HHS Public Access

Author manuscript

Psychol Assess. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Psychol Assess. 2015 December ; 27(4): 1160–1170. doi:10.1037/pas0000115.

Posttraumatic stress disorder among refugees: Measurement invariance of Harvard Trauma Questionnaire scores across global regions and response patterns

Andrew Rasmussen, PhD[1],
Department of Psychology; Fordham University, Bronx, NY

Jay Verkuilen,
Program in Educational Psychology, Center for the Advanced Study in Education; City University of New York, New York, NY

Emily Ho, and
Department of Psychology; Fordham University, Bronx, NY

Yuyu Fan
Department of Psychology; Fordham University, Bronx, NY

Abstract

Despite the central role of posttraumatic stress disorder (PTSD) in international humanitarian aid work, there has been little examination of the measurement invariance of PTSD measures across culturally defined refugee subgroups. This leaves mental health workers in disaster settings with little to support inferences made using the results of standard clinical assessment tools, such as the severity of symptoms and prevalence rates. We examined measurement invariance in scores from the most widely used PTSD measure in refugee populations, the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), in a multinational and multilingual sample of asylum seekers from 81 countries of origin in 11 global regions. Clustering HTQ responses to justify grouping regional groups by response patterns resulted in three groups for testing measurement invariance: West Africans, Himalayans, and all others. Comparing log-likelihood ratios showed that while configural invariance seemed to hold, metric and scalar invariance did not. These findings call into question the common practice of using standard cut-off scores on PTSD measures across culturally dissimilar refugee populations. In addition, high correlation between factors suggests that the construct validity of scores from North American and European measures of PTSD may not hold globally.

Keywords

posttraumatic stress disorder; culture; refugees; measurement invariance; Harvard Trauma Questionnaire

[1]Correspondence should be directed to the first author's attention via mail at Fordham University, 441 E. Fordham Rd., Dealy 226, Bronx, NY 10458 or via electronic mail at arasmussen@fordham.edu.

Introduction

The globalization of trauma psychology through international disaster relief and humanitarian aid efforts has resulted in mental health professionals using assessments of posttraumatic stress disorder (PTSD) in settings far afield from the cultural contexts in which they were developed. Although the experience of intense emotional distress following traumatic events is likely to be universal, there is evidence to suggest that the expression of that distress is subject to substantial cultural variation (Hinton & Kirmayer, 2013; Marsella, Friedman, & Spain, 1996; Rasmussen, Keatley, & Joscelyn, 2014). And yet, assessments of PTSD that follow the construct as it appears in North American and European nosology are applied widely in non-Western samples with little critique. The past 20 years have seen numerous studies in which individuals within refugee populations endorse PTSD symptoms on questionnaires (de Jong et al., 2001; Fox, & Tang, 2000; Ichikawa, Nakahara, & Wakai, 2006; Sachs, Rosenfeld, Lhewa, Rasmussen, & Keller, 2008; Shrestha et al., 1998) and structured clinical interviews (Neuner, Schauer, Klaschik, Karunakara, & Elbert, 2004; Rasmussen, Rosenfeld, Reeves., & Keller, 2007), and these responses are often positively correlated with the number of potentially traumatic events (PTEs) that they report (Cardozo, Vergara, Agani, & Gotway, 2000; Fawzi et al., 1997; Marshall, Schell, Elliott, Berthold, & Chun, 2005; Mollica et al., 1999). It is often assumed that scores from these assessments thus have comparable meaning, despite the radically different populations in which they are used. Due to a paucity of studies in the literature on these measures concerning cross-cultural validity (van Ommeren, 2003), it is not known whether this application of PTSD scores across heterogeneous populations leads to reasonable inferences concerning symptom severity and diagnoses.

Assessing the comparability of scores is a particularly important step in the adaptation of psychological measures across cultures, perhaps most clearly outlined by Geisinger (1994; also see Hambleton, Merenda, & Spielberger 2006). Of particular importance to Geisinger (1994) are two broad issues (which are relevant to several steps): Having culturally faithful versions of the instrument and ensuring scores across populations maintain psychometric properties such as adequate reliability and construct validity. To date most PTSD measures used in refugee and asylum seeking populations have been translated and back-translated into relevant languages and have established adequate internal reliability (Cronbach's alpha). Although a few studies have examined the construct validity of responses from these measures in as much as they have fit confirmatory factor analyses (CFA) to test configural invariance of the predominant North American and European models of PTSD, none have compared construct validity across populations within the same sample. One of the key psychometric techniques to identify violations of construct validity is to examine all aspects of measurement invariance.

Psychometric and clinical significance of measurement invariance

Measurement invariance is central to the validity of quantitative measures. Measurement invariance, also known as measurement equivalence, is a statistical property that gauges the degree to which responses to a survey or questionnaire are similarly related to latent variables across different conditions or populations. Thus, measurement invariance is

necessary to support inferences based on scale scores across multiple groups (Millsap, 2011). Measurement invariance is usually modeled using CFA, where factors represent latent variables and factor loadings represent item scores' contribution to those latent variables. The extensive literature on measurement invariance defines three main types: configural, metric, and scalar. Each type of invariance adds further constraints to the previous type, and thus represents a set of nested models.

Configural invariance, the least restrictive form of invariance, requires that measured symptoms have the same dimensional structure across groups. In a factor analytic approach, item scores must load onto the same factors across groups within a given sample, although the size of loadings may differ in magnitude. Concern about whether or not PTSD "looks the same" across cultural groups has given rise to a CFA literature that directly addresses configural invariance in a variety of refugee samples, and in general responses are associated with one another in ways that are similar to associations seen in European and North American samples. The most common structure is a four-factor model consisting of reexperiencing, avoidance, numbing, and hyperarousal symptoms (Palmieri, Marshall, & Schell, 2007; Rasmussen, Smith, & Keller, 2007; Vinson & Chang, 2012). This four-factor model (4F model) has been interpreted using terms from the *Diagnostic and Statistical Manual, Fourth Edition* (DSM-IV; APA), 1994), and was the model for changes appearing in the current DSM-5 (APA, 2013). That evidence for configural invariance of the 4F model has been found in multiple culturally defined samples suggests that the symptoms that define *reexperiencing*, *avoidance*, *negative cognitions and moods*, and *hyperarousal* are the same across culturally-defined samples, though the relative contribution of symptoms to each symptom cluster may vary. Because the contributions of the item scores may vary, configural invariance does not support strong interpretations about individuals or groups, but only general inferences concerning content validity.

Metric invariance presumes configural invariance but further requires that the strength of relationships between item scores and latent variables is consistent across groups, i.e., that the loadings of items on factors across groups are equal. In cross-cultural PTSD research, this would mean that the relative contribution of specific symptoms to symptom clusters (e.g., intrusive imagery's association with the latent re-experiencing variable) would be uniform across cultural groups. Metric invariance can support comparisons of change in scores across groups over time, though not the comparisons of the level or magnitude of scores.

