Title
Reliability and validity of the Perceived Stress Scale-10 in Hispanic Americans with English or Spanish language preference.

Journal
Journal of health psychology, 24(5)

ISSN
1359-1053

Authors
Baik, Sharon H
Fox, Rina S
Mills, Sarah D
et al.

Publication Date
2019-04-01

DOI
10.1177/1359105316684938

Peer reviewed

JHP

Article

Journal of Health Psychology
2019, Vol. 24(5) 628–639

DOI: 10.1177/1359105316684938
journals.sagepub.com/home/hpq
SAGE

Reliability and validity of the Perceived Stress Scale-10 in Hispanic Americans with English or Spanish language preference

Sharon H Baik[1,2], Rina S Fox[1,2], Sarah D Mills[1,2], Scott C Roesch[1,3], Georgia Robins Sadler[1,2,4], Elizabeth A Klonoff[1,3] and Vanessa L Malcarne[1,2,3]

Abstract

This study examined the psychometric properties of the Perceived Stress Scale-10 among 436 community-dwelling Hispanic Americans with English or Spanish language preference. Multigroup confirmatory factor analysis examined the factorial invariance of the Perceived Stress Scale-10 across language groups. Results supported a two-factor model (negative, positive) with equivalent response patterns and item intercepts but different factor covariances across languages. Internal consistency reliability of the Perceived Stress Scale-10 total and subscale scores was good in both language groups. Convergent validity was supported by expected relationships of Perceived Stress Scale-10 scores to measures of anxiety and depression. These results support the use of the Perceived Stress Scale-10 among Hispanic Americans.

Keywords

Hispanic, reliability, scale, stress, validation

Introduction

The 14-item self-report Perceived Stress Scale (PSS; Cohen et al., 1983) is widely used to assess the degree to which situations in one's life are appraised as stressful (Cohen et al., 1983). As a global stress measure, the PSS items are general in nature rather than event-specific, and evaluate the extent to which individuals perceive their lives to be "unpredictable, uncontrollable, and overloading" (Cohen et al., 1983: p. 387). Respondents rate the frequency of their feelings and thoughts about life events and situations over the previous month using a five-point scale ranging from (0) Never to (4) Very Often. The PSS yields a total score that describes overall perceived stress. A shorter

[1]SDSU/UC San Diego Joint Doctoral Program in Clinical Psychology, USA
[2]UC San Diego Moores Cancer Center, USA
[3]San Diego State University, USA
[4]UC San Diego School of Medicine, USA

Corresponding author:
Vanessa L Malcarne, SDSU/UC San Diego Joint Doctoral Program in Clinical Psychology, 6363 Alvarado Court, Suite 103, San Diego, CA 92120-4913, USA.
Email: vmalcarne@mail.sdsu.edu

10-item version of the PSS (PSS-10; Cohen and Williamson, 1988) was derived by removing the four items with the lowest factor loadings (Items 4, 5, 12, and 13) from the original scale (PSS). The PSS-10 was recommended by the scale developers for use in future research, as it demonstrated psychometric properties comparable to the original, 14-item version.

The psychometric properties of the PSS-10 were originally evaluated in a large national sample of 2387 American adults. Cohen and Williamson (1988) reported that scores on the PSS-10 demonstrated adequate internal consistency reliability ($\alpha = .78$); moderate concurrent criterion validity with the amount of stress experienced during an average week ($r = .39$, $p < .001$) and the frequency of stressful life events within the past year ($r = .32$, $p < .001$); and adequate convergent validity as evidenced by expected negative associations with perceived health status ($r = -.22$, $p < .001$) and positive associations with psychosomatic symptoms (rs = .28–.34, $p < .001$) and health service utilization ($r = .22$, $p < .001$). Since then, other studies have similarly reported that the PSS-10 has good internal consistency reliability (e.g. Barbosa-Leiker et al., 2013; Golden-Kreutz et al., 2004; Reis et al., 2010) and adequate convergent validity based on associations with measures of physical and mental health (e.g. Mitchell et al., 2008; Roberti et al., 2006; Wu and Amtmann, 2013).

The dimensionality of the PSS-10 has received considerable empirical attention, although the underlying structure of the measure remains controversial. Studies have consistently identified a two-factor structure with six negatively worded items (Items 1–3, 6, 9, and 10) comprising the first factor and four positively worded items (Items 4, 5, 7, and 8) comprising the second factor (e.g. Barbosa-Leiker et al., 2013; Cohen and Williamson, 1988; Golden-Kreutz et al., 2004; Reis et al., 2010; Roberti et al., 2006). Cohen and Williamson (1988) originally argued that the distinction between the two factors was irrelevant, as factor structure corresponded to item directionality. They conceptualized perceived stress, as measured by the PSS-10, as a single construct that can be evaluated by both negatively and positively worded items. Other researchers, however, have posited that the two-factor model reflects distinct negative and positive components of the stress experience (Barbosa-Leiker et al., 2013; Golden-Kreutz et al., 2004; Roberti et al., 2006). Specifically, the two factors represent the negative feelings associated with stress ("Stress") and positive feelings counter to stress ("Counter Stress"), consistent with Folkman's (1997) modified stress theory. Factor 1 (negatively worded items) has also been termed "Perceived Helplessness" (Roberti et al., 2006) and "Negative Stress" (Reis et al., 2010). Factor 2 (positively worded items) has been termed "Perceived Self-Efficacy" (Roberti et al., 2006) and "Positive Stress" (Reis et al., 2010).

