousal model fit the data significantly better than the Dysphoria model, S–Bχ^2_{diff} (18, N = 403) = 33.74, $p < .05$, as well as the Emotional Numbing model, S–Bχ^2_{diff} (18, N = 403) = 38.88, $p < .01$ Interestingly, the BIC value difference between the Emotional Numbing model and Dysphoria model reached 7.071 (>6), suggesting that the Dysphoria model fit better.

3.5. Invariant factor variances

The chi-square difference tests indicated that the additional constraint of equal factor variances across two waves did not affect the model fit, $p > .05$ for all models. Furthermore, the fit indices supplied additional evidence, with changes in TLI and CFIs for four models being less than .005 and RMSEA also revealing good fit. Additionally, chi-square difference testing showed that the Dysphoric Arousal model fit significantly better than the Dysphoria model when factor variances were constrained to be equal across time, S–Bχ^2_{diff} (17, N = 403) = 33.57, $p < .01$, and the Emotional Numbing model, S–Bχ^2_{diff} (17, N = 403) = 29.28, $p < .05$. BIC

Table 4
Parameter comparison among three alternative models.

Model comparison	M0	M1	M2	M3	M4	M5
M_N vs. M_{DA}						
ΔS–Bχ^2	25.555	26.290	30.000*	38.881**	29.280*	25.242*
Δdf	17	18	18	18	17	13
M_D vs. M_{DA}						
ΔS–Bχ^2	39.923**	35.124**	33.911*	33.743*	33.570**	30.535**
Δdf	17	18	18	18	17	13
M_D vs. M_N						
BIC_{diff}	12.67	10.109	4.269	−7.071	4.592	6.314

Note. ΔS–Bχ^2, difference between more constrained model and less constrained model; BIC_{diff}, difference between non-nested model; M_{DA}, Dysphoric Arousal model; M_N, PTSD-numbing model; M_D, PTSD-dysphoria model; M0, configural invariance; M1, metric invariance; M2, scalar invariance; M3, error variance invariance; M4, factor variances invariance; M5, factor covariances invariance.

* $p < .05$.
** $p < .01$.

Table 5
Standardized factor loadings for four models by time.

DSM-IV PTSD criterion	*DSM-IV*	King's model	Simms's model	Elhai's model
B1. Intrusive thoughts	.595/.667	.607/.678	.603/.673	.605/.675
B2. Nightmares	.440/.516	.439/.520	.440/.520	.438/.519
B3. Reliving trauma/flashback	.696/.703	.689/.698	.690/.703	.690/.701
B4. Emotional cue reactivity	.648/.697	.650/.700	.650/.698	.650/.699
B5. Sudden physical reaction	.572/.615	.568/.611	.569/.611	.568/.611
C1. Avoidance of thoughts/feelings	.573/.610	.759/.764	.755/.756	.761/.765
C2. Avoidance of reminders	.564/.574	.740/.701	.744/.708	.738/.700
C3. Specific amnesia	.479/.468	.481/.464	.473/.456	.478/.460
C4. Loss of interest	.548/.574	.590/.609	.570/.589	.587/.607
C5. Feeling distant	.645/.634	.712/.697	.681/.672	.709/.695
C6. Feeling numb	.669/.670	.691/.707	.685/.705	.695/.712
C7. Hopelessness	.440/.509	.458/.528	.459/.528	.460/.529
D1. Trouble sleeping	.487/.505	.486/.503	.469/.485	.479/.494
D2. Irritable/angry	.559/.564	.562/.568	.546/.549	.573/.582
D3. Poor/difficulty concentration	.607/.574	.609/.575	.604/.571	.625/.592
D4. Overly alert	.589/.666	.590/.668	.658/.738	.654/.733
D5. Exaggerated startle response	.538/.564	.530/.557	.580/.598	.584/.601

Note. R, re-experiencing; A, avoidance; N, numbing; H, hyper-arousa; D, dysphoria.
The second time loadings followed by first time loadings.
All loadings are significant at $p < 001$.

value differences between the Emotional Numbing model and Dysphoria models was 4.59 (<6), suggesting no difference between them.

3.6. Invariant factor covariances

Finally, the equality of the correlations between the factors was further imposed to test structural invariance over time. The chi-square test at $p < .01$ significant level indicated that there were no substantial differences between unconstrained correlations and their constrained counterparts. Furthermore, differences between the fit indices were very subtle, with changes in CFI and TLI less than .001 for the four models. According to the aforementioned criteria, all of this evidence supported the invariant factor covariances in the four models.