The third and strongest type of invariance that is testable given the data researchers usually collect is called *scalar invariance*. Scalar invariance requires that configural and metric invariance hold and adds that the relationship between the items scores and latent variables also agree in overall level – i.e., the level of item endorsement and scale scores, which on clinical scales represents symptom severity. Scalar invariance is a necessary characteristic for most of the practical uses of assessment tools, from clinical inferences about individuals' diagnoses and the burden of disease within a population. In humanitarian aid practice, scalar invariance is necessary to support inferences comparing the prevalence of PTSD between different culturally-defined groups and supporting the use of specific scores to identify

clinical cases – e.g., that the cut-off score of 2.5 on the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992) indicates probable PTSD across groups.

Lack of measurement equivalence across cultures is well-documented in areas of psychological inquiry outside of clinical assessment (Henrich, Heine, & Norenzayan, 2010; Steenkamp & Baumgartner, 1998). In general this literature supports the conclusion that differing response patterns and therefore a lack of scalar invariance are problematic for comparing scores from psychological assessments across culturally defined populations. Findings from the clinical literature suggests that similar issues may pose particular problems in the clinical assessment of PTSD includes widely varying PTSD scores across post-conflict settings (de Jong et al., 2001), extremely low scores among Tibetan refugees (Lhewa, Banu, Rosenfeld, & Keller, 2007; Sachs, Rosenfeld, Lhewa, Rasmussen, & Keller, 2008), high scores among Latino combat veterans within the United States (Pole, Best, Metzler, & Marmar, 2005), and severity differences between Mexican and U.S. hurricane survivors (Norris, Perilla, & Murphy, 2001). To date there has been no direct empirical test of scalar invariance in PTSD scores among refugees or asylum seekers. Beyond the configural invariant baseline, studies have yet to compare measurement invariance of PTSD models across culturally defined groups within a single sample.

The Current Study

The current study examines the cross-cultural measurement invariance of HTQ (Mollica, et al., 1992) scores in a diverse sample of treatment seeking asylum seekers in [removed to permit masked review]. The HTQ is the most frequently used measure of PTSD in refugee and asylee populations around the world. The 16-item PTSD section of the HTQ has been used with mulitple groups in multiple war-affected settings (e.g., the former Yugoslavia; Mollica, et al., 1999) and has the most robust findings for internal and test-retest reliability in the refugee literature (Hollifield et al., 2002).

There were two main parts to the current study. First, we attempted to determine the possible regional or cultural groupings that were in the data in terms of response patterns. In any study of invariance it is necessary to determine what groups to compare. In a typical invariance study there are clearly identifiable subgroups, such as males or females, or groups of individuals responding to three different assessment forms. These groups are usually stated *a priori* (e.g., by policy concerns). Cross-culturally, however, the number of possible groups is potentially quite large. From the literature, we expected East Asian participants to be distinct in terms of response style, implying a violation of scalar invariance. From work with West African populations (Rasmussen, Smith, et al., 2007) we suspected that their particular response patterns might also indicate configural differences. We had no particular hypotheses about additional groups, and thus our analysis at this point is best termed *semi-confirmatory*. In order to provide further guidance in determining the number of groups to be compared without overly relying on a specific model, we used nonparametric methods to group individuals by similarity of symptom profiles based on response patterns across region of origin. Specifically, we used K-means cluster analysis (Lattin, Carroll, & Green, 2003), a nonparametric categorization method that makes relatively weak assumptions about the data to group observations by a set of variables (here PTSD symptoms). We compared

cluster membership across regions in order to examine whether regional groupings represented distinct response profiles. We then grouped participants based on clusters, i.e. region of origin consistent with shared response patterns. Since disparate cultures may share response styles, we felt justified in allowing otherwise dissimilar cultures to be grouped together empirically for this study.

Second, we considered the measurement invariance of the 4F model using the groups identified in the first step. We used multigroup CFA to test configural, metric, and scalar invariance of HTQ scores. Because McDonald and Ho (2002) note that it is possible for goodness-of-fit statistics to look quite reasonable while important model parameters fit poorly and Chen (2007) notes that goodness-of-fit statistics are frequently insensitive to violations of invariance when groups are of unequal size, we relied on the corrected likelihood-ratio chi square tests (LR-X^2) to statistically compare the models.

Methods

Sample

Participants were 878 survivors of torture and other human rights abuses who completed an intake assessment as part of treatment at a clinic specializing in the medical and psychosocial care of refugees and asylum seekers. Participants were accepted to the clinic after being positively identified as survivors of torture based upon criteria set by the United Nations Convention against Torture (United Nations, 1984). The semi-structured intake interview was designed to elicit a detailed trauma narrative, including the number and types of PTEs (up to five persecution events), medical and psychological treatment history, demographic information, and standardized clinical assessments that included the PTSD section of the HTQ. We grouped reported PTE types into 22 categories according to guidelines provided by Human Rights Documents International (HURIDOCS; Dueck & Aida, 1993), an international system used to document human rights abuses. The data for the current study was drawn from a five-year period. The use of this archival data for secondary analysis was approved by the Institutional Review Board of the New York University School of Medicine (where the first author was employed at the time of retrieval).

Of 878 cases accepted to the clinic, 518 (59%) were male; 328 (37%) were Muslim, 293 (33%) Christian, 196 (22%) Buddhist, 22 (3%) endorsed other religions and 13 (2%) were unaffiliated (26, 3% were missing information on religion). The largest of the 11 represented global regions were West Africa (n = 307, 35%), Himalayan Asia (n = 188, 21%), and Central Africa (n = 122, 14%). The intersection of gender, religion, and region are presented in Table 1. Countries represented in the sample are presented in Supplemental Table 1. Mean age at interview was 34.90 years (SD = 9.92).

Harvard Trauma Questionnaire

The HTQ (Mollica, et al., 1992) is comprised of three sections: a list of PTE types, a 16-item symptom list that corresponds to the 17 symptoms of PTSD in the DSM-IV, and a supplemental symptom section designed to change according to the culturally-based expressions of distress within the population of interest. In order to make comparisons

across respondents, the clinic from which the data were drawn and the current study utilized the 16-item PTSD section alone (items appear in Table 5, below). The HTQ uses a four-point relative severity response scale. Respondents endorse how much each symptom has bothered them in the past week: *not at all*, *a little bit*, *quite a bit*, or *extremely*. The HTQ total score is an average score, with 2.5 suggested as the clinical cut-off score indicating that a respondent has a high likelihood of PTSD (Mollica et al., 1992). In addition to the English original, translated standard versions were available for administration in French and Spanish (the survey had been translated and back-translated by clinic interpreters and French and Spanish speaking staff and pilot and field tested with good reliability; see Hooberman, Rosenfeld, Rasmussen & Keller, 2010; Rasmussen et al., 2007), Tibetan (Lhewa, et al., 2007), Arabic (Shoeb, Weinstein, & Mollica, 2007), and Cambodian (Mollica, et al., 1992). The HTQ was administered using English or one of these standard versions in 725 cases (83%; there were 48 cases with missing data for language of administration): English (n = 322, 36.7%), French (n = 234, 26.7%), Tibetan (n = 148, 16.9%), Spanish (n = 13, 1.5%), and Arabic (n = 8, 0.9%; no HTQs were administered using Cambodian). Other language needs (n = 105, 12%) were met by professional health interpreters trained in working with the population and in interpreting the English-language HTQ. For full information on the use of versions of the HTQ by country of origin, see Supplemental Table 1.