Recently, two studies proposed a bifactor model of the PSS-10 comprised of a single underlying general factor of perceived stress and two domain-specific factors composed of negatively (Factor 1) and positively (Factor 2) worded items (Jovanović and Gavrilov-Jerković, 2015; Wu and Amtmann, 2013). In a bifactor model, all items load onto a general factor representing the target construct (perceived stress) intended to be assessed by a measure (PSS-10), as well as one of the domain-specific factors. The general factor accounts for the common variance among all observed indicators (PSS-10 items), and the domain-specific factors account for the unique variance of subsets of indicators above and beyond the variance explained by the general factor (Brown, 2013; Rios and Wells, 2014). Both studies exploring this structure reported that the bifactor model of the PSS-10 demonstrated a better fit than the two-factor or one-factor models in an American sample of 446 adults with multiple sclerosis (Wu and Amtmann, 2013) and in Serbian clinical ($N = 157$ psychiatric outpatients) and non-clinical ($N = 458$ university students and adults) samples (Jovanović and Gavrilov-Jerković, 2015). Wu and Amtmann (2013) also concluded that the PSS-10 was sufficiently unidimensional, as the general factor of perceived stress accounted for greater variance than either of the domain-specific factors, and thus supported the use of the PSS-10 total score. Conversely, Jovanović and Gavrilov-Jerković (2015) suggested that both the total and subscale scores of the PSS-10 could be used, although variance explained by the general or domain-specific factors was not evaluated or reported. Further examination of the scale's dimensionality using a bifactor approach is needed to determine whether the PSS-10 best represents a single dimension (total score) or different dimensions (subscale scores) of perceived stress, or whether both total and subscale scores should be used.

The PSS, in its 14- and 10-item versions, has been widely used across diverse samples and translated into 30 languages, including Spanish (Cohen, 2014). Currently, there are two validated Spanish language versions of the PSS, one that was developed in Spain and validated in a sample of 440 Spanish adults (European Spanish PSS-14; Remor, 2006) and another that culturally adapted the European Spanish PSS-14 for use in Mexico by revising the wording of items to be more culturally appropriate for Mexican samples, and was validated in a sample of 365 Mexican college students (Mexican Spanish PSS-14; Ramírez and Hernández, 2007). Both Spanish language versions of the PSS-14 demonstrated adequate internal consistency reliability (European Spanish: $\alpha = .81$; Mexican Spanish: $\alpha = .83$) and convergent validity evidenced by correlations with scores on measures of anxiety (European Spanish: $r = .64$, $p < .001$), depression (Mexican Spanish: $r = .55$, $p = .001$), and emotional exhaustion (Mexican Spanish: $r = .52$, $p = .001$). Additionally, exploratory and confirmatory factor analyses of the Mexican Spanish PSS-14 confirmed the two-factor structure found in the original English version of the PSS-14 (Cohen and Williamson, 1988). Although both European Spanish (Remor, 2006) and Mexican Spanish (Ramírez and Hernández, 2007) versions of the PSS-14 have been used in studies of Hispanic Americans in the United States (e.g. Chavez-Korell and Torres, 2013; Detjen et al., 2007; Segrin and Badger, 2013), the psychometric properties of these measures have not been evaluated outside of Spain or Mexico. Furthermore, there are no known studies that have validated the English or Spanish versions of the PSS-10 for use in Hispanic Americans, and the measurement equivalence of the different language versions has not yet been established. Without demonstrating such measurement equivalence, it is unknown if observed differences in PSS-10 scores across language groups reflect true differences in stress, or differences in how these groups define, experience, and communicate stress (Corral and Landrine, 2010; Geisinger, 1994). Given the large and growing Hispanic American population in the United States, a valid and reliable measure of perceived stress is needed for this population.

The primary purpose of this study was to conduct a psychometric evaluation of the English (Cohen and Williamson, 1988) and Spanish ("Dr. Cohen's Scales," 2015) versions of the PSS-10 for Hispanic Americans with different language preferences. First, factorial validity of the one-factor, two-factor, and bifactor models of the PSS-10 was tested to determine the best fitting model in Hispanic Americans with English or Spanish language preference. Next, the measurement and structural invariance of the best fitting model of the PSS-10 was tested between English and Spanish language preference subgroups. Based on previous factor analytic studies, a two-factor solution corresponding to item directionality was hypothesized to best fit the data for both language preference subgroups. Once the factor structure was established, the internal consistency reliability of the PSS-10 total and subscale scores was examined for the full sample and for both language preference subgroups separately. Finally, convergent validity of the English and Spanish versions of the PSS-10 was evaluated by examining correlations with measures of depression and anxiety. For both language preference subgroups, the PSS-10 full scale and the negatively-worded-item-subscale were hypothesized to be strongly and positively correlated with measures of depression and anxiety, while the positively-worded-item-subscale was hypothesized to be moderately and negatively correlated with measures of depression and anxiety.