Additionally, chi-square difference tests showed that the Dysphoric Arousal model fit significantly better than the Dysphoria model when factor covariances were constrained to be equal across time, S–Bχ^2_{diff} (13, $N = 403$) = 25.24, $p < .05$, as well as the Emotional Numbing model, S–Bχ^2_{diff} (15, $N = 403$) = 30.54, $p < .01$. BIC value differences between the Emotional Numbing model and Dysphoria models was 6.31 (>6), suggesting Emotional Numbing model fit the data better than Dysphoria model.

3.7. Factor loadings and correlations

The standardized factor loadings for the tested models are presented in Table 5 by time of assessment. All factor loadings for the four models were moderate to high and significant ($p < .001$ for all) and ranged from .44 to .76. The results demonstrated that the D4–D5 symptoms were referred to as a separate factor in the Dysphoric Arousal model and PTSD-dysphoric model with higher factor loadings than that in the *DSM-IV* model and PTSD-numbing model. D1–D3 items were treated as an independent factor in the Dysphoric Arousal model with higher factor loadings than those of the other three models across the two time points.

The factor correlations for three alternative models are presented in Table 6. These correlations were moderate to high across time in three models. The dysphoric arousal factor as a new construct correlated with the other four factors at a moderate to high level, as well as different with the correlations between the anxious arousal factor and other factors in Dysphoric Arousal model or hyperarousal factor in the PTSD-dysphoria model; and also different with the correlations between the hyperarousal factor and other factors in the PTSD-numbing model.

Taken together, it is appropriate that D1–D3 and D4–D5 were simultaneously treated as independent factors in PTSD-Dysphoric Arousal model.

Table 6
Intercorrelation matrix of latent factors for the Elhai's model, PTSD-numbing model and PTSD-dysphoria model.

Elhai's model	Reexperiencing	Avoidance	Numbing	Dysphoric arousal
Avoidance	.745			
Numbing	.688	.610		
Dysphoric arousal	.721	.658	.904	
Anxious arousal	.728	.619	.758	.848

PTSD-numbing	Reexperiencing	Avoidance	Numbing	Hyper-arousal
Avoidance	.745			
Numbing	.687	.610		
Hyper-arousal	.762	.683	.877	1

PTSD-dysphoria	Reexperiencing	Avoidance	Dysphoria	Hyper-arousal
Avoidance	.747			
Dysphoria	.717	.638		
Hyper-arousal	.727	.639	.804	1

Note. All correlations are significant at $p < .01$.

4. Discussion

Although an increasing number of studies have confirmed the factor structure of PTSD symptoms through the use of a cross-section and/or single sample analysis, few studies have investigated the factorial invariance of PTSD symptoms across time (Elhai & Palmieri, 2011). In order to fill this gap in knowledge, in our current study, we used a longitudinal dataset to test the longitudinal factorial invariance of four representative models in a sample of adolescent earthquake survivors.

Substantial evidence consistently indicates that the 3-factor model defined in *DSM-IV* fits worse compared to other PTSD models (Elhai & Palmieri, 2011). In the present study, the fit indices of the configural invariance testing were in line with this conclusion; however, the chi-square differences tests of structural invariance for the 3-factor model rejected the hypothesis that the more restrictive models were significantly different from the less-constrained model. Additionally, most of the factor loadings of the 3-factor model were lower than those of other alternative models. Like in most cross-section investigations (e.g., Elhai et al., 2011; Engdahl et al., 2011; Wang et al., 2009), these results led us to conclude that compared with other more well-fitting models, the 3-factor model was unacceptable.

Similar to cross-sectional measurement invariance findings, configural and metric invariance were found for both 4-factor models (Krause et al., 2007; Mansfield et al., 2010; Simms et al., 2002) and new proposed 5-factor model, suggesting that symptoms corresponding to common factors have the same meaning across time (Gregorich, 2006; Meredith & Teresi, 2006). It is very important to understand the meaning of underlying factors across time. Our results failed to achieve scalar invariance, indicating that observed scores across time points were not the same for the identical underlying factors (Cheung & Rensvold, 2002; Gregorich, 2006; Meredith & Teresi, 2006); in other words, any differences in raw scores that could be observed between two time points should be mainly interpreted in terms of an intercept difference rather than a factor mean difference. A meaningful comparison of both observed means and variances requires strict factorial invariance or error variance invariance, and thus our results did not support those comparisons between two time points. Structural invariance including factor variance and covariance invariance held consistent, meaning that associations among different constructs were fairly stable (Pitts et al., 1996).

