



Published in final edited form as:

*J Trauma Stress*. 2008 August ; 21(4): 369–376. doi:10.1002/jts.20349.

## Spanish and English Versions of the PTSD Checklist – Civilian Version (PCL-C): Testing for Differential Item Functioning

Jeremy N.V. Miles, Grant N. Marshall, and Terry L. Schell

RAND Corporation, 1776 Main Street, P.O. Box 2138, Santa Monica CA, 90407-2138

Jeremy N.V. Miles: jmiles@rand.org; Grant N. Marshall: grantm@rand.org; Terry L. Schell: tschell@rand.org

### Abstract

Interpretation of ethnic differences in PTSD is predicated on demonstration that differences are not due to measurement bias. This is difficult when multiple languages are used in the assessment. This study used confirmatory factor analysis to examine possible differential item functioning (DIF) across English and Spanish versions of the PTSD Checklist-Civilian Version (PCL-C). Data were derived from two assessments of Hispanics ( $Ns = 304, 213$ ), who were hospitalized with physical injuries. After correction for multiple testing, univariate tests revealed no statistically significant DIF effects; multivariate tests revealed some indication of DIF at the initial assessment only. This bias was inconsistent across waves and unlikely to be substantively consequential, indicating that the two versions of the PCL-C were generally equivalent.

---

Hispanics comprise the largest ethnic minority group in the United States (US Census Bureau, 2006). Moreover, recent migration has resulted in a sizable influx of Latin American immigrants and refugees with limited proficiency in English (Shin & Bruno, 2003). A growing body of data suggests that Hispanic Americans appear more likely than their non-Hispanic counterparts both to develop posttraumatic stress disorder (PTSD) and to experience more extreme symptoms of PTSD in response to traumatic stress (Adams & Boscarino, 2006; Galea et al., 2002; Kulka, Fairbank, Jordan, Weiss & Cranston, 1990; Ortega & Rosenheck, 2000; Pole et al., 2001). Yet, most research studies comparing Hispanic and non-Hispanic Americans have either been confined to English-speaking Hispanics (e.g., Pole et al., 2001) or have not examined whether Spanish and English translations of symptom severity measures possess equivalent psychometric properties in both languages (e.g., Galea et al., 2002). Given this apparent ethnic disparity in PTSD and the large number of Hispanics who are native Spanish speaking, it is critical to demonstrate that the Spanish versions of PTSD screening instruments are unbiased relative to English versions. Otherwise, problems in instrument translation may erroneously be interpreted as evidence of substantive ethnic disparities in mental health.

In recent years, a small number studies have investigated the equivalence of Spanish and English versions of assessment instruments measuring PTSD symptom severity (Norris, Perilla, & Murphy, 2001; Orlando & Marshall, 2002; Marshall, 2004). Yet, each of these investigations has shortcomings for determining if the Spanish version is unbiased relative to the English version. Marshall (2004) and Norris et al. (2001), for example, demonstrated equivalent factor loadings across samples of Spanish- and English-speaking respondents but did not examine whether latent means and item intercepts were equivalent. The latter step is a prerequisite for evaluating if the group means are equal. Similarly, by comparing Hispanic respondents who completed a PTSD symptom severity instrument in Spanish with a mixed group of Hispanic and non-

---

<sup>1</sup>Strictly speaking, Norris et al. (2001) conducted an international study comparing English-speaking US citizens and Spanish-speaking citizens of Mexico. Thus, some of the issues we address here are not directly applicable.

Hispanic respondents who completed the Posttraumatic Stress Disorder Symptom Checklist (Weathers et al., 1993) in English, Marshall (2004) may have partially confounded the instrument version with ethnicity. Finally, Orlando and Marshall (2002) did look at item-level bias across English and Spanish versions using Item Response Theory. However, this analysis presumes that PTSD items comprise a single underlying construct. This assumption may be questionable inasmuch as numerous studies have demonstrated that the structure of PTSD symptoms is multidimensional (e.g. Andrews, Joseph, Shevlin, & Troop, 2006; King, Leskin, King, & Weathers, 1998; Palmieri, Marshall, & Schell, 2007); violation of this supposition may compromise the ability to identify bias in the individual scale items.