Methods

Participants

Participants were a community sample of 436 self-identified Hispanic American adults with either an English (n=210) or Spanish (n=226) language preference. Individuals were eligible for inclusion if they self-identified as Hispanic American, were 21 years of age or older, were residents of the United States, and were sufficiently literate in either English or Spanish to complete informed consent procedures and study questionnaires.

Measures

PSS-10. As described above, the PSS-10 is a 10-item self-report measure of global perceived stress (Cohen and Williamson, 1988; "Dr. Cohen's Scales," 2015). A total score ranging from 0 to 40 is computed by reverse scoring the four positively worded items and then summing all the scale items. Higher scores indicate greater levels of perceived stress. Subscale scores were computed by summing the six negatively worded items (Items 1, 2, 3, 6, 9, and 10) for Factor 1 ("Negative") and the four positively worded items (Items 4, 5, 7, and 8) for Factor 2 ("Positive"), with higher scores indicating greater negative distress/stress feelings and greater positive stress feelings and coping abilities, respectively.

Generalized Anxiety Disorder-7. The Generalized Anxiety Disorder (GAD-7) is a self-report measure that assesses the presence and severity of anxiety-related symptoms (García-Campayo et al., 2010; Spitzer et al., 2006), as outlined by the *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition—Text Revision* (DSM-IV-TR; American Psychiatric Association (APA), 2000). Respondents rate the occurrence of each anxiety symptom over the past two weeks on a four-point scale ranging from (0) Not at All to (3) Nearly Every Day. Total scores range from 0 to 21, with higher scores indicating greater severity of anxiety. The GAD-7 was recently validated for use among Hispanic Americans with an English or Spanish language preference (Mills et al., 2014). Internal consistency reliability was good for the total sample (α=.93) and for both language preference groups (English: α=.91; Spanish: α=.94).

Patient Health Questionnaire-9. The Patient Health Questionnaire-9 (PHQ-9) is a self-report measure of depression (Huang et al., 2006; Spitzer et al., 1999). The PHQ-9 items correspond to the diagnostic criteria for depression as outlined by the DSM-IV-TR (APA, 2000). Respondents rate the occurrence of nine depressive symptoms over the past two weeks on a four-point scale ranging from (0) Not at All to (3) Nearly Every Day. Total scores range from 0 to 27, with higher scores indicating a greater endorsement of depressive symptoms. Both English and Spanish versions of the PHQ-9 have been validated for use in Hispanic Americans (Merz et al., 2011). In this study, internal consistency reliability was good for the total sample (α=.90) and for both language preference groups (αs=.90).

Procedure

This study used data that were collected from a community-based cross-sectional study examining the validity of English and Spanish language measures among Hispanic Americans. Prior to human subject enrollment, study protocols and materials were reviewed and approved by the sponsoring universities' Institutional Review Boards. Participants were recruited from Southern California using diverse recruitment strategies, including flyer distribution, word of mouth, health fairs, and engagement with community and religious leaders. Eligible participants provided written informed consent and completed survey questionnaires in their preferred language of English or Spanish. Participants were given US$75 for their time and effort.

Data analysis

To test the measurement and structural invariance across English and Spanish language preference groups, multigroup confirmatory factor analysis (CFA) was used to test formal hypotheses of model parameter invariance across groups. A series of increasingly restrictive, nested models were fit to the data, with each model imposing additional equality constraints and being tested against the less restrictive model to determine the level of invariance across groups. First, to determine the best fitting model, the one-factor, two-factor, and bifactor models of the PSS-10 were estimated and tested separately for the English and Spanish language preference groups. For the one-factor model, all 10 items of the PSS-10 were specified to load onto a single factor (perceived stress). For the two-factor model, six negatively worded items (Items 1, 2, 3, 6, 9, and 10) were specified to load onto the first factor labeled "Negative," and four positively worded items (Items 4, 5, 7, and 8) were specified to load onto the second factor labeled "Positive." For the bifactor model, all items were specified to load onto a single general factor of perceived stress and onto one of the domain-specific factors ("Negative" or "Positive") indicated in the two-factor model. After determining the best fitting factor solution for both language preference groups, the *configural invariance model*, which is the least restrictive, tested whether the overall factor structure was equivalent across language preference groups by examining the pattern of free and fixed model parameters without any equality constraints imposed. Next, the *metric invariance model* constrained each item's factor loading to equivalence across groups and examined whether the relations between items and factors are equivalent in both groups. This level of invariance tested whether each item on the PSS-10 loaded equivalently onto the same factor in both English and Spanish language preference groups. Metric invariance, also known as weak measurement invariance, indicates equal measurement units of the scale and implies that respondents in both groups interpret the items in the same way. Then, the *scalar invariance model* additionally constrained the item intercepts and tested whether items have the same intercept (item means) across language preference groups. Scalar invariance, or strong measurement invariance, indicates that item scores from both groups have equivalent measurement metric and the same scalar, and allows for direct comparisons of factor means across groups. A lack of scalar invariance suggests the presence of a systematic bias in the response patterns between groups. Finally, the *factor variance/covariance invariance model* (structural invariance), the most restrictive model, included additional constraints on factor variances and covariances to determine whether the relationship between latent factors was equivalent across language preference groups. This level of invariance tested whether English and Spanish language preference groups use the same range of the continuum of PSS-10 scores and whether the association between factors is equivalent across language groups.