Previous studies, using cross-sectional data, have provided evidence that either the emotional numbing or dysphoria four-factor PTSD models offer the best fitting of the underlying factor structure of PTSD symptoms (Elhai & Palmieri, 2011; Yufik & Simms, 2010). We found additional empirical support for these four-factor models of PTSD. The fit indices for the configural invariance testing suggested that the four factor alternative models fit the data very well. Moreover, the further invariance analysis demonstrated that the two models satisfied metric invariance and structural invariance, but the BIC_{diff} result supported the Numbing model rather than the Dysphoria model at three of five parameters constrained levels. It was surprising that there were no substantial differences in terms of scalar invariance and factor variances invariance model between them, as well as the Dysphoria model fit better than the Numbing model when Error Variances were constrained to be equal. Specifically, our study was the first one that simultaneously compared the longitudinal factorial invariance of the two models, with results offering mort longitudinal support for the Numbing model two models than for the Dysphoria model.

In contrast to previous studies of the longitudinal factorial invariance of PTSD symptoms (Baschnagel et al., 2005; Krause et al., 2007) that identified the PTSD-dysphoria model as the superior model, our results did demonstrate the PTSD-dysphoria model to fit significantly worse than the PTSD-numbing model. Specially, the newest Dysphoric Arousal model proposed by Elhai et al. (2011) was the best fitting model.

In this newest Dysphoric Arousal model, symptoms were referred to as the dysphoric arousal (D1–D3) represent a separate construct from the Emotional Numbing model's arousal factor, and the Dysphoria model's dysphoria factor. In several cross-sectional investigations, the model was identified as the best fitting model (Elhai et al., 2011; Wang, Long, et al., 2011; Wang, Zhang, et al., 2011). The results of configural invariance tests were consistent with those limited cross-section investigations (Elhai et al., 2011; Wang, Long, et al., 2011; Wang, Zhang, et al., 2011) – namely, that the newest proposed Dysphoric Arousal model was the best fitting model. Additionally, the findings of measurement invariance testing did reveal similar results that the Dysphoric Arousal model fit significantly better than the Dysphoria models, as well as at the same level with the Emotional Numbing model. Furthermore, the findings of structural invariance testing demonstrated that the Dysphoric Arousal model fit significantly better than the Dysphoric models and Emotional Numbing model.

The pattern of factor loadings of the present study was similar with the results of Elhai et al. (2011) that PTSD's D4–D5 symptoms were referred to as a separate factor with higher factor loadings than the *DSM-IV* and PTSD-numbing models, as well as D1–D3 were treated as an independent factor with higher factor loadings than the other three models at two time points. This finding offers further support for the newest proposed model with longitudinal superiority. However, we made such a conclusion with caution because the fit indices of the 3 types of invariance involved only minute differences (ΔCFIs < .005) between them (Cheung & Rensvold, 1999, 2002). The differences between the factor loadings were subtle as well. Until now, few published studies have tested the longitudinal structural invariance of PTSD symptoms, and more investigations should be conducted to confirm or reject our conclusions.

The PTSD-numbing model specifies only 2 indicators of avoidance, as well as the PTSD-dysphoria model specifies another factor with only 2 indicators of hyper-arousal. In the same way, the Dysphoric Arousal model specifies only 2 indicators of anxious arousal and only 3 indicators of dysphoric arousal. Previous studies using cross-sectional samples could not confirm that the factorial solution is stable (e.g., Asmundson, Stapleton, & Taylor, 2004; Elhai et al., 2011; Palmieri, Weathers, et al., 2007). Ideally, an analysis would include 4–6 indicators of each putative factor to increase the confidence that the factorial domain is fully measured and the factorial solution is stable (Harvey, Billings, & Nilan, 1985; Hinkin, 1998). The findings in the current longitudinal study suggested that those factors with only 2 indicators were stable across time.

The findings of the present study also expand our knowledge about the longitudinal factorial invariance of PTSD symptoms, specifically in an Asian population. There is no determining distinction between physical and mental disorders, as psychological factors are also implicated in physical illnesses (Lin, 1982). Chinese people have dissimilar perceptions and symptom manifestations of mental illness and psychiatry (Cheung, 1995). Chinese patients may present with symptoms of the alleged physiological function of the "related" bodily organ, rather than with emotional states associated with the mental health condition. This study also lays a preliminary foundation for the generalizability and stability of latent structure of PTSD symptoms identified in the western culture.

