



Published in final edited form as:

Int Psychogeriatr. 2020 September ; 32(9): 1073–1084. doi:10.1017/S1041610220000502.

Evaluation of the Measurement Properties of the Perceived Stress Scale (PSS) in Hispanic Caregivers to Patients with Alzheimer’s Disease and Related Disorders

Jeanne A. Teresi, EdD, PhD^{1,2,3,4}, Katja Ocepek-Welikson, MPhil¹, Mildred Ramirez, PhD^{1,2,4}, Marjorie Kleinman, MS³, Katherine Ornstein, PhD, MPH⁵, Albert Siu, MD⁶, Jose Luchsinger, MD⁷

¹Research Division, Hebrew Home at Riverdale, 5901 Palisade Avenue, Riverdale, New York, 10471;

²Measurement and Data Management Core, Mount Sinai Pepper Older Americans Independence Center, Mount Sinai Medical Center, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029

³Analytic Core, Columbia University Alzheimer’s Disease Resource Center for Minority Aging Research and Stroud Center, New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY 10032

⁴Division of Geriatrics and Palliative Medicine, Weill Cornell Medical Center, 1300 York Avenue, New York, NY 10168

⁵Department of Geriatrics and Palliative Medicine, Institute for Translational Epidemiology Mount Sinai School of Medicine, New York, 1425 Madison Avenue, New York, NY 10029;

⁶Department of Geriatrics and Palliative Medicine, General Internal Medicine, Health Evidence and Policy, Mount Sinai Medical Center, 1 Gustave L. Levy Pl, New York, NY 10029 New York, NY, USA

⁷Columbia University Department of Medicine; PH9 Center, room 210; 630 West 168th Street; New York, NY, USA 10032;

Abstract

Objectives: The Perceived Stress Scale (PSS) is the most widely used measure of perceived stress; however, minimal psychometric evaluation has been performed among Hispanic respondents, and even less among Hispanic caregivers to persons with Alzheimer’s disease and related disorders (ADRD).

Correspondence regarding this article should be addressed to Jeanne Teresi, EdD, PhD, Research Division, Hebrew Home at Riverdale, 5901 Palisade Avenue, Riverdale New York 10471, teresimeas@aol.com; tel : 718 581 1139; fax : 718 543-2477.
Description of authors’ roles

Jeanne A. Teresi, the lead author of the manuscript, was responsible for the design and supervision of the analyses and wrote the background for the paper. Katja Ocepek-Welikson and Marjorie Kleinman contributed to analysis and interpretation of data. Mildred Ramirez, Katherine A. Ornstein and Albert L. Siu contributed to manuscript preparation. Jose A. Luchsinger, the senior author of the manuscript, was the principal investigator of the study, and supervised all aspects of study design and implementation.

Conflict of interest declaration
None

Design: Secondary data analysis.

Setting: New York City, NY, USA

Participants: a sample of 453 community dwelling Hispanic caregivers to patients with ADRD.

Measurements: Latent variable models were used to evaluate the PSS. Exploratory and confirmatory factor analyses were used to examine unidimensionality. Differential item functioning (DIF) was examined for age, education and language using the graded item response model.

Results: The factor and bifactor analyses results supported essential unidimensionality of the item set; however, positively worded items were observed using response item theory to be less informative than the negatively worded items. Reliability estimates were high. Salient DIF was not observed for age, education, or language of interview using the primary DIF detection method. Sensitivity analyses using a second DIF-detection method identified uniform language- DIF for the item, “In the last month, how often have you felt that you were on top of things?” However, the non-compensatory DIF value was below the threshold considered salient.

Conclusions: In summary, the 10-item PSS performed well in a sample of English and Spanish speaking Hispanic caregivers to patients with ADRD. Very little DIF, and none of high magnitude and impact was observed. However the negatively worded items, perhaps because they are more directly reflective of stress, were more informative. In the context of a short-form measure or computerized adaptive test, more informative items are those that would be selected for inclusion.