Overall model fit was evaluated using multiple fit indices, as recommended by Bentler (2007). Because preliminary analyses revealed evidence of multivariate non-normality for the observed measures, the Satorra–Bentler-scaled χ^2 (S-Bχ^2; Satorra and Bentler, 1994) was used, with a non-significant test value ($p > .05$) indicating acceptable model fit. Additionally, three descriptive fit indices were examined: (1) the comparative fit index (CFI; Bentler, 1990), (2) the root mean square error of approximation (RMSEA; Steiger, 1990), and (3) the standardized root mean residual (SRMR; Hu and Bentler, 1999). CFI values greater than .90 were indicative of acceptable model fit, and values greater than .95 were indicative of good model fit. For the RMSEA and SRMR indices, values less than .08 indicated acceptable model fit, and values less than .05 indicated good model fit. A model was deemed to adequately fit the data if at least two descriptive fit indices met criteria for acceptable model fit. For tests of invariance, the S-Bχ^2 difference test (ΔS-Bχ^2; Satorra and Bentler, 2001) was used to statistically test the significance of differences between nested models, with a non-significant value ($p > .05$) indicating that the nested model fit the data as well as the comparison model. However, because the S-Bχ^2 difference test also is sensitive to sample size, changes in descriptive fit indices (ΔCFI, ΔRMSEA, and ΔSRMR) were also examined. Based on Chen's (2007) recommended criteria, $\Delta CFI \leqslant -.010$, $\Delta RMSEA \geqslant .015$, and $\Delta SRMR \geqslant .010$ signified a decrement in model fit, indicating a lack of invariance. The factor structure and invariance analyses were conducted using Mplus Version 7.2 (Muthén and Muthén, 1998-2012), and missing data were handled via the maximum likelihood estimation with robust standard errors method, which uses all available data points.

Once the factor structure was determined, Cronbach's alpha was used to evaluate the internal consistency reliability of total and subscale scores for the English and Spanish language versions of the PSS-10. Independent samples *t*-tests were used to compare PSS-10 total and subscale scores across language preference groups. Convergent validity (i.e. relationship between a measure and conceptually related constructs; Groth-Marnat, 2009) was evaluated via bivariate correlations between scores on the PSS-10 and scores on the GAD-7 and the PHQ-9. Based on past research findings, levels of perceived stress were hypothesized to be significantly and moderately positively correlated with measures of anxiety (GAD-7) and depression (PHQ-9). Correlations between Factor 1 ("Negative") subscale scores and GAD-7 and PHQ-9 total scores were hypothesized to be stronger than those between Factor 2 ("Positive") and GAD-7 and PHQ-9 scores, and in the same direction as the PSS-10 total scores. In addition, due to the similarity of item content between the PSS-10 (e.g. felt nervous and "stressed") and GAD-7 (e.g. felt nervous or anxious), PSS-10 total scores and Factor 1 ("Negative") subscale scores were hypothesized to correlate more strongly with scores on the GAD-7 than with scores on the PHQ-9. Reliability and convergent validity of the PSS-10 were evaluated using SPSS version 22.0 (IBM Corp., 2013).

Results

Descriptive analyses

Sample characteristics for the English and Spanish language preference groups can be found in Table 1. The average age of the total sample was 42.5 years old (standard deviation (SD) = 14.07). There were minimal missing data on the PSS-10, GAD-7, and PHQ-9 (no item had more than 2% missing data) so pairwise deletion of missing data was used (e.g. means and SDs were computed based on all available data for each variable; correlations were based on each pair of variables with non-missing data). For the total sample, the mean PSS-10 total score was 15.14 ($SD = 6.52$, range: 0–33), Factor 1 ("Negative") score was 10.03 ($SD = 5.12$, range: 0–24), and Factor 2 ("Positive") score was 10.86 ($SD = 3.25$, range: 0–16). Mean PSS-10 total and factor scores for both language groups are reported in Table 1. Mean PSS-10 total scores were not statistically different between language preference groups, $t(421) = .30$, $p = .764$. Mean English "Negative" scores did not significantly differ from mean Spanish "Negative" scores, $t(426) = 1.73$, $p = .084$. However, mean "Positive" scores were higher in the English language preference group, $t(428) = 2.07$, $p = .039$.