Procedures

Exploratory analysis to classify participants by regions—In order to define comparison subsamples, we took an iterative, bottom-up approach to classifying individuals. We began by classifying participants by country and then grouped contiguous countries with small sample sizes into regions according to HURIDOCS country codes, which provide regional classification based on cultural and historical information within codes. We then grouped contiguous regions with small sample sizes into 11 larger regions. Countries within regional groups are presented in Supplemental Table 1.

Following regional classification, we examined univariate statistics. We generated a region by religion by gender matrix in order to examine dependence between the three. We examined the associations between the number of PTE types reported and HTQ scores, and specific PTE types reported by more than 5% of the sample and HTQ scores. To consider whether there might be systematic differences in reported PTEs by region of origin and gender, we ran a linear regression with region and gender interacted on the total number of reported PTEs. To consider whether there might be systematic differences in HTQ responses due to administrative (i.e., standard translation) differences, we examined whether the availability of standard versions of the HTQ was associated with total HTQ scores. In order to examine a potential positive linear association between number of PTEs and symptom severity, we modeled regressions predicting symptom severity using grand mean-centered PTEs within global regions.

Exploratory analysis to classify item response profiles—To consider the relationship between the HTQ items and regional classifications in a parsimonious way, we made use of K-means cluster analysis of raw item scores (using R's "kmeans" module; R Development Core Team, 2008). K-means clustering requires users to select the number of

groups *a priori*. The algorithm finds homogeneous groups based on separating the means within group, with cases with some missing data being assigned to a separate cluster. We tried several different values of K and used random initial starts to protect against local optima (Steinley, 2006). In general, we preferred to have too many clusters rather than too few. To provide a rough interpretation of the clusters, we considered their relationship to 4F model averages, *reexperiencing*, *avoidance*, *numbing*, and *hyperarousal*. We then examined these clusters across regional groups using cross-tabulation to determine whether and how regional groups might be associated with response patterns. Regional groups with like response patterns were then grouped together.

Confirmatory analysis to examine measurement invariance—We considered measurement invariance for the three global regional response pattern groups identified by our classification analysis (Himalayan, West African, and Other) identified in the data previously. Mplus 7.11 (Muthén & Muthén, 2013) was used to fit the models using full information maximum likelihood with the Satorra-Bentler correction. In order to avoid the problems associated with relying on aggregate fist statistics in judging violations of invariance across groups of unequal size (Chen, 2007; McDonald & Ho, 2002), we relied on corrected LR X^2 tests to statistically compare the three models.

Results

Exploratory results

The 11 × 6 × 2 classification system of region by religion by gender matrix resulted in a total of 132 possible combinations (presented in Table 1). Stratifying by gender, each table for region × religion was strongly dependent (Cramér's V = 0.56 for males, 0.51 for females), with the dependence being consistent with global patterns (e.g., Himalayans were very likely to be Buddhist, West Africans to be Muslim). Given that religion and region were so strongly associated and that region is more likely to be culturally consistent than religion, we focused subsequent analyses primarily on region.

During the intake process 798 of the 878 (91%) respondents provided reliable information on number of PTEs. The mean number of reported PTEs was 2.87 (SD = 1.52). A large majority of the 798 respondents reported beatings (n = 673, 87%); other reports included threats of death or injury by authorities (n = 295, 41%), forced performance of degrading behavior (n = 22, 18%), prolonged deprivation of food, water, or sensory stimuli (n = 129, 15%), rape (n = 119, 17%), and immobilization (e.g., with ropes; n = 54, 8%). All other types were reported by less than 5% of the sample. Total number of PTEs was weakly but significantly associated with HTQ scores (r = .17, p < .001). Only one major PTE type was associated with HTQ scores – rape. Those participants who reported rape had higher HTQ scores (M = 2.82, SD = 0.57) than those who did not (M = 2.47, SD = 0.65; t(184.35 df) = 5.85, p < .001). The linear regression with region and gender interacted on the total number of reported PTEs is presented as Supplemental Table 2. The R^2 for this regression was quite modest, 0.08. No region-by-gender groups were statistically different from others.

K-means cluster analysis of items suggested K = 8, which generated 8 clusters. Table 2 presents subscales scores for each cluster. Cluster 1 consisted of low *reexperiencing*,

numbing, *avoidance* and *hyperarousal* subscores. Clusters 2 and 5 had higher *reexperiencing*, *avoidance* and *hyperarousal* scores but relatively low (below 2.5) *numbing* subscores. Cluster 3 was uniformly high on all subscores, and cluster 4 was similar to cluster 3 but with lower (although still severe) *numbing* subscores. Clusters 6 and 7 were both intermediate clusters, with the difference being that cluster 6 had somewhat lower subscores (on average half a point lower). Cluster 8 grouped all cases with missing data together and was thus a residual category; cluster 8's average subscores were very similar to cluster 2. Varying the number of clusters by 1–2 groups did not substantially affect the results.

Table 3 presents the 11 regional subsamples by cluster membership. Clusters and regional subsamples were strongly dependent (LR-X^2 = 128.43, 70 *df*, $p < 0.001$). To determine the nature of this dependence, we examined adjusted residuals, the standardized values that indicated misfit compared to the independence model. The main source of large residuals was the Himalayan group. Most notably the Himalayan group was over-represented in clusters with lower subscale scores (most notably in clusters 1 and 6). To further examine this dependence we excluded the Himalayan group. The subsequent model fit independence (LR-X^2 = 77.93, 63 *df*, $p = 0.098$); however, large adjusted residuals remained for the West African group (as suggested by the statistical trend towards dependence). West Africans were overrepresented in lower reporting clusters (clusters 1, 6 and 7; they were also overrepresented in one higher reporting cluster, cluster 5). The primary distinction in these data, therefore, appeared to be between Himalayan participants, West African participants, and all others (Other). Alpha reliability for HTQ total scores within Himalayan, West African, and Other subgroups was high (α = .89, .89, .86, respectively).

Covariation by group

The availability of standard versions of the HTQ (i.e., use of the English, French, Spanish, Tibetan or Arabic standard versions) was not associated with total HTQ scores (t(828 *df*) = 0.602, $p = .55$), indicating that the observed HTQ response patterns were not due to differences in survey administration. HTQ administration did differ across global regional response pattern groups, with Others' administration less likely to have used a standard version (n = 292, 81%) than both Himalayans (n = 172, 97%) and West Africans (n = 261, 89%; χ^2 (2 *df*) = 27.64, $p < .001$). However, due to these differences not resulting in HTQ mean differences, this variable was not examined further. Rape was associated with global regional response pattern group, with those in the Other category reporting higher rates (n = 87, 27%) than Himalayans (n = 4, 2%) and West Africans (n = 28, 13%; χ^2 (2 *df*) = 50.30, $p < .001$); of note, there were substantial missing data (n = 171) for analyses examining rape.

For subscales representing the 4F model of PTSD, internal reliability in the full sample was variable, with alphas for *reexperiencing*, *numbing*, *hyperarousal* being adequate (α = .74, .71, .76, respectively), and for *avoidance* being marginal (α = .63). This pattern was somewhat different across the three subgroups: scores from Himalayans, West Africans, and Others had adequate internal reliability concerning *re-experiencing* (α = .68, .76, .73, respectively), *numbing* (α = .74, .69, .67, respectively), and hyperarousal (α = .78, .75, .71, respectively); for *avoidance*, Himalayans and West Africans' scores were reliable (α = .74, and .70, respectively), but Others' scores were not (α = .48).