Prior to determining whether the PTSD scores of Hispanics and non-Hispanics differ from each other given a comparable degree of trauma exposure, it is first necessary to establish that the PTSD instrument actually measures the equivalent construct in each group. One method of assessing equivalency is by means of Differential Item Functioning (DIF; Holland & Wainer, 1993). Also known as item bias, DIF is a potential problem that occurs for all studies that use multiple item measures. If one item is endorsed differentially by members of different groups who have the same true score on a particular measure, then the error component of the score ceases to be random, and the scores will be biased when comparing across those groups.

To give a hypothetical example, PTSD measures usually contain an item assessing “feeling numb.” It may be the case that members of one group are less likely to endorse the feeling numb item, despite having the same level of PTSD symptoms on the other 16 items in the scale. When these groups are compared, researchers are likely to conclude that one group has lower levels of PTSD than the other group. In reality, the groups have the same degree of overall PTSD but somewhat different endorsement of feeling numb. When comparing symptoms across cultural and language groups, it is common to have items with different meanings or interpretations for the people in the different groups. When comparing Hispanic to non-Hispanic groups, for example, it is possible that symptoms such as “numbing” are not interpreted equivalently, particularly when ethnicity is strongly correlated with the language of administration. It is critical, therefore, that researchers who investigate ethnic disparities in PTSD demonstrate that the individual instrument items are not biased across the language versions of the instrument. Otherwise, relatively innocuous problems in translation may be erroneously interpreted as reflective of ethnic disparities in mental health.

Using a sample of individuals who were hospitalized for treatment of injuries stemming from physical trauma, we examined Spanish and English translations of the PTSD Checklist – Civilian Version (Weathers et al., 1993) to test for differential item functioning using confirmatory factor analysis. Data from two measurement occasions were evaluated to assess the generalizability over time of any patterns of DIF that emerged. To avoid confounding language and ethnicity, we included only those respondents who identified themselves as Hispanic, and who reported being able to speak Spanish.

## Method

### Participants

The data that are reported in this paper were part of a larger study examining the psychological impact of physical trauma. Between February 2004 and August 2006, research staff attempted to screen all consecutive hospital admissions for a blunt or penetrating trauma at four trauma centers in Los Angeles County. After an individual was deemed medically capable, research staff conducted a brief interview with potential participants to determine eligibility. Patients were eligible for the larger study if they (a) were over 18 years old, (b) had been hospitalized for surgical treatment of sudden physical injuries not sustained from family violence or attempted suicide; (c) spoke either English or Spanish; and (d) had no severe cognitive

impairment. Multiple attempts were made to approach patients who were initially away from their rooms, receiving visitors, or not sufficiently alert to respond.

Of persons screened, 850 were determined to be eligible. Of these, 677 (80%) completed a baseline interview. Respondents were compensated \$25 for participation, and informed consent was obtained. The resulting sample of the larger study was representative of all Trauma Center admissions in Los Angeles County on age, gender, ethnicity, injury severity, and injury type (Trauma and Emergency Medical Information System, 2007).

For the current purpose, an analytic sample was selected from the larger sample. This sample was restricted to 304 participants who self-identified as “Latino or Hispanic” and who were not English monolingual. This latter subsample was excluded to try to create analytic groups that were relatively comparable on factors other than language of administration. At the initial interview, 111 of 304 (36.2%) respondents completed the interview in Spanish; at the follow up interview, 71 of 213 (33.3%) interviews were completed in Spanish.

Males comprised 245 (80%) of the sample, and the mean age at the time of the incident was 30.1 ( $SD = 10.6$ , range 18 to 65). The most common injuries were attributable to being shot with a gun (24%), hit by a moving vehicle while driving a car or motorcycle (20%), being stabbed with a knife or another object (13%), hitting a non-moving object (such as a wall) as either a car driver (14%) or passenger (9%). The most common injury sites were the legs or feet (57%), arms or hands (54%), chest, stomach or groin (53%) and back (42%).

The research was approved and monitored by the Institutional Review Boards of the RAND Corporation, the University of Southern California School of Medicine, the University of California at Los Angeles Medical School, the Drew University School of Medicine, and the California Hospital Medical Center.