Single-group CFA models

CFA was used to examine one-factor, two-factor, and bifactor models of the PSS-10 for the English and Spanish language preference groups (see supplementary table 1 for the goodness-of-fit indices). The one-factor model did not fit well statistically (English: S-Bχ^2 (35, $N = 210$) = 211.18, $p < .001$; Spanish: S-Bχ^2 (35, $N = 226$) = 212.46, $p < .001$) or descriptively (English: CFI = .76, RMSEA = .16, SRMR = .11; Spanish: CFI = .73, RMSEA = .15, SRMR = .15) for either language preference group. The two-factor model, comprised of "Negative" and "Positive" factors, provided acceptable fit in both English (S-Bχ^2 (34, $N = 210$) = 83.19, $p < .001$; CFI = .93, RMSEA = .08, SRMR = .05) and Spanish (S-Bχ^2 (34, $N = 226$) = 48.35, $p = .05$;

Table 1. Sample characteristics.

Age, *M* (*SD*)	English (*n*=210)		Spanish (*n*=226)	
	38.50 (13.74)		46.24 (13.37)	
Gender, *n* (%)				
Female	107 (51.0%)		112 (49.6%)	
Male	103 (49.0%)		114 (50.4%)	
Education, *n* (%)				
Less than high school	13 (6.2%)		108 (47.8%)	
High school/trade school	39 (18.6%)		48 (21.2%)	
Some college/associates degree	81 (38.6%)		41 (18.1%)	
Bachelor's degree	57 (27.1%)		17 (7.5%)	
Postgraduate	18 (8.6%)		7 (3.1%)	
Do not know/missing	2 (1.0%)		5 (2.2%)	
Employment status, *n* (%)				
Employed	141 (67.1%)		105 (46.5%)	
Unemployed	30 (14.3%)		42 (18.6%)	
Homemaker	6 (2.9%)		30 (13.3%)	
Student/retired/disabled	19 (9.0%)		30 (13.3%)	
Social security/SSI	4 (1.9%)		9 (4.0%)	
Do not know/missing	10 (4.8%)		10 (4.4%)	
Country of birth, *n* (%)				
United States	131 (62.4%)		32 (14.2%)	
Mexico	52 (24.8%)		134 (59.3%)	
Other	9 (4.3%)		4 (1.8%)	
Missing	18 (8.6%)		56 (24.8%)	
Annual household income, *n* (%)				
Less than US$15,000	37 (17.6%)		77 (34.1%)	
US$15,000–US$34,999	44 (20.9%)		75 (33.2%)	
US$35,000–US$49,999	39 (18.6%)		29 (12.8%)	
US$50,000–US$75,000	41 (19.5%)		11 (4.9%)	
More than US$75,000	34 (16.2%)		9 (4.0%)	
Do not know/missing	15 (7.1%)		25 (11.1%)	
Marital status, *n* (%)				
Married	95 (45.2%)		116 (51.3%)	
Not married	115 (54.8%)		109 (48.2%)	
Missing	0 (0.0%)		1 (0.4%)	
Children, *n* (%)				
Yes	121 (57.6%)		144 (63.7%)	
No	89 (42.4%)		81 (35.8%)	
Missing	0 (0.0%)		1 (0.4%)	
	Scale range	*M* (*SD*)	Scale range	*M* (*SD*)
PSS-10				
Total score	0–33	15.24 (6.62)	0–31	15.05 (6.44)
"Negative"	0–24	10.46 (4.88)	0–22	9.61 (5.31)
"Positive"	3–16	11.19 (2.85)	0–16	10.55 (3.57)
GAD-7	0–21	4.17 (4.64)	0–21	5.68 (5.74)
PHQ-9	0–23	4.50 (5.09)	0–26	4.87 (5.39)

SSI: social security income; *M*: mean; *SD*: standard deviation; PSS-10: Perceived Stress Scale-10; GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9.

Table 2. Goodness-of-fit indices for configural, metric, scalar, and factor variance covariance invariance models of the Perceived Stress Scale-10.

Model	S-Bχ^2	*df*	*p*	CFI[a]	RMSEA[b]	SRMR[b]	ΔS-Bχ^2	Δ*df*	Δ*p*	ΔCFI[c]	ΔRMSEA[d]	ΔSRMR[e]
1. Configural	138.332	68	<.001	.947	.069	.055	–	–	–	–	–	–
2. Metric	148.420	76	<.001	.945	.066	.062	9.673	8	.289	.000	−.004	.005
3. Scalar	164.514	84	<.001	.939	.066	.064	16.150	8	.040	−.006	.000	.002
4. Factor variance covariance	191.925	87	<.001	.920	.074	.097	22.307	3	<.001	−.019	.008	.033

S-Bχ^2: Satorra–Bentler χ^2statistic; *df*: degrees of freedom, CFI: comparative fit index; RMSEA: root mean square error of approximation; SRMR: standardized root mean residual.
[a]Acceptable fit > .90, Good fit > .95.
[b]Acceptable fit < .08, Good fit < .05.
[c]Improved fit ⩽ .010.
[d]Improved fit ⩽ .015.
[e]Improved fit ⩽ .010.