For Himalayans, PTEs significantly predicted HTQ scores, β = 0.27, $t(160) = 3.62$, $p < .001$; $R^2 = .08$, $F(1,160) = 13.08$, $p <. 001$. For West Africans, PTEs did not significantly predict HTQ scores, β = .07, $t(253) = 1.10$, $p = .27$. For Others, PTEs significantly predicted HTQ scores, β = 0.13, $t(370) = 2.56$, $p = .011$, but very little of the variance, $R^2 = .02$, $F(1,370) = 6.54$, $p = .01$. Intercepts were roughly equal for Himalayans (β = 2.20), West Africans (β = 2.50), and Others (β = 2.70).

Measurement invariance

CFA statistics for configural, metric, and scalar invariance models are presented in Table 4. The models converged to proper solutions in all cases. Testing against the saturated model indicated that the configural model did not fit relative to the saturated model (corrected LR-X^2 = 369.58 on 294 *df*, *p*-value = 0.0018). This is unsurprising as it is commonly found in practice. However, root-mean-square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), and the standardized root-mean-residual (SRMR) all indicated that these models were reasonable according to standard guidelines (e.g., McDonald & Ho, 2002). Thus, most of the misfit seen in the covariance residuals of the configural model was limited to few indicators. We provide goodness-of-fit statistics for reference in Table 4, but caution that although these statistics appear to show reasonable fit for each model, due to the problems associated with using fit statistics for invariance testing reported in the Introduction and Methods sections, they should not be interpreted too strongly. We used the corrected LR X^2 tests to statistically compare the three models. The residuals from the configural model are presented in Supplemental Table 3 for further reference. Examining these residuals resulted in three notable observations: (1) the Himalayan group's model did not fit as well as the others, due to a mild floor effect in their indicators; (2) the other two fit well; (3) these differences in fit were not particular to specific symptom clusters.

Using the LR X^2 of the metric-to-configural model suggested that metric invariance did not appear to hold (corrected LR X^2 = 48.19 on 24 *df*, *p*-value = 0.0024). To interpret, we considered the parameter estimates from the least constrained configural model, presented in Table 5. As evident from an inspection of the factor loadings, several were different across groups. Three differed by .200 or more between at least two groups: one *re-experiencing* item, one *avoidance* item, and one *hyperarousal* item (in Table 5: R4/5, A2, and H2, respectively). Most notably, factor loadings comprising the *numbing* factor for West Africans were systematically larger (directly related to the very large correlations between several factors within this regional group; see Table 6). (See also below regarding the factor correlations.)

Unsurprisingly the scalar model is also rejected (corrected LR X^2 = 117.95 on 48 df, p-value < 0.001). This was apparent not only in the LR X^2 test, but also in the diversity of intercept values (presented in Table 5). Some indicators had similar intercepts and others were quite different. In particular, one *re-experiencing* item and one *hyperarousal* item (in Table 5: R3 and H1) had intercepts that differed by half a point or more on the HTQ across groups. As values were averaged across large numbers of participants on the four-point HTQ response scale, differences of half a point or more were deemed clinically significant. Intercept patterns were systematic across groups, indicating that differences within subsamples did

not offset one another. Total score mean box plots are presented in Figure 1; group distributions are presented relative to the 2.5 cut-off score used to suggest probable PTSD in many studies.

To check that the lack of metric and scalar invariance was not due to a potential relationship between higher incidence of rape and membership in the Other group, we reran the CFAs on the subset of respondents who reported that they were not raped. A similar pattern of results emerged. According to hypothesis testing, metric invariance held reasonably well but scalar invariance did not; however, due to missing data, cell sizes were rather small. Model output is presented in Supplemental Table 4.

Correlations among latent variables in each group (Table 6) were extraordinarily high, particularly for the West African group. The extremely high correlation between the *numbing* and *hyperarousal* factors suggested that these two were close to collapsing into one another, comprising a statistically improper solution (Dillon, Kumar, & Mulani, 1987). Correlations were lowest for the *avoidance* factor, *r* ranging from .50 – .75. This pattern of very high factor correlations was also observed for supplementary analyses using only those subjects who did not report raped, lending further credence to differential symptom manifestation across cultural groups. It also suggested that the 4F model needs to be treated with caution for these data. For this reason we did not pursue modeling partial invariance, where stronger invariance can be shown to hold for some set of items but not others.

Discussion

Clinical implications of a lack of invariance

At the individual level, assessments such as the HTQ help clinicians triage patients, target symptoms, and track treatment outcomes. At the group level, assessments provide information about the prevalence of disorders, subpopulations that need treatment resources, therapeutic modalities that are more effective, and mental health information about patient populations in general. If a PTSD measure is to be used for any of these purposes with individuals from different culturally defined populations, it is essential that scores from it have cross-cultural construct validity: they must be configurally invariant, metric invariant, and if used to compare populations with respect to PTSD phenomenology, also scalar invariant. To date, the HTQ has been validated in different cultures only by examining basic psychometric properties (i.e., the first three or four steps in Geisinger, 1997). Ours is the first study we know of that has examined its metric and scalar invariance.

Based on differences in likelihood ratio chi-square tests (Chen, 2007; McDonald & Ho, 2002), our findings for HTQ scores were that configural invariance appeared to hold, but metric and scalar invariance did not. In other words, consistent with other literature (Palmieri et al., 2007; Rasmussen et al., 2007; Vinson & Chang, 2012) the basic content validity of PTSD as represented by the HTQ appears reasonable, but substantial differences in the contribution of specific symptoms to symptom dimensions and baseline intercepts across groups threaten the validity of cross-cultural comparisons. These differences were not attributable to specific items, systematic differences between groups in number, types of traumatic events, or differences in administration (i.e., interpreted or using standard

versions), suggesting a closer examination of the assessment is needed before the 16-item portion of the HTQ is used for extensive cross-cultural comparisons. These findings demand attention, calling into question using the HTQ to compare the reported level of trauma severity across different cultural groups, particularly the use of the commonly cited 2.5 clinical cut-off score for probable PTSD. The lack of scalar invariance suggests that using a single cut-off score is simply not a valid procedure for cross-cultural samples. At this time, we recommend that the HTQ should only be used to compare severity of PTSD symptoms across populations from different cultures with strong caution, and only in cases where such comparison is absolutely necessary.

Findings as they relate to the literature

Although the PTSD literature does include discussion of response style as it relates to inaccurate responding (i.e., malingering; e.g., Morel, 1998), culturally defined response style has largely been ignored. However, configural invariance with large differences in response style found in the current study is consistent with the small body of work examining PTSD factor structure among non-European origin populations (e.g., (Palmieri, et al., 2007; Rasmussen, Smith, et al., 2007) and, although not specific to PTSD, the larger literature on culturally defined response style (Byrne & Campbell, 1999; Heine, Lehman, Peng, & Greenholtz, 2002; Smith, 2004). Lower intercepts among Tibetans in the current study are generally consistent with a tendency to suppress affect among East Asians in general (Iwata, Roberts, & Kawakami, 1995; Noh, Kaspar, & Chen, 1998), and low HTQ scores for Tibetan asylum seekers in particular reported elsewhere (Lhewa, et al., 2007; Sachs et al., 2008) is evidence of a one-sided extreme response style towards the low or mild end of the scale. Although quite different culturally, similar patterns of response have been observed in scores from depression measures among Koreans (Cho & Kim, 1998), Japanese (Iwata & Roberts, 1996), and Chinese respondents (Li & Hsiao-Rei Hicks, 2009; Lin, 1989). Response style differences throughout the (non-PTSD) clinical literature suggests that scalar invariance is not just a problem for the HTQ or PTSD scales in general, but perhaps most clinical diagnoses relying on item scores.