This study has several limitations. The study relied on the use of a single self-reported measure of PTSD symptoms; the use of self-reported PTSD symptoms has been shown to yield a PTSD structure that differs slightly from that of findings obtained with interview data (Palmieri, Weathers, et al., 2007). The sample may not be representative of the general Asian population. We were unable to evaluate model fit across different trauma types, due to the limitations of sample size. Previous studies has revealed little evidence that different types of trauma might moderate the structure of PTSD for two four-factor models (e.g., Elhai et al., 2009; Yufik & Simms, 2010). Wang, Zhang, et al. (2011) found that the newly proposed five factor model was better than the four-factor model when examining data from both earthquake victims and violent riot witnesses. Due to less research published on the robustness of the five factor model in various trauma samples, further research with a variety of trauma populations is needed and will add to our understanding of the latent structure of PTSD, especially in Asian populations.

In summary, the present study promotes our understanding of the longitudinal factorial invariance of PTSD symptoms in young survivors after serious traumatic events. Our results show that the two 4-factor intercorrelated models and the newest proposed 5-factor model were superior to the model defined in *DSM-IV*, as well as Dysphoric Arousal model was better than the Dysphoric models and at the same level with Emotional Numbing model. The sufficiently good fit of the alternative models warrant their future investigation.

References

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

Armour, C., Elhai, J. D., Layne, C. M., Shevlin, M., Durakovic-Belko, E., Djapo, N., et al. (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

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, 10.1.007/s10960-004-5795-7.

Barnes, V. A., Treiber, F. A., & Ludwig, D. A. (2005). African-American adolescents' stress responses after the 9/11/01 terrorist attacks. *Journal of Adolescent Health*, *36*, 201–207. doi:10.1016/j.jadohealth.2004.02.035

Baschnagel, J. S., O'Connor, R. M., Colder, C. R., & Hawk, L. W. (2005). Factor structure of posttraumatic stress among western New York undergraduates following the September 11th terrorist attack on the World Trade Center. *Journal of Traumatic Stress*, *18*, 677–684. doi:10.1002/jts.20076

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

Calderoni, M. E., Alderman, E. M., Silver, E. J., & Bauman, L. J. (2006). The mental health impact of 9/11 on inner-city high school students 20 miles north of Ground Zero. *Journal of Adolescent Health*, *39*, 57–65. doi:10.1016/j.jadohealth.2005.08.012

Cheung, F. M. (1995). Facts and myths about somatization among the Chinese. In: T. Y. Lin, W. S. Tseng, & E. K. Yeh (Eds.), *Chinese societies and mental health* (pp. 156–166). Hong Kong: Oxford University Press.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*, *25*(1-27) doi:10.1177/014920639902500101

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/S15328007SEM0902_5

DuHamel, K. N., Ostroff, J., Ashman, T., Winkel, G., Mundy, E. A., Keane, T. M., et al. (2004). Construct validity of the posttraumatic stress disorder checklist in cancer survivors: analyses based on two samples. *Psychological Assessment*, *16*, 255–266. doi:10.1037/1040-3590.16.3.255

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*, 340–345. doi:10.1016/j.janxdis.2010.10.007

Elhai, J. D., Ford, J. D., Ruggiero, K. J., & Frueh, B. C. (2009). Diagnostic alterations for posttraumatic stress disorder: examining data from the National Comorbidity Survey replication and national survey of adolescents. *Psychological Medicine*, *39*, 1957–1966. doi:10.1017/S0033291709005819

Elhai, J. D., Gray, M. J., Docherty, A. R., Kashdan, T. B., & Kose, S. (2007). Structural validity of the Posttraumatic Stress Disorder Checklist among college students with a trauma history. *Journal of Interpersonal Violence*, *22*, 1471–1478. doi:10.1177/0886260507305569

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*, 849–854. doi:10.1016/j.janxdis.2011.04.007

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

Engdahl, R. M., Elhai, J. D., Richardson, J. D., & Frueh, B. C. (2011). Comparing posttraumatic stress disorder's symptom structure between deployed and non-deployed veterans. *Psychological Assessment*, *23*, 1–6. doi:10.1037/a0020045

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11, Suppl 3), S78–S94. doi:10.1097/01.mlr.0000245454.12228.8f