Keywords

stress measurement; differential item functioning; Latinx caregivers; dementia

Introduction

Perceived stress refers to appraisal of threats arising from a stressor for which individuals lack the resources to cope with the demands of the stressor (Lazarus and Folkman, 1984). Chronic stress has been defined as demanding and distressing experiences on a daily basis lasting for 6 or more months (Epel et al., 2018). Chronic stress exposure may be more prevalent among individuals from lower socio-economic status groups and among racial and ethnic minority groups who may experience discrimination (American Psychological Association [APA], 2017). One type of stressful exposure is caregiving to persons with dementia, including Alzheimer’s disease and related disorders (ADRD). Such caregiving can result in chronic stress because of the cumulative reaction to daily stressful demands. Stress has been found to relate to disease biomarkers (Blackburn et al., 2015) and overall adverse health and mental health outcomes (APA, 2017). Caregivers as contrasted with noncaregivers are more likely to have comorbid mental and physical health conditions (Pinquart and Sörensen, 2003) and among those reporting high caregiver emotional strain, increased mortality (Schulz and Beach, 1999).

The most widely used scale to measure perceived stress is the ten-item Perceived Stress Scale (PSS; Cohen et al., 1983). The original measure contained 14 items and was developed among English-speaking respondents, most of whom were college students. Because four

items did not perform well in a validation study with a probability sample of English-speakers in the United States of American, a ten item version was later developed (Cohen and Williamson, 1988), and is that most widely used. Typical items in the 10-item scale are: “In the last month, how often have you felt that you were unable to control the important things in your life?”; “In the last month, how often have you felt nervous and stressed?” The ten-item version was used in these analyses. The original English version by Cohen and Williamson (1988) and the Spanish version translated by Remor (2006) were used in the study. A word or a phrase was modified in 6 of the items of the Spanish version in order to make them more culturally appropriate for the Spanish speaking sample, mainly of Dominican descent.

Dimensionality: One issue raised in several studies is the dimensionality of the measure, and whether the item set is represented best as unidimensional (e.g., Cohen and Williamson, 1988; Mitchell et al., 2008) or as two subscales. One subscale is purportedly comprised of positively worded items, e.g., things going your way; able to control irritations in your life; felt on top of things, and the other negatively worded items, e.g., unable to control important things in your life; nervous and stressed; and difficulties were piling up so high that you could not overcome them (see Barbosa-Leiker et al., 2013; Golden-Kreutz et al., 2004). Negative items assess distress related to a stressor directly; whereas the positive items may measure coping and resilience. However, it has also been argued that despite the findings of two factors, the measure is essentially unidimensional (e.g., Cohen and Williamson, 1988; Mitchell, et al., 2008; Wu and Amtmann, 2013). The implication of such findings is that the positive and negative factors are measurement artifacts induced by the item format as positively and negatively worded. For example, the factor structure of the PSS in the Spanish language was examined with one, two, and bifactor models ($n = 5,176$; Perera et al., 2017). One general factor (10 items) was established; however, additional variance was explained by 4-item positively-worded factor. Taylor (2015) examining a confirmatory factor analysis (CFA) estimated with polychoric correlations found support for a two-factor solution. These authors concluded that the reverse-worded (positive) items may yield unique substantive information.

In terms of validity of the two factor solution with positively and negatively worded PSS items, Baik et al. (2017) studied a sample of community resident self-identified Hispanics; 210 were interviewed in English (average age 38.5) and 226 in Spanish (average age 46.2). They found that the negative subscale items correlated positively, while the positive item subscale correlated negatively with measures of depression and anxiety. Jiang et al. (2017) examined the 14-item PSS in a sample ($n = 663$) of White (65%), Black (28%) and other (7%) older people in the Einstein Aging Project; they found that the positive subscale was uniquely predictive of development of amnesic mild cognitive impairment.

Little psychometric evaluation has been performed on this measure among Hispanic respondents, and even less among Latino caregivers to persons with Alzheimer’s disease. One recent study (Deeken et al., 2018) examined the fourteen-item PSS in German patients with dementia and their caregivers using factor analyses. The results supported a two-factor solution corresponding to the positively and negatively worded items in both patients and caregivers.