Although not as stark as the lack of scalar invariance, the lack of metric invariance uncovered in the current study is also troubling. For the Himalayan and West African groups, three factors out of the 4F model showed differential item loadings, suggesting that the HTQ PTSD section of the assessment may not fully encapsulate all post-traumatic symptoms that may be manifest in different culturally defined populations. This may imply that specific items need to be adapted, or may imply that the construct of PTSD is not the best representation of posttraumatic psychopathology across different cultures. There is a small but growing literature defining posttraumatic responses from culturally emic (i.e., cultural insiders') perspectives, from Khmer *baksbat* in Cambodia (Chimm, 2012) to Mandinka *kidja faro* in Gambia (Fox, 2003), Rwandan *ihakamuka* (Hagengimana & Hinton, 2009), and Masalit *hozun* and *majnun* in the Darfur region of Sudan (Rasmussen, Katoni, Keller, & Wilkinson, 2011). There have even been attempts to measure such locally-relevant expressions and compare their measurement to PTSD measures (Jayawickreme, Jayawickreme, Atanasov, Goonasekera, & Foa, 2012). Notably, the HTQ itself was originally constructed to have an emic section to be constructed from ethnographic research

prior to using it in a setting that supplemented the 16 items measuring DSM PTSD. Although Mollica's original study (Mollica, et al., 1992) described the development of the ethnographically-derived section for use in Cambodian refugee camps, most studies using the HTQ since then have either applied the Cambodian-specific section to non-Cambodian groups or ignored the emic piece altogether in order to compare disparate populations.

The high correlation between latent variables observed in the current study suggests that internal inconsistency within factors and response style differences are not the only concerns related to measurement invariance cross-culturally. Though CFA models have become common in the literature on PTSD, high correlation between factors calls into question the phenomenological distinctiveness of the factors. Factor correlations throughout the PTSD literature are also quite high. Yufik & Simms (2010) found mean factor correlations around 0.80 in their meta-analysis of 40 studies, only slightly lower than what was found in the current study. Even if the factors are theoretically reasonable, sixteen indicators, or even DSM-5's 20 PTSD indicators, may not be enough to measure four factors well. It may be that other models better account for variance in the experience of generalized trauma-related distress. The generalized bifactor model (Chen, West, & Sousa, 2006), in which one "common distress" factor is supplemented by factors particular to symptom classes, might be a more appropriate approach to conceptualizing and measuring posttraumatic distress than the standard four-factor model or other standard latent variable models that exist in the PTSD literature (e.g., the four-factor dysphoria model proposed by Simms, Watson, & Doebbelling, 2002). Cohen and Bolt (2005) also note that manifest groups may not capture the problem of invariance violations and note that a mixture approach may find different groupings. The advantages and disadvantages of these approaches are a matter for further research.

Limitations

This study has a number of limitations. First and foremost, our regional response pattern groups are not likely good proxies for culturally specific subsamples. A critique can easily be made that in balancing regional grouping with subsample size we were too concerned with the latter. Although we made attempts to avoid confirmation bias, to avoid grouping regions by available sample sizes, and to monitor configural invariance at each step of our classification process, it is certainly possible that we began with too few participants from some regions to capture meaningful variance. In other words, in spite of our attempts to guard against groups comprising radically different response patterns, and in spite of our reasonable assumption that disparate cultural groups might very well have similar response styles, clearly grouping Central Africans with Latin Americans and the seven other regional groups might mask considerable cultural hetereogeneity in HTQ responses. Absence of evidence is not evidence of absence. We in no way wish to suggest that these groups are all culturally "the same" in ways other than their response patterns on the HTQ.

Other noteworthy limitations concern translation and administration: language of assessment differed across groups and for a sizeable minority of cases the HTQ was interpreted during administration. Given that standard versions were available for large majorities of each regional group and that mean HTQ scores did not differ by whether or not these forms were

available we conclude that administration procedures had no large effects on our findings related to measurement invariance. However, we acknowledge that more standardization would likely have strengthened findings. Although it is difficult to conceive of cross-cultural research with refugees done using measures in a single language—i.e., avoiding translation altogether—in the future researchers should avoid interpreted versions for the sake of the clarity of findings.

Although we believe that our findings speak to cultural differences as they relate to response style, we urge caution in applying them globally. Our findings represent analyses of the responses of treatment-seeking torture survivors to HTQ items at one clinic in the United States. Suffice it to say, the participants are not a random sample of asylum seekers or of refugees in general, let alone representative of cultures represented within regions. Indeed, the largest refugee populations today are from Central Asia and the Middle East, and our sample included few participants from either region. That aside, we believe that our data represent the most diverse dataset in which PTSD measurement invariance has been explored to date. Further research using similarly diverse datasets is necessary. Finally, using only one measure across multiple populations fails to tap into culturally bound interpretations of distress (referred to above) that may be much more relevant to the phenomenology of posttraumatic experiences than is represented on standardized measures of PTSD.

Conclusions

The current study examined the HTQ's measurement invariance, a necessary (though not sufficient) element of any measure's construct validity. For configural aspects, HTQ scores appear invariant; however, metric and scalar invariance did not hold. This points to potential differences in symptomology across different cultures and global regions that the HTQ may not adequately capture. Culturally differing response styles, well known in some subfields of psychology, must become part and parcel of cross-cultural clinical assessment in disaster and post-conflict psychology as well. Relief resources are too few and humanitarian aid efforts too important to ignore them.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project was supported in part by Award Number K23HD059075 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NIH/NICHD), awarded to the first author. The authors would like to thank Howard T. Everson, Ph.D. of the City University of New York for his comments on a revision of this manuscript.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed.. Washington, DC: American Psychiatric Association; 1994. *text rev*.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed.. Washington, DC: American Psychiatric Association; 2013.

Byrne B, Campbell TL. Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. Journal of Cross-Cultural Psychology. 1999; 30:555–574.

Cardozo BL, Vergara A, Agani F, Gotway CA. Mental Health, Social Functioning, and Attitudes of Kosovar Albanians Following the War in Kosovo. JAMA: Journal of the American Medical Association. 2000; 284(5):569–577. [PubMed: 10918702]

Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modelling. 2007; 14:464–504.

Chen FF, West SG, Sousa KH. A comparison of bifactor and second-order models of quality of life. Multivariate Behavioral Research. 2006; 41(2):189–225.

Chimm S. Baksbat (broken courgage): A trauma-based cultural syndrome in Cambodia. Medical Anthropology: Cross-Cultural Studies in Health and Illness. 2012; 32(2):160–173.