Grouzet, F. M. E., Otis, N., & Pelletier, L. G. (2006). Longitudinal cross-gender factorial invariance of the academic motivation scale. *Structural Equation Modeling*, *13*, 73–98. doi:10.1207/s15328007sem1301_4

Harvey, R. J., Billings, R. S., & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: good news and bad news. *Journal of Applied Psychology*, *70*, 461–468. doi:10.1037/0021-9010.70.3.461

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Method*, *1*, 104–121. doi:10.1177/109442819800100106

Hoyt, T., & Yeater, E. A. (2010). Comparison of posttraumatic stress disorder symptom structure models in Hispanic and White college students. *Psychological Trauma: Theory, Research, Practice, and Policy*, *2*, 19–30. doi:10.1037/a0018745

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:10.1037/104 0-3590.10.2.90

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

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*, 165–175, 10.1037/1040-3590.19.2.165.

Li, H., Wang, L., Shi, Z., Zhang, Y., Wu, K., & Liu, P. (2010). Diagnostic utility of the PTSD Checklist in detecting PTSD in Chinese earthquake victims. *Psychological Reports*, *107*, 733–739. doi:10.2466/03.15.20.PR0.107.6.733-739

Lin, K. M. (1982). Traditional Chinese medical health beliefs and their relevance for mental illness and psychiatry. In: A. Kleinman, & T. Lin (Eds.), *Normal and abnormal behavior in Chinese culture* (pp. 95–111). London: Reidel Publishing Company.

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

Marshal, G. N. (2004). Posttraumatic Stress Disorder Symptom Checklist: factor structure and English-Spanish measurement invariance. *Journal of Traumatic Stress*, *17*, 223–230. doi:10.1023/B:JOTS.0000029265.56982.86

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592. doi:10.1037/0021-9010.93.3.568

Meade, A. M., Lautenschlager, G. J., & Hecht, J. E. (2005). Establishing measurement equivalence and invariance in longitudinal data with item response theory. *International Journal of Testing*, *5*, 279–300. doi:10.1207/s15327574ijt0503_6

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11 (Suppl. 3)), S69–S77. doi:10.1097/01.mlr.0000245438.73837.89

Miles, J. N. V., Marshall, G. N., & Schell, T. L. (2008). Spanish and English versions of the PTSD Checklist-Civilian Version (PCL-C): testing for differential item functioning. *Journal of Traumatic Stress*, *21*, 369–376. doi:10.1002/jts.20349

Muthén, B. O. & Muthén, L. (2006). *Chi-square difference testing using the Satorra–Bentler scaled chi-square*. Available from: <http://statmodel.com/chidiff.shtml> Accessed 01.05.11.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Palmieri, P. A., Marshall, G. N., & Schell, T. L. (2007). Confirmatory factor analysis of posttraumatic stress symptoms in Cambodian refugees. *Journal of Traumatic Stress*, *20*, 207–216. doi:10.1002/jts.20196

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

Pitts, S. C., West, S. G., & Tein, J. Y. (1996). Longitudinal measurement models in evaluation research: examining stability and change. *Evaluation and Program Planning*, *19*, 333–350. doi:10.1016/S0149-7189(96)00027-4

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., 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, 10.1037/0021-843X.111.4.637.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. doi:10.1177/109442810031002

Wang, M., Dai, X., & Wan, J. (2009). Factor structure of PTSD Checklist: a confirmatory factor analysis study in adolescents from earthquake region. *Chinese Journal of Clinical Psychology*, *17*, 420–423 (in Chinese).

Wang, L., Long, D., Li, Z., & Armour, C. (2011). Posttraumatic stress disorder symptom structure in Chinese adolescents exposed to a deadly earthquake. *Journal of Abnormal Child Psychology*, *34*, 749–758. doi:10.1007/s10802-011-9508-4

Wang, L., Zhang, J., Shi, Z., Zhou, M., Li, Z., Zhang, K., et al. (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*, 771–776. doi:10.1016/j.janxdis.2011.03.011

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993). The PTSD Checklist: reliability, validity, and diagnostic utility. Paper presented at the Annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

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*, 596–606. doi:10.1002/da.20837

Wu, K. K., Chan, S. K., & Yiu, V. F. (2008). Psychometric properties and confirmatory factor analysis of the posttraumatic stress disorder checklist for Chinese survivors of road traffic accidents. *Hong Kong Journal of Psychiatry*, *18*, 144–151. Available from: <http://hkjpsych.com/journal_file/0804_V18N4-p 144-51.pdf> Accessed 01.02.11

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