CFI = .98, RMSEA = .04, SRMR = .06) language preference groups. Finally, the bifactor model, which included a single general factor (perceived stress) and two specific domain factors (“Negative” and “Positive”), provided good fit for the English language preference group (S-Bχ^2 (25, N = 210) = 35.74, p = .08; CFI = .99, RMSEA = .05, SRMR = .03). However, the bifactor model was unable to be estimated for the Spanish language preference group due to negative residual variance, indicating model misspecification, from one of the factors. Because only the two-factor solution demonstrated acceptable model fit for both language preference groups, tests of measurement and structural invariance were only conducted on the two-factor model. For the two-factor model in the full sample, standardized factor loadings were statistically significant and ranged from 0.17 to 0.81 for the “Negative” factor and 0.60 to 0.80 for the “Positive” factor. Interfactor correlation was small (r = .25, p < .001), and this finding that the two PSS-10 factors were not highly associated provides additional evidence against the one-factor model and the use of a total score.

Multigroup CFA models

Fit indices for the configural, metric, scalar, and factor variance covariance invariance models across language for the two-factor model of the PSS-10 are presented in Table 2.

Configural invariance. Configural invariance was examined by fitting the two-factor model to the data for the English and Spanish language preference groups, without any equality constraints imposed. For both language groups, the baseline two-factor model fit adequately according to the CFI, RMSEA, and SRMR. Factor loadings from English and Spanish baseline models are included in supplementary table 2. All unstandardized factor loadings were statistically significant and in the same direction for both languages (λs = .81–1.31, ps < .001), providing further support for the configural invariance of the two-factor model.

Metric invariance. Metric invariance was examined by constraining all factor loadings to equivalence across English and Spanish language preference groups. The metric invariance model fit the data well (CFI value > .90, RMSEA and SRMR values < .08), indicating that factor loadings were invariant across language preference groups. When the metric invariance model was compared to the less restrictive configural invariance model, results demonstrated no statistical ($p > .05$) or descriptive (all Δ values < .01) differences in model fit. Thus, the metric invariance model was deemed a superior fit to the data, and weak measurement invariance was satisfied.

Scalar invariance. Scalar invariance was examined by constraining factor loadings and item intercepts to equivalence across English and Spanish language preference groups. The scalar invariance model fit the data well (CFI value > .90, RMSEA and SRMR values < .08), indicating that both factor loadings and item intercepts were invariant across language groups. When this constrained model was compared to the less restrictive metric invariance model, no descriptive differences (all Δ values < .01) were noted in model fit. Thus, strong measurement invariance was met.

Factor variance/covariance invariance. Factor variance/covariance invariance was examined by constraining factor loadings, item intercepts, factor variances, and factor covariances to equivalence across language preference groups. The factor variance/covariance invariance model demonstrated acceptable model fit according to the CFI and RMSEA. When this model was compared to the less restrictive scalar invariance model, descriptive criteria (ΔCFI = −.019, ΔRMSEA = .008, ΔSRMR = .033) indicated a decrement in model fit, indicating a lack of factor variance/covariance invariance. Thus, the scalar invariance model was considered the superior fit to the data.

Internal consistency reliability

Internal consistency reliability for the PSS-10 total scores was adequate for the full sample ($\alpha = .82$), English language preference group ($\alpha = .87$), and Spanish language preference group ($\alpha = .78$). Alphas for Factor 1 ("Negative"; 6 items) scores were adequate for the full sample and both language groups (all αs = .89). Alphas for Factor 2 ("Positive"; 4 items) scores were also adequate for the full sample ($\alpha = .78$), English language preference group ($\alpha = .77$), and Spanish language preference group ($\alpha = .78$).

Convergent validity

Correlations between scores on the PSS-10 and scores on the convergent validity measures are presented in supplementary table 3. As anticipated, PSS-10 total and subscale scores were significantly correlated with scores on measures of anxiety (GAD-7) and depression (PHQ-9), in the hypothesized directions, for the total sample and for the English and Spanish language preference groups separately, supporting convergent validity. Specifically, PSS-10 total and Factor 1 ("Negative") scores correlated strongly and positively with anxiety (rs = .64–.71) and depression (rs = .56–.64), while PSS-10 Factor 2 ("Positive") scores were less strongly (small-to-moderate correlations) and negatively correlated with anxiety (rs = −.25 to −.46) and depression (rs = −.24 to −.46). Comparisons of correlations between language preference groups indicated that the Spanish language preference group had smaller correlations between the "Positive" subscale and anxiety ($z = -2.52$, $p = .01$) and depression ($z = -2.46$, $p = .01$) than the English language preference group.

Discussion

In this study, the one-factor, two-factor, and bifactor models of the PSS-10 were evaluated separately for the English and Spanish language preference groups. The one-factor model, which reflected Cohen and Williamson's (1988) original conceptualization of a unidimensional measure of perceived stress, showed poor fit for both language preference groups. The two-factor model, in which factor structure corresponded to item directionality (Factor 1: negatively worded items; Factor 2: positively worded items), provided acceptable fit in both language preference groups. The bifactor model, reflecting a two-factor structure of negatively and positively worded items with a single underlying general factor of perceived stress, provided adequate fit for the English language preference group but did not converge to the data for the Spanish language preference group. Consequently, the two-factor model was the only structure that provided acceptable fit in both English and Spanish language preference groups.