Measurement equivalence: Measurement equivalence studies involve examining measurement invariance using factor analytic models and or differential item functioning (DIF). Both approaches assume an underlying latent trait, e.g., stress measured by a set of items. DIF using item response theory (IRT) models posit that items perform in the same manner across studied groups, conditional on the trait. In other words, the probability of response is the same across the trait measured for e.g., males and females, different age or education groups. Few studies have examined measurement equivalence of the PSS across different language and ethnic groups. Using a random sample of students in Mexico, the factor structure of the Spanish version (Remor, 2006) of the 14-item PSS was compared with that observed for the English version (Cohen & Williams, 1988). The internal consistency was adequate ($\alpha = .83$) and confirmatory factor analysis corroborated the two-factor structure with the first factor comprised of positive and the second of negative items (Ramírez & Hernández, 2007). In a heterogeneous sample of 440 Spanish adults, the European Spanish 14- and 10-item PSS (author translated) demonstrated adequate reliability; the internal consistencies were $\alpha = .81$ and $\alpha = .82$, respectively. The test-retest reliability estimates were $r = .73$ and $r = .77$, respectively (Remor, 2006).

Taylor (2015) examined DIF in the ten-item version of the PSS using ordinal logistic regression and found no salient gender bias in a national adult sample ($N = 1,236$; 56% female; $M\ age = 54.48$, $SD = 11.69$; 77% self-identified White, 17.31% Black and/or African American, 2.5% “other”). They also tested configural, metric, and scalar invariance using multiple-group CFA, and concluded that measurement invariance was established. They observed that the latent mean difference between males and females for the negatively phrased item set was significant, but no significant mean differences were observed on the factor with positively worded items. These authors also examined the information functions from item response theory separately for the positively and negatively worded items, and observed lower item-level information for one of the positively worded items, “ability to control irritations in life”. Perera et al. (2017) performed multi-group CFA, comparing Spanish and English groups using a bifactor model. They established configural, metric, scalar, and residual invariance. Baik et al. (2017) examined measurement equivalence of the PSS by language, and established configural, metric, and scalar invariance between English and Spanish speakers specifying a two-factor model.

A smaller number of studies have examined differential item functioning among ethnically diverse groups. In a nationally representative sample ($N = 2,264$), DIF was assessed for each of the 10 items for sex (59% female), race (85% White, 8% African American, 4% Hispanic, 3% other), and education (48% greater than a high school education). Two items functioned differently by ethnicity. White non-Hispanic respondents reported higher perceived stress. Four items exhibited DIF by sex, four items by education, and five of the 10 items displayed DIF. However, the author concluded that the 10-item scale was valid (Cole, 1999).

Finally, very few studies of the PSS have used item response theory latent variable models to examine DIF in ethnically diverse groups. In a study (Sharp et al., 2007) of 312 people with asthma, IRT was used to assess DIF by race (21% African American, 33% Hispanic) and literacy. Literacy groupings included 33% with lower literacy as measured by Rapid Estimate of Adult Literacy in Medicine (REALM; Davis et al., 1993) scores equivalent to an

that a single general trait explains most of the common variance but that group traits explain additional common variance for item subsets (Reise et al., 2010). Final models used orthogonal rotation, and polychoric correlations estimated in MPlus (Muthén and Muthén, 2011).

Additionally, the explained common variance (ECV) from a bifactor model was estimated in R (R Core Team, 2013; Revelle, 2015; Rizopoulos, 2009) and MPlus (Muthén and Muthén, 2011). Finally a CFA model with polychoric correlations was tested in MPlus (Muthén and Muthén, 2011) in order to confirm the general pattern of the loadings.

Generalized, standardized local dependency (LD) chi-square statistics (Chen and Thissen, 1997) were calculated to test the IRT assumption of local independence. IRT item parameters and LD statistics were estimated using Item Response Theory for Patient Reported Outcomes (IRTPRO), version 2.1 (Cai et al., 2011).

To evaluate the model fit the following statistics were used: the root mean square error of approximation (RMSEA) and the comparative fit index (CFI; Bentler, 1990). The general guidelines for the fit index are CFI > 0.95, and the RMSEA < 0.06. Model fit for the IRT models was examined using RMSEA from IRTPRO (Cai et al., 2011) software.

Reliability and Information

Cronbach's alpha and ordinal alpha based on polychoric correlations (Gadermann et al., 2012; Zumbo et al., 2007) were estimated. McDonald's (McDonald, 1999) omega total (ω_t), a reliability estimate based on the proportion of total common variance explained was calculated. Finally, IRT-based reliability measures were examined at selected points along the underlying latent continuum. IRT-based information functions were also examined. The latter are used in selection of short-form versions of measures and provide information about precision of the measure at selected points along the measure continuum.

In addition to the baseline data, stability of reliability and dimensionality results were examined over three follow-up waves of data ($n = 343, 301, \text{ and } 219$; See Appendix Table A1).