## Measures

PTSD symptom severity was assessed using the 17-item PTSD Checklist–Civilian Version (PCL; Weathers, et al., 1993). The PCL has been used in diverse samples including hospitalized physical injury survivors and possesses solid psychometric properties (Andrykowski, Cordova, Studts, & Miller, 1998; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Denson, Marshall, Schell, & Jaycox, 2007; Zatzick et al., 2007). Participants rated the degree to which they were bothered by each symptom on a scale ranging from 1 (*not at all*) to 5 (*extremely*), with possible scores ranging from 17 to 85. Symptoms were assessed with respect to the injury (e.g., “how much have you been bothered by repeated, disturbing dreams of *the injury*”). For the first assessment, the timeframe for the PCL was “since the injury”; for the follow-up assessment, the timeframe was “in the past week”. At the first assessment, participant total scores averaged 40.6 ( $SD = 15.6$ ); at 6-month follow-up total scores averaged 39.7 ( $SD = 16.7$ ).

The translation procedures have been described previously (Orlando & Marshall, 2002); the Spanish translation is available in Appendix 1.

## Procedure

All participants were assessed using face-to-face structured interviews conducted by trained lay interviewers; items were read to the participants and their verbal responses were recorded by the interviewer. The initial interview took place within several days of hospital admission. Follow-up interviews were conducted in participants' home approximately six months later. The interviews took approximately 60 minutes and were conducted in either English or Spanish as chosen by the participant. The complete interview inquired about a range of topics including demographic and pretrauma characteristics, peritraumatic thoughts and emotions, posttraumatic thoughts, emotions, and behaviors (cf. Denson, Marshall, Schell, & Jaycox, 2007), as

well as mental health service utilization. For the current study, only symptoms of posttraumatic distress were examined.

## Data Analysis

We first tested whether initial scores on the PCL-C or other demographic variables were predictive of dropout, to determine if there were important sample differences between the baseline and follow up interviews that would impact DIF analyses. Using logistic regression, we ran separate models for each predictor, and regressed follow up status (followed up successfully vs dropped out) on sex, age, PCL-C total score, PCL-C subscale scores for criterion B, C and D, and source of injury (assault vs other).

We used a MIMIC (Multiple Indicator Multiple Cause; see Bollen, 1989) model approach to estimate any DIF effects between Spanish and English versions of the PCL-C. Using this approach, a confirmatory factor analysis is fit to the data, and the factors are regressed on a predictor (in this case, language of presentation). The null hypothesis that the factors mediate the relationship between the predictor and the items is tested. If this null hypothesis is rejected, then direct paths are required from the predictor to the items, indicating the presence of DIF. If the null hypothesis is not rejected, then we do not have evidence that the items are functioning differently – that is the mean score of each item is changing in line with the other items.

In the first stage, we fit a confirmatory factor analysis model to the dataset to establish a well-fitting baseline model, shown in Figure 1. Establishing an appropriate baseline model is crucial, as a poorly fitting baseline model can have the effect of introducing apparent DIF as an artifact. Most confirmatory factor analyses studies of PTSD measures have found that one of two 4-factor models is usually found to provide the best fit to the data. The first of these models, proposed by King, et al. (1988) posits that DSM-IV items C1 and C2 form a separate factor measuring avoidance, and that the remainder of the Criterion C items are more appropriately called Numbing. The alternate model, proposed by Simms, Watson, & Doebbeling (2002), retains the two item avoidance factor proposed by King, et al., but incorporates a Dysphoria factor consisting of items C3 to C7 (i.e., the items that are labeled Numbing, by King, et al.) and items D1 to D3 (the first three items of the hyperarousal factor); in the Simms et al. model, items D4 and D5 remain in the hyperarousal cluster. For the baseline model, we used a hybrid model in which the King et al. (1998) and Simms, Watson, and Doebbeling (2002) four-factor models are nested. In other words, the hybrid model contains all of the loadings from the four-factor models posited by Simms et al. (2002) and King et al. (1998). This creates a hybrid model comprised of Reexperiencing, Avoidance, Numbing/Dysphoria and Hyperarousal symptom clusters. Guided by the modification indices, we added two parameters to ensure that the baseline model was able to account for the observed data. The resulting model was fit to both the baseline and follow-up waves of data in order to identify DIF.