Using this two-factor model structure, multi-group CFA was conducted to test the measurement and structural invariance of the English and Spanish versions of the PSS-10. Results demonstrated configural, metric, and scalar invariance of the PSS-10 across language preference groups. Factor variance covariance invariance of the PSS-10 was not supported, indicating that the correlational relationship between the two latent factors was not equivalent across language preference groups. The lack of structural invariance, however, does not signify that the PSS-10 is inappropriate for use among English and Spanish language preference groups, as the critical prerequisite for cross-group comparisons is measurement invariance (metric and scalar invariance) (Cheung and Rensvold, 2002).

In addition to factorial validity, PSS-10 scores demonstrated good convergent validity. Scores on the PSS-10 were significantly correlated in the expected directions with scores on the GAD-7 and PHQ-9, such that greater levels of perceived stress (PSS-10 total scores) and negative stress feelings (PSS-10 "Negative" scores) were associated with greater reports of anxiety and depressive symptoms, while greater levels of positive stress feelings (PSS-10 "Positive" scores) were associated with less reports of anxiety and depressive symptoms, in both English and Spanish language preference groups. This is consistent with the convergent validity analyses from previous studies (e.g. Ramírez and Hernández, 2007; Remor, 2006; Roberti et al., 2006; Wu and Amtmann, 2013). These results provide further support for two distinct PSS-10 subscales that are differentially related to clinical syndromes, as "Negative" scores were more strongly related to GAD-7 and PHQ-9 than "Positive" scores. Interestingly, while "Positive" scores were significantly and negatively correlated with GAD-7 and PHQ-9 scores for both groups, as predicted, the strength of associations differed between language preference groups. Smaller correlations were found in the Spanish language preference group, indicating that perceived coping abilities and positive stress feelings ("Positive" subscale) were less associated with anxiety and depressive symptoms in Spanish-speaking participants than English-speaking participants. Despite increasing support for two distinct factors, most studies have used the PSS-10 as a unidimensional measure of perceived stress, arguing that the two factors reflect the format of the items rather than conceptual differences (e.g. Mitchell et al., 2008; Ramírez and Hernández, 2007; Reis et al., 2010) and used the PSS-10 total score rather than the two subscale scores. Additional research is needed to examine the convergent validity of the two PSS-10 factors and their relationships with theoretically related constructs (e.g. anxiety, depression, perceived health status) and to determine whether to use the subscale scores, total score, or both among English- and Spanish-speaking Hispanic Americans.

The results of this study should be interpreted within the context of study limitations. First, the study participants were predominantly Mexican American and recruited from one geographic region in Southern California, limiting the generalizability of findings to other Hispanic American subgroups. Therefore, the factorial validity and measurement invariance of the PSS-10 should be further examined in Hispanic Americans with differing characteristics from this study, so that findings can be generalized to the greater Hispanic American community. In addition, although this study found support for a distinct two-factor model in both language preference groups, the fit of the bifactor model could not be examined in the Spanish language preference group, precluding model comparison between the two-factor and bifactor models. This lack of model convergence may reflect issues with the model specification, specifically identification with one of the group factors, and/or that an underlying single factor of distress may not apply in this Spanish language preference sample. Given the superior fit of the bifactor model for the English language preference group, it may be possible that a unidimensional structure best supports the data for only English-speaking participants and two distinct factors ("Negative" and "Positive") of perceived stress best explain the data for Spanish-speaking participants. Future research should examine the fit and appropriateness of the bifactor model of the PSS-10, compared to the two-factor model, in larger, more diverse English- and Spanish-speaking samples. Furthermore, consistent with prior exploration of PSS-10 convergent validity, this study used two common and psychometrically sound measures of depression and anxiety. However, there are other variables that have been used for convergent validity in prior studies that future studies should also evaluate (e.g. emotional exhaustion).

Despite these limitations, this study is the first known study to evaluate the psychometric properties of the English and Spanish language versions of the PSS-10 in a sample of Hispanic American adults. The findings support the use of the PSS-10 as a reliable and valid measure of perceived stress in Hispanic Americans with either an English or Spanish language preference. Additionally, the study provides additional support for the PSS-10 two-factor model and use of the subscale scores.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by grants from the National Cancer Institute of the National Institutes of Health (R25CA130869, P30 CA023100, U56 CA92079/U56 CA92081, and U54 CA132379/U54 CA132384). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies. Sharon H. Baik was supported by the University of California, San Diego Fellowship.

References

American Psychiatric Association (APA) (2000) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)* (4th edn) text revision. Washington, DC: APA.

Barbosa-Leiker C, Kostick M, Lei M, et al. (2013) Measurement invariance of the Perceived Stress Scale and latent mean differences across gender and time. *Stress and Health* 29: 253–260.

Bentler PM (1990) Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238–246.

Bentler PM (2007) On tests and indices for evaluating structural models. *Personality and Individual Differences* 42: 825–829.