Differential Item Functioning

The graded response model (Samejima, 1969) was used for DIF detection in both primary and sensitivity analyses. The item characteristic curve that relates the probability of an item response to the underlying state, e.g., stress, measured by the item set is characterized by two parameters: a location (severity) parameter (denoted b), and a discrimination parameter, proportional to the slope of the curve (denoted a). An item shows DIF if people from different subgroups but at the same level of the attribute (denoted θ) have unequal probabilities of endorsement. Uniform DIF is detected when the b parameters differ and non-uniform DIF when the a parameters differ among groups. Group differences in IRT item parameters were examined using the Wald test (Lord, 1980), accompanied by magnitude measures. Orthogonal contrasts were used. The final p values were adjusted using Bonferroni (1936) methods.

Latent variable ordinal logistic regression analyses using the graded response model in lordif (Choi et al., 2011) was the DIF sensitivity analysis approach. This method uses IRT models to estimate the conditioning latent stress variable, and a logistic regression approach for DIF detection.

Evaluation of DIF magnitude and impact.

Expected item scores, estimated as the sum of the weighted (by the response category value) probabilities of scoring in each of the possible categories for the item were calculated. The non-compensatory DIF (NCDIF) index (Raju et al., 1995) in DFIT (Raju et al., 2009) was used to quantify the difference in the average expected item scores. Details of the methods are presented elsewhere (Kleinman and Teresi, 2016). Aggregate impact was evaluated by comparing expected scale score functions between groups.

Results

IRT Model Assumptions: Dimensionality

The principal components analyses were performed for the total sample ($n = 453$), for age 19 to 54 ($n = 255$), for age 60–92 ($n = 197$), for high school or less education ($n = 244$), for education greater than high school ($n = 201$) and for language of interview (English $n = 154$ and Spanish $n = 299$). The ratio of the first to the second eigenvalue was 4.8 for the total sample and ranged from 4.0 (for ages 60 to 92) to 5.5 (for ages 19 to 59). A cutoff of 4 in ratios of the first to the second eigenvalue has often been used to indicate unidimensionality. The explained variance for the first eigenvalue was 54% for the total sample and ranged from 50% to 57% (see Appendix Table A2). These results suggest that the measure is essentially unidimensional, an assumption underlying IRT models.

The bifactor CFA model additionally confirmed the unidimensionality of the item set. All items but one (PSS1 – “In the last month, how often have you been upset because of something that happened unexpectedly?”) evidenced higher loadings (λ 's) on the general factor compared with the group factors. The range of λ 's on the factor of the negatively worded items was from 0.05 to 0.34 and 0.84 for the first item, “upset because something happened unexpectedly”, and the range for the factor of positively worded items was 0.28 to 0.48. In contrast, the range of the loadings on the general factor was 0.51 to 0.82 (see Table 2). The model fit statistics were good: RMSEA = 0.044; CFI = 0.996.

The item loadings (λ 's) for the one-factor solution for the total sample ranged from 0.57 to 0.82. Examination of the two-factor solution factor structure matrix showed that all positively worded items loaded on the second factor with higher λ 's (0.64 to 0.82) than on the first; however, these loadings were similar to those of a one factor solution (0.57 to 0.71). Additionally, the loadings (λ 's) for the positively worded items were also relatively high on the first factor of the two-factor solution, ranging from 0.42 to 0.52. These results add confirmatory support for essential unidimensionality (see Appendix Table A3).

Evaluation of the LD statistics (not shown) within each of the subgroups examined showed no violations. All LD chi-square values were below the recommended cutoff of 10.

Reliability

The classical test theory reliability coefficient alpha was 0.88 (unstandardized and standardized) for the total sample. The corrected item-total correlations ranged from 0.48 for the item, “In the last month, how often have you felt confident about your ability to handle your personal problems?” (reverse-coded) to 0.73 for the item, “In the last month, how often have you felt nervous and stressed?” (see Table 3).

The internal consistency and dimensionality estimates for the baseline measurements calculated in the psych package in R, based on the polychoric correlations, were high across follow-up waves of data. The baseline values were ordinal alpha = 0.902, McDonald’s Omega Total = 0.904 and ECV = 68.335. The follow-up statistics were consistent with the baseline results with McDonald’s omegas of about 0.90 across four waves of data collection (see Appendix, Table A1).