We investigated DIF within each of the two waves of data by regressing the four latent variables onto a dichotomous indicator reflecting the language in which the interview was administered. DIF for each item was assessed by adding a direct path from this language variable to each item and comparing the fit to the baseline model using the Wald test. This step was done 17 times at each wave, once for each item. Analyses were performed using maximum likelihood estimation with the Satorra-Bentler scaled  $\chi^2$  correction for non-normality (Satorra & Bentler, 1994) as implemented in Mplus 4.2 (Muthén & Muthén, 2006). This method is robust to violations of assumptions regarding normality; standard errors are adjusted for normality, but parameter estimates are equal to those obtained from maximum likelihood estimation.

This procedure tests the whether the effect of language on any item is mediated entirely by the latent variable of interest. In particular, a statistically significant direct path from language to the item in question indicates that language is affecting the item response beyond the amount

that would be expected by the true differences in the latent variable. In other words, if the two groups have equal means on the latent variable dimension, then they would be expected to have approximately equal means on all of the items that are indicators of that variable. If the latent variable means are equal, and the item means are not equal, then this finding indicates that language has a direct effect on the item.

To correct for  $p$ -value deflation associated with multiple tests, we adjusted the  $p$ -values using the sequentially selective step-up Bonferroni (Hochberg, 1988) as recommended for structural equation models (Cribbie, 2007). As an additional check, we conducted a multivariate test in which we compared a model that exactly fits the associations between language and all items to a model that allows language effects only via the latent variables. This 13  $df$  test detects if significant DIF exists somewhere in the instrument, but does not identify the location or magnitude of the DIF. For that reason, we use the multivariate model as a gateway test to determine if further correction is needed, but actual DIF estimates are taken from the univariate models.

## Results

For descriptive purposes, we calculated the proportion of participants who met screening criteria for PTSD as determined by two criteria. First, we used the Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition criteria (DSM-IV; American Psychiatric Association, 1994), and required that a symptom be endorsed “moderately” or greater to be clinically meaningful (Weathers et al., 1993). We also used the criterion that required positive cases to score at least 50 points on the PCL (Weathers et al., 1993). At initial assessment, 33.8% ( $n = 103$ ) met DSM-IV criteria, excluding duration; 29.3% ( $n = 89$ ) met criteria as determined by the 50-point cutoff. At 6-month follow-up, 36.6% ( $n = 78$ ) met DSM-IV criteria, and 25.4% ( $n = 57$ ) met the 50-point criteria.

To examine predictors of retention, we created a dichotomous indicator of dropout, and used logistic regression to determine the extent to which baseline measures predicted subsequent attrition. None of the predictors approached statistical significance ( $p$ 's > .40).

The first stage of the modeling process was to establish a base model to test DIF. The base model consisted of four correlated factors with some cross loadings incorporated so that both the King et al. (1998) and Simms et al. (2002) models were nested within the base model. At baseline, this model proved an adequate, although not excellent, fit to the data,  $\chi^2(110) = 234$ , CFI = .93, RMSEA = .061, and SRMR = .047, and a slightly better fit at follow up,  $\chi^2(110) = 161$ , CFI = .97, RMSEA = .047, and SRMR = .045.

We examined the residuals, modification indices and expected parameter change statistics, and incorporated two additional parameters (shown in Figure 1). In particular, we added a cross-loading from the Avoidance factor to Item C4 (loss of interest) and a correlation between the unique variances of Items D4 (hypervigilance) and D5 (i.e., startle response); the latter two items comprised the hyperarousal factor as represented in the Simms et al. (2002) model. These additions resulted in an improvement in model fit to  $\chi^2(108) = 209$ , CFI = .94, RMSEA = .056, SRMR = .04 at baseline, and to  $\chi^2(108) = 157$ , CFI = .97, RMSEA = .046, SRMR = .047 at follow up. This model provided a baseline from which to examine the possibility of DIF in both baseline and follow-up data.

Once the base model had been established, we then tested for DIF. We included a dichotomous indicator reflecting the language in which the interview was completed (i.e., Spanish or English), and regressed the four latent variables onto this indicator at each wave. We then tested DIF for each item by fitting a separate model for each item which tested a direct path from language to the item in question. The results of these univariate tests are shown in Table 1.