Cho MJ, Kim KH. Use of the Center for Epidemiologic Studies Depression (CESD) Scale in Korea. Journal of Nervous and Mental Disease. 1998; 186:304–310. [PubMed: 9612448]

de Jong JTVM, Komproe IH, van Ommeren M, El Masri M, Araya M, Khaled N, Somasundaram D. Lifetime events and posttraumatic stress disorder in four postconflict settings. JAMA. 2001; 286:555–562. [PubMed: 11476657]

Dillon WR, Kumar A, Mulani N. Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. Psychological Bulletin. 1987; 101(1):126–135.

Dueck, J.; Aida, M. HURIDOCS standard formats: A tool for documenting human rights violations. Oslo, Norway: Human Rights Information and Documentation Systems International (HURIDOCS); 1993.

Fawzi MCS, Pham T, Lin L, Nguyen TV, Ngo D, Murphy E, Mollica RF. The Validity of Posttraumatic Stress Disorder Among Vietnamese Refugees. Journal of Traumatic Stress. 1997; 10(1):101–108. [PubMed: 9018680]

Fox SH. The Mandinka Nosological System in the Context of Post-Trauma Syndromes. Transcultural Psychiatry. 2003; 40(4):488–506. [PubMed: 14979464]

Fox SH, Tang SS. The Sierra Leonean refugee experience: Traumatic events and psychiatric sequelae. Journal of Nervous and Mental Disease. 2000; 188(8):490–495. [PubMed: 10972567]

Geisinger KF. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. Psychological Assessments. 1994; 6(4):304–412.

Hagengimana, A.; Hinton, DE. Ihahamuka, a Rwandan syndrome of response to the genocide: Blocked flow, spirit assault, and shortness of breath. In: Hinton, DE.; Good, BJ., editors. Culture and panic disorder. Stanford: Stanford University Press; 2009. p. 205-229.

Heine SJ, Lehman DR, Peng KP, Greenholtz J. What's wrong with cross-cultural comparisons of subjective Likert scales? The reference group effect. Journal of Personality and Social Psychology. 2002; 82(6):903–918. [PubMed: 12051579]

Henrich J, Heine SJ, Norenzayan A. The wierdest people in the world? Behavioral and brain sciences. 2010; 33:61–135. [PubMed: 20550733]

Hinton DE, Kirmayer LJ. Local responses to trauma: Symptom, affect, and healing. Transcultural Psychiatry. 2013; 50(5):607–621. [PubMed: 24142932]

Hollifield M, Warner TD, Lian N, Krakow B, Jenkins JH, Kesler J, Westermeyer J. Measuring trauma and health status in refugees: a critical review. JAMA: The Journal Of The American Medical Association. 2002; 288(5):611–621. [PubMed: 12150673]

Hooberman J, Rosenfeld B, Rasmussen A, Keller A. Resiliencein trauma-exposed refugees: The moderating effect of coping style on resilience variables. American Journal of Orthopsychiatry. 2010; 80:557–563. [PubMed: 20950296]

Ichikawa M, Nakahara S, Wakai S. Cross-cultural use of the predetermined scale cutoff points in refugee mental health research. Social Psychiatry & Psychiatric Epidemiology. 2006; 41(3):248–250. [PubMed: 16518569]

Iwata N, Roberts C, Kawakami N. Japan-U.S. comparison of responses to depression scale items among adult workers. Psychiatry Research. 1995; 58(3):237–245. [PubMed: 8570779]

Iwata N, Roberts RE. Age differences among Japanese on the Center for Epidemiologic Studies Depression Scale: an ethnocultural perspective on somatization. Social Science & Medicine. 1996; 43:967–974. [PubMed: 8888466]

Jayawickreme N, Jayawickreme E, Atanasov P, Goonasekera MA, Foa EB. Are culturally specific measures of trauma-related anxiety and depression needed? The case of Sri Lanka. Psychological Assessment. 2012; 24(4):791–800. [PubMed: 22429206]

Lattin, JM.; Carroll, DC.; Green, PE. Analyzing multivariate data. Pacific Grove, CA: Thomson Brooks/Cole; 2003.

Lhewa D, Banu S, Rosenfeld B, Keller AS. Validation of a Tibetan Translation of the Hopkins Symptom Checklist-25 and the Harvard Trauma Questionnaire. Assessment. 2007; 14:223–230. [PubMed: 17690379]

Li Z, Hsiao-Rei Hicks M. The CES-D in Chinese American women: Construct validity, diagnostic validity for major depression, and cultural response bias. Psychiatry Research. 2009; 175:227–232. [PubMed: 20006386]

Lin N. Measuring depressive symptomatology in China. Journal of Nervous and Mental Disease. 1989; 177:121–131. [PubMed: 2918295]

Marsella, AJ.; Friedman, MJ.; Spain, EH. Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications. In: Marsella, AJ.; Friedman, MJ.; Gerrity, ET.; Scurfield, RM., editors. Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications. Washington, D.C.: American Psychological Association; 1996. p. 105-129.

Marshall GN, Schell TL, Elliott MN, Berthold SM, Chun C-A. Mental Health of Cambodian Refugees 2 Decades After Resettlement in the United States. JAMA: Journal of the American Medical Association. 2005; 294(5):571–579. [PubMed: 16077051]

McDonald RP, Ho M-HR. Principles and practice in reporting structural equation analyses. Psychological Methods. 2002; 7(1):64–82. [PubMed: 11928891]

Millsap, RE. Statistical approaches to measurement invariance. New York, NY: Routledge: 2011.

Mollica RF, Caspi-Yavin Y, Bollini P, Truong T, Tor S, Lavelle J. The Harvard Trauma Questionnaire. Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. The Journal Of Nervous And Mental Disease. 1992; 180(2):111–116. [PubMed: 1737972]

Mollica RF, McInnes K, Sarajlic N, Lavelle J, Sarajlic I, Massagli MP. Disability Associated With Psychiatric Comorbidity and Health Status in Bosnian Refugees Living in Croatia. JAMA. 1999; 282(5):433–439. [PubMed: 10442658]

Morel KR. Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. Journal of Personality Assessment. 1998; 70(2):299–314. [PubMed: 9697332]

Neuner F, Schauer M, Klaschik C, Karunakara U, Elbert T. A Comparison of Narrative Exposure Therapy, Supportive Counseling, and Psychoeducation for Treating Posttraumatic Stress Disorder in an African Refugee Settlement. Journal of Consulting and Clinical Psychology. 2004; 72(4): 579–587. [PubMed: 15301642]

Noh S, Kaspar V, Chen X. Measuring depression in Korean immigrants: ASsessing validity of the translated Korean version of CES-D scale. Cross-Cultural Research. 1998; 32(4):358–377.