Brown TA (2013) Latent variable measurement models. In: Little TD (ed.) *The Oxford Handbook of Quantitative Methods: Statistical Analysis*, vol. 2. New York: Oxford University Press, pp. 257–280.

Chavez-Korell S and Torres L (2013) Perceived stress and depressive symptoms among Latino adults: The moderating role of ethnic identity cluster patterns. *The Counseling Psychologist* 42: 230–254.

Cheung GW and Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling* 9: 233–255.

Cohen S (2014) PSS: Frequently asked questions. Available at: http://www.psy.cmu.edu/~scohen/scales.html

Cohen S and Williamson G (1988) Perceived stress in a probability sample of the United States. In: Spacapam S and Oskamp S (eds) *The Social Psychology of Health*. Newbury Park, CA: SAGE, pp. 31–67.

Cohen S, Kamarck T and Mermelstein R (1983) A global measure of perceived stress. *Journal of Health and Social Behavior* 24: 385–396.

Corral I and Landrine H (2010) Methodological and statistical issues in research with diverse samples: The problem of measurement equivalence. In: Landrine H and Russo NF (eds) *Handbook* *of Diversity in Feminist Psychology*. New York: Springer, pp. 83–136.

Detjen MG, Nieto FJ, Trentham-Dietz A, et al. (2007) Acculturation and cigarette smoking among pregnant Hispanic women residing in the United States. *American Journal of Public Health* 97: 2040–2047.

Dr. Cohen's Scales (2015) Available at: http://www.psy.cmu.edu/~scohen/scales.html

Folkman S (1997) Positive psychological states and coping with severe stress. *Social Science & Medicine* 45: 1207–1221.

García-Campayo J, Zamorano E, Ruiz MA, et al. (2010) Cultural adaptation into Spanish of the Generalized Anxiety Disorder-7 (GAD-7) scale as a screening tool. *Health and Quality of Life Outcomes* 8: 8.

Geisinger KF (1994) Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment* 6: 304–312.

Golden-Kreutz DM, Browne MW, Frierson GM, et al. (2004) Assessing stress in cancer patients: A second-order factor analysis model for the Perceived Stress Scale. *Assessment* 11: 216–223.

Groth-Marnat G (2009) *Handbook of Psychological Assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Hu LT and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal: A Multidisciplinary Journal* 6: 1–55.

Huang FY, Chung H, Kroenke K, et al. (2006) Racial and ethnic differences in the relationship between depression severity and functional status. *Psychiatric Services* 57: 498–503.

IBM Corp. (2013) *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Jovanović V and Gavrilov-Jerković V (2015) More than a (negative) feeling: Validity of the Perceived Stress Scale in Serbian clinical and non-clinical samples. *Psihologija* 48: 5–18.

Merz EL, Malcarne VL, Roesch SC, et al. (2011) A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity & Ethnic Minority Psychology* 17: 309–316.

Mills SD, Fox RS, Malcarne VL, et al. (2014) The psychometric properties of the Generalized Anxiety Disorder-7 Scale in Hispanic Americans with English or Spanish language preference. *Cultural Diversity & Ethnic Minority Psychology* 20: 463.

Mitchell AM, Crane PA and Kim Y (2008) Perceived stress in survivors of suicide: Psychometric properties of the Perceived Stress Scale. *Research in Nursing & Health* 31: 576–585.

Muthén LK and Muthén BO (1998–2012) *Mplus User's Guide* (7th edn). Los Angeles, CA: Muthén & Muthén.

Ramírez MTG and Hernández RL (2007) Factor structure of the Perceived Stress Scale (PSS) in a sample from Mexico. *The Spanish Journal of Psychology* 10: 199–206.

Reis RS, Hino AA and Anez CR (2010) Perceived Stress Scale: Reliability and validity study in Brazil. *Journal of Health Psychology* 15: 107–114.

Remor E (2006) Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). *The Spanish Journal of Psychology* 9: 86–93.

Rios J and Wells C (2014) Validity evidence based on internal structure. *Psicothema* 26: 108–116.

Roberti JW, Harrington LN and Storch EA (2006) Further psychometric support for the 10-item version of the Perceived Stress Scale. *Journal of College Counseling* 9: 135–147.

Satorra A and Bentler PM (1994) Corrections to test statistics and standard errors in covariance structure analysis. In: Von Eye A and Clogg CC (eds) *Latent Variables Analysis: Applications for Developmental Research*. Thousands Oaks, CA: SAGE, pp. 399–419.

Satorra A and Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66: 507–514.

Segrin C and Badger TA (2013) Interdependent psychological distress between Latinas with breast cancer and their supportive partners. *Journal of Latina/o Psychology* 1: 21–34.

Spitzer RL, Kroenke K and Williams JBW (1999) Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA* 282: 1737–1744.

Spitzer RL, Kroenke K, Williams JBW, et al. (2006) A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine* 166: 1092–1097.

Steiger JH (1990) Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 25: 173–180.

Wu SM and Amtmann D (2013) Psychometric evaluation of the Perceived Stress Scale in multiple sclerosis. *ISRN Rehabilitation* 2013: 608356.