The first column of Table 1 shows the mean difference in the item scores that would be expected between Spanish and English versions of a given item if the two groups had equal means on all factors. Thus, it indicates the degree and direction of bias in that item when used as an indicator of the factor. For example, for item B1 at baseline, i.e., intrusive thoughts, an individual completing the interview in Spanish would have a mean score that was 0.05 points higher than an English interview respondent with the same level of overall Reexperiencing symptoms, as determined by the other items. As shown in Table 1, no individual DIF effects were statistically significant at either baseline or follow up.

As a second approach to assessing DIF, we conducted a single multivariate test rather than a series of univariate tests. This test has greater power than correcting univariate tests for multiple testing when outcomes are negatively correlated, as they are with DIF (Cole, Maxwell, Arvey, & Salas, 1994; Miles, 2003). At baseline, this comparison gave  $\Delta\chi^2(13) = 25, p = .026$ , suggesting the presence of DIF in some item/items; at follow up, the comparison gave  $\chi^2(13) = 13, p = .48$ , indicating no significant evidence of DIF. To follow up on the significant test on the baseline data, we adjusted the item that showed the greatest evidence of DIF and conducted the test again. Specifically, we subtracted 0.42 from item B3 for respondents who completed the interview in Spanish. There was no significant evidence of DIF at baseline after this one adjustment;  $\Delta\chi^2(13) = 17, p = .20$ . Overall, minimal evidence of item bias was identified; for the single item in which there was some suggestion of bias at baseline, no evidence of bias was found at follow up.<sup>2</sup>

## Discussion

This study assessed the comparability of English and Spanish versions of the PCL-C (Weathers et al., 1993) using a strategy that addressed several shortcomings of previous research. In particular, we used structural equation modeling which did not require making the assumption that PTSD symptoms reflected a single underlying construct (cf. Orlando & Marshall, 2002). Moreover, this approach, conducted entirely among Hispanics in the US, removed some of the possible confounding effects of ethnicity. In addition, we replicated the analyses across two time points and, unlike some of the earlier work (Orlando & Marshall, 2002; Marshall, 2004), we included symptom assessments that occurred after the DSM-IV duration criteria for PTSD.

Using a series of univariate tests, we found no evidence of DIF using the  $p < .05$  level when we corrected for multiple tests. A more powerful multivariate test found slight evidence for DIF at baseline, but no evidence of DIF at 6-month follow-up. In addition, the only item with potential DIF at baseline (item B3) had a considerably smaller DIF effect at 6-month follow-up. Only one item had a comparable DIF effect at both assessments, i.e., D5, and this effect did not approach statistical significance. Overall, we found little or no evidence of consistent DIF across two administrations of the PCL-C. Even the largest DIF effect estimated at baseline was less than half of a scale point on one item. An effect of this magnitude would have no impact on observed rates of PTSD when using the PCL-C as a screening instrument. Our use of two waves of data demonstrated that the Spanish and English versions of the PCL are largely equivalent when administered shortly after the trauma as well as when used several months later.

We consider three implications of these results to be noteworthy. First, this study provides evidence of the functional equivalence of this version of the PCL-C in both Spanish and English

<sup>2</sup>We calculated attrition weights to ensure that the follow up sample was equivalent to the baseline sample on a range of baseline variables (including age, gender, education, hostility, language, treatment site, assault, injury severity, PTSD, depression, time between hospitalization and interview, duration of hospitalization, percent missing to all survey items). We repeated the analysis using these weights; where results were substantively equivalent we interpreted the unweighted analyses to maximize power. The results did not change when attrition weights were applied.

versions when used as a screening instrument for identifying persons with possible PTSD, suggesting that the translation procedure was successful and that the scale can be applied in populations of Spanish and English speakers. In addition, the results indicate that if differences in rates of probable PTSD are observed between Spanish and English speakers, they are unlikely to be artifacts attributable to biases across the English and Spanish versions of the instruments. Finally, even when using the PCL-C as a continuous measure (i.e., using it as a measure of symptom severity), there would appear to be little need to correct for DIF.