Table 4 presents the IRT reliability estimates along the levels of the perceived stress attribute (denoted theta in IRT). The estimates were limited to the theta levels for which there were respondents. The estimates for the total sample and all subgroups were high from theta -1.2 to 2.0 ranging from 0.90 to 0.93. The lowest estimates were at theta -2.4 (0.75 to 0.82) and then at theta 2.4 to 2.8 (0.81 to 0.88) across all groups. The average reliability estimate for the total sample was 0.89 and ranged from 0.88 to 0.90 for subgroups (see Table 4). IRT model fit statistics estimated in IRTPRO were good (< 0.01 to 0.06; see Table 4).

Item and Scale Information

The test information function was slightly bimodal with the first peak of 11.83 at theta -0.4 and second of 11.94 at theta 1.2 (see Appendix, Figure A1). The two most informative items were: “In the last month, how often have you felt nervous and stressed?” (information = 2.04 at theta -0.8) and “In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?” (information = 1.99 at theta -0.4). All four positively- worded items provided the least amount of information, and the following two provided the lowest: “In the last month, how often have you felt confident about your ability to handle your personal problems?” and “In the last month, how often have you been able to control irritations in your life?” (information = 0.52 for both at theta 0.0 and -0.4 ; see Appendix, Figure A2).

Differential item functioning

DIF was performed for age, education, and language of interview. DIF was not significant using the Wald test evaluating the individual items against the DIF-free anchor set. The item, “In the last month, how often have you felt confident about your ability to handle your personal problems?” showed non-uniform DIF for the education groups only before the Bonferroni correction for multiple comparisons.

The second DIF method used in sensitivity analyses, based on ordinal logistic regression using *lordif* software in R (Choi et al., 2011) detected DIF for the language of the interview comparisons. Uniform DIF in the item, “In the last month, how often have you felt that you were on top of things?” was identified by the chi-square criterion and additionally by the

non-compensatory DIF criterion in R; however, the NCDIF value (0.0621) was below the threshold value (0.0960 for five response categories) to be considered salient (see Table 5 and Appendix Figure A3). In general, the magnitude and impact of DIF were minimal (see Appendix Figures A4 and A5).

Scale Distribution

Appendix Table A4 presents the distribution of the perceived stress summary score mapped to the estimated theta (θ) for the total sample. The median of the sum score distribution was between 27 and 28 ($\theta = 0.062$ and $\theta = 0.166$) while the median of the theta distribution was at $\theta = -0.041$ (sum score = 26). The number of cases was the highest between the sum score of 29 to 32 ($\theta = 0.270$ and $\theta = 0.583$).

Discussion

While some information is available regarding the performance of the PSS in Hispanic samples (Baik et al., 2017; Perera et al., 2017; Ramírez and Hernández, 2007; Remor, 2006), few studies have used latent variable models to study scale performance in caregivers to patients with ADRD.

High reliability estimates were observed across methods for all groups. The classical test theory estimate for the total sample was high (alpha of 0.89). The IRT-derived estimates were also high (0.88 to 0.90 across age, education, and language of interview subgroups). The IRT-reliabilities were also fairly high across the range of the stress theta distribution, with somewhat lower values at the less-stressed tail. Factor analyses-derived reliability estimates (McDonald's omega; McDonald, 1999) were about 0.90 across several waves of data.

Numerous investigators (e.g., Cohen et al., 1983; Deeken et al., 2018; Ezzati et al., 2014) have observed a better fit of a two factor model to that of a unidimensional model using the 14-item version. Similar results were found with the 10-item scale (Baik et al., 2017; Barbosa-Leiker et al., 2013; Reis et al., 2010; Taylor, 2015). Often the two-factors were comprised of positively and negatively worded items. Methods for investigation of whether positively and negatively worded items are reflective of different underlying dimensions or are measurement artifacts have been the subject of numerous methodological studies in other fields and include examination of different types of factor models (Maydeu-Olivares and Coffman, 2006; Meredith and Teresi, 2006; Reise, 2012; Teresi et al., 2017). In this study a bifactor approach was used to examine the IRT essential unidimensionality assumption.