Norris FH, Perilla JL, Murphy AD. Postdisaster stress in the United States and Mexico: A cross-cultural test of the multicriterion conceptual model of posttraumatic stress disorder. Journal of Abnormal Psychology. 2001; 110(4):553–563. [PubMed: 11727945]

Palmieri PA, Marshall GN, Schell TL. Confirmatory factor analysis of posttraumatic stress symptoms in Cambodian refugees. Journal of Traumatic Stress. 2007; 20(2):207–216. [PubMed: 17427910]

Pole N, Best SR, Metzler T, Marmar CR. Why are Hispanics at greater risk for PTSD? Cultural Diversity and Ethnic Minority Psychology. 2005; 11(2):144–161. [PubMed: 15884985]

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rasmussen A, Katoni B, Keller AS, Wilkinson J. Psychological Distress among Darfur Refugees: Hozun and Majnun. Transcultural Psychiatry. 2011; 48(4):392–415. [PubMed: 21911508]

Rasmussen A, Keatley E, Joscelyn A. Posttraumatic stress in humanitarian disaster settings outside North America and Europe: A review of the emic trauma literature. Social Science & Medicine. 2014; 109:44–54. [PubMed: 24698712]

Rasmussen A, Rosenfeld B, Reeves K, Keller AS. The effects of torture-related injuries on psychological distress in a Punjabi Sikh sample. Journal of Abnormal Psychology. 2007; 116(4): 734–740. [PubMed: 18020719]

Rasmussen A, Smith HE, Keller AS. Factor Structure of PTSD symptoms among West and Central African refugees. Journal of Traumatic Stress. 2007; 20(3):271–280. [PubMed: 17597123]

Sachs E, Rosenfeld B, Lhewa D, Rasmussen A, Keller AS. Entering exile: Trauma, mental health, and coping among Tibetan refugees arriving in Dharamsala, India. Journal of Traumatic Stress. 2008; 21(2):199–208. [PubMed: 18404641]

Shoeb M, Weinstein H, Mollica R. The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. International Journal of Social Psychiatry. 2007; 53:447–463. [PubMed: 18018666]

Shrestha NM, Sharma B, Ommeren MV, Regmi S, Makaju R, Komproe I, Jong JTVMd. Impact of Torture on Refugees Displaced Within the Developing World: Symptomatology Among Bhutanese Refugees in Nepal. JAMA. 1998; 280:443–448. [PubMed: 9701080]

Simms LJ, Watson D, Doebbelling BN. Confirmatory factor analyses of posttraumatic stress symptoms in deployed and non-deployed veterans of the Gulf War. Journal of Abnormal Psychology. 2002; 111(4):637–647. [PubMed: 12428777]

Smith PB. Acquiescent response bias as an aspect of cultural communication style. Journal of Cross-Cultural Psychology. 2004; 35(1):50–61.

Steenkamp JEM, Baumgartner H. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research. 1998; 25(1):78–107.

Steinley D. Profiling local optima in K-means clustering: Developing a diagnostic technique. Psychological Methods. 2006; 11(2):178–192. [PubMed: 16784337]

van Ommeren M. Validity issues in transcultural epidemiology. British Journal of Psychiatry. 2003; 182:376–378. [PubMed: 12724237]

United Nations. Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment. Geneva: United Nations; 1984.

Vinson GA, Chang Z. PTSD symptom structure among West African War trauma survivors living in African refugee camps: A factor-analytic investigation. Journal of Traumatic Stress. 2012; 25(2): 226–231. [PubMed: 22522740]

Yufik T, Simms LJ. A meta-analytic investigation of the structure of posttraumatic stress disorder symptoms. Journal of Abnormal Psychology. 2010; 119(4):764–776. [PubMed: 21090877]

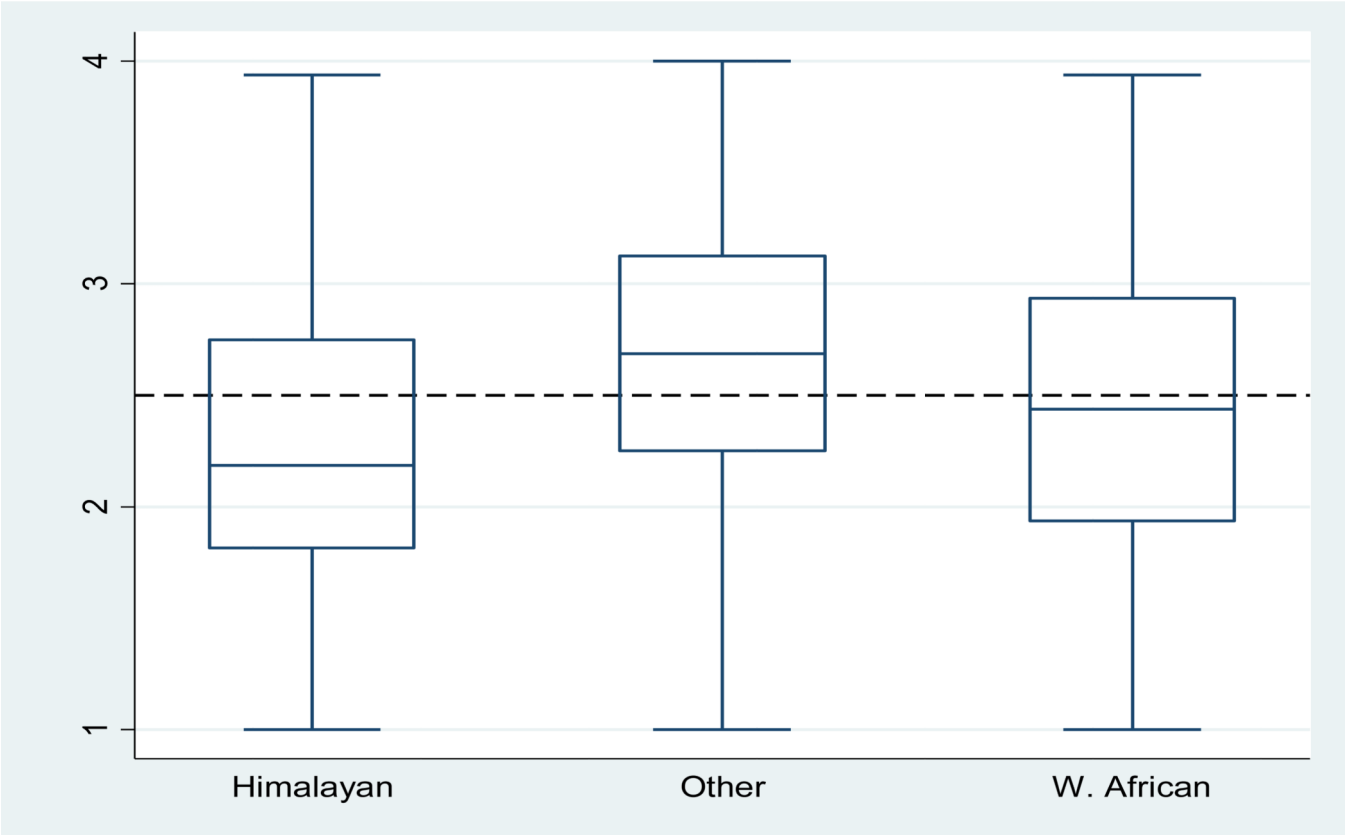

Figure 1.
Distribution of Harvard Trauma Questionnaire scores by regional group

Table 1

Regional and religious demographics by gender

Males Region	Buddhist	Christian	Muslim	Other	Unaffiliated	Unknown	Total	Percent[a]
Afro-Caribbean	1	7	0	1	0	0	9	1.0%
Balkans	0	4	3	0	3	1	11	1.3%
Central Africa	0	68	3	2	0	1	74	8.4%
Eastern Europe	0	17	2	5	1	2	27	3.1%
East & South Africa	0	4	4	0	0	0	8	0.9%
Himalayan region	121	1	0	0	0	4	126	14.4%
Latin America	2	15	0	0	2	1	20	2.3%
MENA[b]	0	0	8	2	1	0	11	1.3%
Other Asia	1	3	8	0	1	1	14	1.6%
South Asia	3	0	24	1	0	0	28	3.2%
West Africa	0	36	149	1	0	4	190	21.6%
Total within males	128	155	201	12	8	14	518	59.0%