Although the results of this study demonstrate the broad equivalence of the English and Spanish translations of the PCL-C, a number of limitations should be noted. First, these results are based on Hispanic survivors of physical trauma who sustained injuries of sufficient severity to require hospitalization. Additional research is needed on persons who experienced other types of trauma. Similarly, the population of Hispanic individuals with traumatic injuries requiring hospitalization in Los Angeles County is demographically different from the general Hispanic population or the populations affected by other traumas (e.g., rape). In particular, the population we studied includes comparatively more males and younger individuals than does the general population. The Spanish speaking Hispanic population of Los Angeles is largely of Mexican origin and may have been established for a longer time than Spanish speaking residents of other metropolitan areas in the US (Frey, 1999). To the extent that language use differs across these demographic groups, caution should be exercised in generalizing from these findings to other groups of trauma survivors.

Second, the baseline interviews occurred approximately one week after the trauma, and therefore patients did not meet the duration criteria for PTSD diagnosis. However, the baseline findings were replicated in symptoms measured at 6-months, reducing concerns about generalizability due to the timing of administration.

The third limitation concerns the nature of our replication. We observed a similar pattern of findings at two different assessment points with the same individuals. A true replication would utilize a new, independent sample of individuals. Thus, additional research is warranted to determine whether our findings hold in another sample. Nonetheless, analysis of the two waves provides more information about the equivalence of the PCL-C than either wave used by itself.

In summary, although appropriate caution should be exercised, the results of this investigation indicate that data obtained from Spanish and English versions of the PCL-C are likely to be equivalent for both research and clinical purposes.

## Acknowledgments

This research was supported by grants R01MH56122 and R01MH071636 from the National Institute of Mental Health and R01AA014246 from the National Institute on Alcohol Abuse and Alcoholism. The views expressed are those of the authors and do not necessarily reflect those of the sponsors or RAND. We express appreciation to Drs. Howard Belzberg, Henry Cryer, Gudata Hinika, Peter Meade, and Vivek Shetty for facilitating data collection. We thank the RAND Survey Research Group and Harris Interactive for their assistance with data collection. We gratefully acknowledge the generosity of the trauma survivors who participated in this study.

## Appendix: Spanish Translation of the PCL-C

Ahora voy a leer una lista de problemas y síntomas que a veces tiene la gente después de una lesión (herida). Dígame cuánto le ha molestado cada una de estas cosas desde que ocurrió la lesión (herida).

**Desde la lesión (herida), ¿cuánto le ha molestado \_\_\_\_\_?**

	Paranada	Unpoco	Más o menos	Bastante	Extremadamente
a. tener recuerdos, pensamientos perturbadores o imágenes que se repiten de la lesión?	1	2	3	4	5
b. tener sueños perturbadores y que se repiten de la lesión?	1	2	3	4	5
c. actuar o sentir de repente como si la lesión ocurriera otra vez (como si lo a vivir)?	1	2	3	4	5
d. sentirse muy disgustado (preocupado o afligido) cuando algo le recuerda la lesión (herida)?	1	2	3	4	5
e. tener reacciones físicas (como latidos fuertes del corazón, le cuesta respirar, suda mucho) cuando algo le recuerda la lesión (herida)?	1	2	3	4	5
f. evitar pensar o hablar sobre la lesión (herida) o evitar sentir algo que que ver con eso?	1	2	3	4	5
g. evitar actividades o situaciones porque le recuerdan cuando estaba siendo (herido)?	1	2	3	4	5
h. tener dificultad para recordar lo que pasó durante el accidente (sin contar lo que no podría recordar por estar inconsciente)?	1	2	3	4	5
i. perder interés en las actividades que antes disfrutaba?	1	2	3	4	5
j. sentirse distante o aislado (alejado) de otras personas?	1	2	3	4	5
k. sentir insensibilidad emocional o incapacidad de sentir amor por sus seres queridos?	1	2	3	4	5
l. sentir como si su futuro será más corto [o interrumpido] de alguna manera?	1	2	3	4	5
m. tener dificultad para quedarse dormido o seguir durmiendo?	1	2	3	4	5
n. sentirse irritado o tener arrebatos de coraje?	1	2	3	4	5
o. tener mucha dificultad para concentrarse?	1	2	3	4	5
p. estar siempre muy “alerta”, vigilante o en guardia?	1	2	3	4	5
q. sentirse sobresaltado o asustado por cualquier cosa?	1	2	3	4	5