In the current study of the 10-item scale among a sample of Hispanic caregivers to patients with Alzheimer's disease and related disorders, it was observed that all items loaded > 0.50 on the general factor from a bifactor model. Similar to the findings of Wu and Amtmann (2013), we found that the measure was essentially unidimensional and that a bifactor model fit adequately, producing loadings on the general factor larger than those on the domain-specific group factors. Additionally, other indices of dimensionality, the eigenvalue ratios and ECV suggested an essentially unidimensional factor, supporting the premise that there may be a methodological artifact related to the direction of the item wording.

Few studies have represented different ethnic and racial groups adequately, and very few studies of measurement equivalence across different ethnic/ racial and language subgroups exist. The current study of the PSS adds to the literature by examining Hispanic caregivers to patients with ADRD. Generally, modern psychometric methods such as item response theory have not been applied, although latent variable models such as factor analyses have been used to examine dimensionality and measurement invariance. For example, in the study by Baik et al. (2017), evidence supported a 2-factor model in samples of Hispanics (interviewed in English and in Spanish). Between groups invariance: configural (number of factors), metric (factor loadings), and scalar (intercepts) was observed. In one study of the Perceived Stress Scale that did use IRT (Sharp et al., 2007), most items showed DIF in a sample of African American adults with low literacy. In the current study, little differential item functioning was observed. IRT Wald tests identified no DIF of high magnitude for age, education, or language of interview and only one item was significant before the correction for multiple comparisons: ability to handle your personal problems. The sensitivity analyses method, ordinal logistic regression with IRT-based parameter estimates identified three items with DIF for language of interview; only one evidenced higher magnitude, but below the threshold for salient DIF. The DIF impact was low. Items with DIF were: felt that things were going your way, felt you were on top of things (higher DIF magnitude), difficulties were piling up. Two of the items with DIF were positively worded. Despite evidence of minor DIF, the item-level magnitude and impact of DIF on the scale scores were minimal. No items were flagged for salient DIF with either the primary or sensitivity analyses.

The study by Jiang et al. (2017) observed that the positive subscale was more predictive of amnesic mild cognitive impairment (MCI) than the negatively worded item subscale. It could be argued that the positive items may be more reflective of coping and resilience. However, in this study, the positive items provided low information as estimated with IRT models. While overall scale-level information was high, some items were not informative. The most informative items were: felt nervous and stressed, difficulties were piling up so high that you could not overcome them, and unable to control the important things in your life. Uninformative items were all positively worded: handle your personal problems, control irritations in your life, things were going your way, and on top of things. Taylor (2015) also observed low information for the item “control irritations in your life”. Similar to previous analyses of positive and negative affect and subjective well-being (Teresi et al., 2017), the positively worded items were less informative, indicating that they are less useful in operationalizing a trait. In other contexts, such as computerized adaptive testing applications, such items would not be among those selected first for administration. They also would not be among those selected for short-form versions. An example of the development of a short-form measure of stress resilience using an item bank can be found in Obbarius et al. (2018).

Limitations of the Study

The analyses have some limitations, including the smaller sample size for those interviewed in English, which may have affected power for DIF detection. Additionally, respondents were Hispanic of Dominican, Puerto Rican, or Mexican descent; however, the sample size did not permit analyses of Hispanic subgroups. Finally, the sample size for males was too

small to examine gender DIF, and DIF was not examined for the Hispanic group as contrasted with a reference group of another ethnic background. An additional limitation is that only measurement properties were examined in this paper, and validity data were not included because such analysis was beyond the scope of this paper. Thus, it was not possible to examine correlates of stress such as caregiver burden and reaction or behaviors of the person with dementia or other characteristics of the caregiver or patient. However, a study of validity will be the focus of future research with this measure in this sample.

Contributions of the Study

This study is the first to examine the PSS in Hispanic caregivers to patients with ADRD. The PSS is increasingly being used in caregiver research and in intervention studies. If results are to be compared with studies with English speakers or with other Hispanic samples, it is important to provide guidance regarding whether such comparisons are substantively legitimate or if differences observed could be due to bias. This paper adds to the small but growing literature (e.g., Deeken, et al., 2018) about the performance of the PSS in dementia patients and their caregivers.