Females Region	Buddhist	Christian	Muslim	Other	Unaffiliated	Unknown	Total	Percent[a]
Afro-Caribbean	1	9	1	0	0	1	12	1.4%
Balkans	0	12	3	0	0	2	17	1.9%
Central Africa	0	46	2	0	0	0	48	5.5%
Eastern Europe	8	31	5	5	2	3	54	6.2%
East & South Africa	0	9	3	0	1	0	13	1.5%
Himalayan region	54	0	0	5	0	3	62	7.1%
Latin America	0	3	0	0	0	1	4	0.5%
MENA[b]	0	0	4	0	0	0	4	0.5%
Other Asia	2	5	8	0	1	0	16	1.8%
South Asia	3	2	7	0	0	1	13	1.5%
West Africa	0	20	94	0	1	1	116	13.2%
Total within females	68	137	127	10	5	12	359	40.9%

[a]Percentages are percentages of the total sample (n = 878).

[b]MENA = Middle East and North Africa

Table 2

K-means cluster subscore means

	K-means cluster							
	1 (n = 110)	**2 (n = 115)**	**3 (n = 102)**	**4 (n = 113)**	**5 (n = 125)**	**6 (n = 143)**	**7 (n = 110)**	**8 (n = 60)**
Reexperiencing	1.68	3.17	3.51	3.50	3.09	2.27	2.47	2.81
Avoidance	1.66	3.00	3.43	3.23	3.03	2.21	2.40	2.65
Numbing	1.27	2.32	3.34	2.80	1.97	1.52	2.49	2.00
Hyperarousal	1.31	2.70	3.40	3.37	2.64	2.09	2.31	2.50

Table 3

Regional subsamples by cluster membership

	K-means cluster																
	1		2		3		4		5		6		7		8		Total
Region	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**
Afro-Caribbean	1	4.8	0	0.0	4	19.0	2	9.5	4	19.0	5	23.8	3	14.3	2	9.5	21
Balkans	1	3.6	3	10.7	5	17.9	5	17.9	6	21.4	4	14.3	0	0.0	4	14.3	28
Central Africa	10	8.2	28	23.0	13	10.7	16	13.1	17	13.9	12	9.8	15	12.3	11	9.0	122
Eastern Europe	4	4.9	11	13.6	14	17.3	17	21.0	10	12.3	12	14.8	10	12.3	3	3.7	81
East & South Africa	3	14.3	3	14.3	2	9.5	3	14.3	2	9.5	3	14.3	3	14.3	2	9.5	21
Himalayan region	36	19.1	12	6.4	14	7.4	12	6.4	24	12.8	54	28.7	23	12.2	13	6.9	188
Latin America	2	8.3	4	16.7	6	25.0	2	8.3	3	12.5	1	4.2	5	20.8	1	4.2	24
MENA[a]	1	6.7	2	13.3	3	20.0	0	0.0	3	20.0	0	0.0	4	26.7	2	13.3	15
Other Asia	1	3.3	8	26.7	4	13.3	4	13.3	3	10.0	4	13.3	4	13.3	2	6.7	30
South Asia	4	9.8	7	17.1	6	14.6	9	22.0	5	12.2	3	7.3	6	14.6	1	2.4	41
West Africa	47	15.3	37	12.1	31	10.1	43	14.0	48	15.6	45	14.7	37	12.1	19	6.2	307
Total	110	12.5	115	13.1	102	11.6	113	12.9	125	14.2	143	16.3	110	12.5	60	6.8	878

[a]MENA = Middle East and North Africa

Table 4

Summary Statistics of the Three Models

Model	LR-X^2	df	ΔLR-χ^2 vs Configural	Δdf	RMSEA	TLI	SRMR
Configural	394.13	294	--	--	.034	.96	.039
Metric	438.60	318	48.19*	24	.036	.96	.051
Scalar	503.88	342	117.95**	24	.040	.95	.057

* $p < 0.01$,

** $p < 0.001$

Table 5

Parameter Estimates from the Configural Model

	Loadings			Intercepts			Uniquenesses		
Indicator	**Other**	**Himal.**	**West Af.**	**Other**	**Himal.**	**West Af.**	**Other**	**Himal.**	**West Af.**
R1: Recurrent thoughts or memories of the most hurtful or terrifying events.	.586	.483	.647	3.69	3.60	3.07	.657	.767	.582
R2. Recurrent nightmares	.665	.710	.639	2.56	2.20	2.31	.558	.496	.592
R3. Feeling as though the event is happening again.	.649	.805	.722	2.41	1.91	2.43	.578	.352	.479
R4./R5. Sudden emotional or physical reaction when reminded of the most hurtful or traumatic events.	.639	.394	.661	3.36	3.42	2.98	.591	.845	.563
A1. Avoiding activities that remind you of the traumatic or hurtful event.	.651	.795	.750	2.58	2.27	2.39	.576	.367	.437
A2. Avoiding thoughts or feelings associated with the traumatic or hurtful event.	.468	.757	.724	2.90	2.65	2.41	.781	.426	.476
N1. Inability to remember parts of the most traumatic or hurtful events.	.336	.426	.293	1.81	1.78	1.72	.887	.819	.914
N2. Less interest in daily activities.	.626	.711	.589	2.46	2.09	2.07	.609	.495	.653
N3. Feeling detached or withdrawn from people.	.692	.753	.666	2.35	1.97	1.94	.521	.433	.557
N4. Unable to feel emotions.	.566	.678	.626	1.95	1.83	1.82	.680	.541	.608
N5. Feeling as if you don't have a future.	.534	.569	.605	2.25	1.85	1.98	.714	.676	.634
H1. Trouble sleeping.	.577	.681	.587	3.16	2.55	2.72	.667	.537	.655
H2. Feeling irritable or having outbursts of anger.	.485	.690	.587	2.21	2.01	1.91	.764	.524	.656
H3. Difficulty concentrating.	.622	.693	.602	2.79	2.32	2.41	.613	.520	.638
H4. Feeling on guard.	.588	.598	.635	2.38	2.03	2.05	.654	.642	.596
H5. Feeling jumpy, easily startled.	.560	.694	.663	2.34	1.91	2.19	.686	.518	.561

R = reexeperiencing; A = avoidance; N = numbing; H = hyperarousal; Himal = Himalayan, West Af. = West African

Table 6

Factor Correlations

Himalayan

	R	A	N	H
R	1			
A	0.50	1		
N	0.80	0.62	1	
H	0.85	0.58	0.88	1

West African

	R	A	N	H
R	1			
A	0.71	1		
N	0.79	0.75	1	
H	0.89	0.75	0.98	1

Other

	R	A	N	H
R	1			
A	0.82	1		
N	0.73	0.56	1	
H	0.93	0.72	0.87	1

R = reexeperiencing; A = avoidance; N = numbing; H = hyperarousal