## References

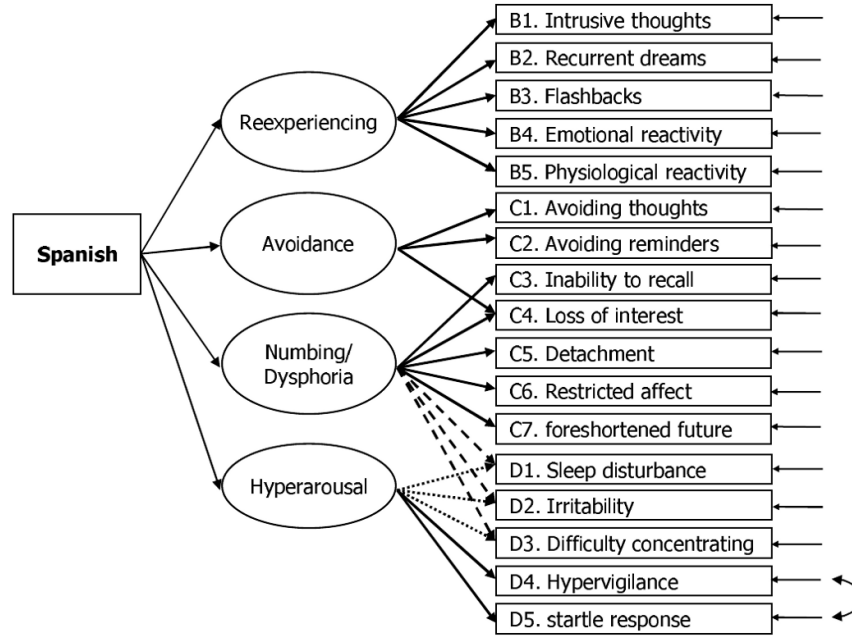
- Adams RE, Boscarino JA. Predictors of PTSD and delayed PTSD after disaster: the impact of exposure and psychosocial resources. *The Journal of Nervous and Mental Disease* 2006;194:485–93. [PubMed: 16840844]
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th. Washington, DC: American Psychiatric Association; 1994.
- Andrews L, Joseph S, Shevlin M, Troop N. Confirmatory factor analysis of posttraumatic stress symptoms in emergency personnel: An examination of seven alternative models. *Personality and Individual Difference* 2006;41:213–224.
- Andrykowski MA, Cordova MJ, Studts JL, Miller TW. Posttraumatic stress disorder after treatment for breast cancer: prevalence of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C) as a



screening instrument. *Journal of Consulting and Clinical Psychology* 1998;66:586–90. [PubMed: 9642900]

- Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD Checklist (PCL). *Behavior Research and Therapy* 1996;34:669–73.
- Bollen, KA. *Structural equations with latent variables*. New York: Wiley; 1989.
- Cole DA, Maxwell SE, Arvey R, Salas E. How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin* 1994;115:465–474.
- Cribbie RA. Multiplicity control in structural equation modeling. *Structural Equation Modeling* 2007;14:98–112.
- Denson TF, Marshall GN, Schell TL, Jaycox LH. Predictors of posttraumatic distress 1 year after exposure to community violence: The importance of acute symptom severity. *Journal of Consulting and Clinical Psychology* 2007;75:683–692. [PubMed: 17907850]
- Frey WH. United States population: Where the new immigrants are. *US Society and Values* 1999;4(2): 25–28.
- Galea S, Ahern J, Resnick H, Kilpatrick D, Bucuvalas M, Gold J, Vlahov D. Psychological sequelae of the September 11 terrorist attacks in New York City. *The New England Journal of Medicine* 2002;346:982–987. [PubMed: 11919308]
- Hochberg Y. A sharper Bonferroni test for multiple tests of significance. *Biometrika* 1988;75:800–802.
- Holland, PW.; Wainer, H. *Differential item functioning*. Hillsdale, NJ: Erlbaum; 1993.
- King DW, Leskin GA, King LA, Weathers FW. Confirmatory factor analysis of the clinician administered PTSD scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment* 1998;10:90–96.
- Kulka, RA.; Fairbank, JA.; Jordan, BK.; Weiss, D.; Cranston, A. *The National Vietnam Veterans Readjustment Study: Table of findings and appendices*. New York: Bruner/Mazel; 1990.
- Marshall GN. Posttraumatic Stress Disorder Symptom Checklist: Factor structure and English-Spanish measurement invariance. *Journal of Traumatic Stress* 2004;17:223–30. [PubMed: 15253094]
- Miles JNV. A framework for power analysis using a structural equation modeling procedure. *BMC Medical Research Methodology* 2003;3:1–11. [PubMed: 12515580]
- Muthén, B.; Muthén, L. *Mplus 4.2*. Los Angeles: Muthén and Muthén; 2006.
- 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;110:553–563. [PubMed: 11727945]
- Orlando M, Marshall GN. Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. *Psychological Assessment* 2002;14:50–59. [PubMed: 11911049]
- Ortega AN, Rosenheck R. Posttraumatic stress disorder among Hispanic Vietnam veterans. *The American Journal of Psychiatry* 2000;157:615–619. [PubMed: 10739422]
- Palmieri PA, Marshall GN, Schell TL. Confirmatory factor analysis of posttraumatic stress symptoms in Cambodian refugees. *Journal of Traumatic Stress* 2007;20:207–216. [PubMed: 17427910]
- Pole N, Best SR, Weiss DS, Metzler T, Liberman AM, Fagan J, Marmar CR. Effects of gender and ethnicity on duty-related posttraumatic stress symptoms among urban police officers. *The Journal of Nervous and Mental Disease* 2001;189:442–8. [PubMed: 11504321]
- Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye, A.; Clogg, CC., editors. *Latent variables analysis: applications for developmental research*. Thousand Oaks: Sage; 1994.
- Shin, HB.; Bruno, R. Language use and English-speaking ability: 2000. Census 2000 brief. 2003 [November 15, 2007]. <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>
- Simms LJ, Watson D, Doebbeling BN. Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology* 2002;111:637–47. [PubMed: 12428777]
- Trauma and Emergency Medical Information System for 2005. Los Angeles County: 2007. unpublished data

- US Census Bureau. Nation's population one-third minority. 2006 [October 23, 2007].  
<http://www.census.gov/Press-release/www/releases/archives/population/006808.html>
- Weathers, FW.; Litz, BT.; Herman, DS.; Huska, JA.; Keane, TM. The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Paper presented at The International Society for Traumatic Stress Studies; 1993.
- Zatzick DF, Rivara FP, Nathens AB, Jurkovich GJ, Wang J, Fan MY, Russo J, Salkever DS, Mackenzie EJ. A nationwide US study of post-traumatic stress after hospitalization for physical injury. *Psychological Medicine* 2007;37:1469–1480. [PubMed: 17559704]



**Figure 1.** Path model for testing Differential Item Functioning. Dashed lines represent parameters found only in the Simms et al. (2002) model, whereas dotted lines represent parameters found only in the King et al. (1988) model. (Note that for clarity, residual variances of the latent variables, and correlations amongst those residual variances are not shown).

**Table 1**  
**Differential Item Functioning Effects and  $p$ -values**

DSM-IV Item Number	Baseline		Follow Up	
	DIF Effect	Corrected $p$ value	DIF Effect	Corrected $p$ value
B1	0.05	.91	0.10	.94
B2	-0.12	.91	-0.11	.94
B3	0.42	.09	0.14	.94
B4	-0.17	.91	-0.13	.94
B5	-0.16	.91	-0.01	.94
C1	0.12	.91	-0.28	.94
C2	-0.09	.91	0.29	.94
C3	-0.03	.91	0.24	.94
C4	-0.09	.91	-0.07	.94
C5	0.14	.91	0.14	.94
C6	-0.06	.91	-0.16	.94
C7	0.01	.91	0.14	.94
D1	-0.30	.25	-0.33	.48
D2	-0.27	.18	0.02	.94
D3	0.25	.25	-0.08	.94
D4	-0.16	.91	-0.19	.94
D5	0.26	.60	0.36	.34