Conclusions and Recommendations

In conclusion, the 10-item PSS was found to perform well among a sample of English and Spanish speaking Hispanic caregivers to patients with Alzheimer's disease and related disorders. Very little DIF, and none of high magnitude and impact was observed. However, the negatively worded items, perhaps because they are more directly reflective of stress, were more informative. In the context of using an item bank to construct a short-form measure or computerized adaptive test, such more informative items are those that would be selected for inclusion. Given the growing Hispanic population and burden of ADRD-related disease experienced by these individuals and their caregivers, it is important to examine the performance of stress measures that are used clinically and in intervention research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Some of these analyses were presented at the preconference on stress and resilience hosted by the Resource Centers for Minority Aging Research at the Annual Meeting of the Gerontological Society of America, November 14, 2018, Boston, Massachusetts.

Support for these analyses was provided by a collaborative effort of the Mount Sinai Claude D. Pepper Older Americans Independence Center (National Institute on Aging, 1P30AG028741, Siu) and the Columbia University Alzheimer's Disease Resource Center for Minority Aging Research (National Institute on Aging, 1P30AG059303, Manly, Luchsinger). Data were from the New York City Hispanic Caregiver Research Program (National Institute of Nursing Research, NINR; 1R01NR0114430, Luchsinger).

References

American Psychological Association, APA Working Group on Stress and Health Disparities (2017). Stress and health disparities: Contexts, mechanisms, and interventions among racial/ethnic minority

and low-socioeconomic status populations. Retrieved from <http://www.apa.org/pi/health-disparities/resources/stress-report.aspx>.

- Baik SH et al. (2017). Reliability and validity of the Perceived Stress Scale-10 in Hispanic Americans with English or Spanish language preference. *Journal of Health Psychology*, 24, 628–639. doi:10.1177/1359105316684938 [PubMed: 28810432]
- Barbosa-Leiker C. et al. (2013). Measurement invariance of the Perceived Stress Scale and latent mean differences across gender and time. *Stress and Health*, 29, 253–260. doi:10.1002/smi.2463 [PubMed: 23027679]
- Bentler PM (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. doi: 10.1037/0033-2909.107.2.238 [PubMed: 2320703]
- Blackburn EH, Epel ES and Lin J. (2015). Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science*, 350(6265), 1193–1198. doi: 10.1126/science.aab3389. [PubMed: 26785477]
- Bonferroni CE (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3–62.
- Cai L., Thissen D. and du Toit SHC (2011). IRTPRO: Flexible, multidimensional, multiple categorical IRT Modeling [Computer software]. Chicago, IL: Scientific Software International, Inc.
- Chen WH and Thissen D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265–289. doi: 10.3102/10769986022003265
- Choi SW, Gibbons LE and Crane PK (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression / item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39, 1–30. doi: 10.18637/jss.v039.i08
- Cohen S, Kamarck T. and Mermelstein R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396. doi: 10.2307/2136404 [PubMed: 6668417]
- Cohen S. and Williamson G. (1988). Perceived stress in a probability sample of the United States. In: Spacapan S., Oskamp S. (Eds.), *The social psychology of health: Claremont symposium on applied social psychology* (pp. 31–67). Newbury Park, CA: Sage.
- Cole RS (1999). Assessment of differential item functioning in the Perceived Stress Scale-10. *Journal of Epidemiology and Community Health*, 53(5), 319–320. doi: 10.1136/jech.53.5.319 [PubMed: 10396541]
- Davis TC et al. (1993). Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Family Medicine*, 25, 391–395. PMID:8349060 [PubMed: 8349060]
- Deeken E., Hausler A., Nordheim J, Rapp M, Knoll N. and Rieckman N. (2018). Psychometric properties of the Perceived Stress Scale in a sample of German dementia patients and their caregivers. *International Psychogeriatrics*, 30, 39–47. doi: 10.1017/S1041610217001387 [PubMed: 28737120]
- Epel E. et al. (2018). More than a feeling: A unified view of stress measurement for population science. *Frontiers in Neuroendocrinology*, 49, 146–169. doi:10.1016/j.yfrne.2018.03.001 [PubMed: 29551356]
- Ezzati A., Jiang J., Katz MJ, Sliwinski MJ, Zimmerman ME and Lipton RB (2014). Validation of the Perceived Stress Scale in a community sample of older adults. *International Journal of Geriatric Psychiatry*, 29, 645–652. doi: 10.1002/gps.4049 [PubMed: 24302253]
- Gadermann AM, Guhn M. and Zumbo BD (2012). Estimating ordinal reliability of Likert-type and ordinal response data: A conceptual, empirical and practical guide. *Practical Assessment, Research and Evaluation*, 17, 1–13.
- Golden-Kreutz DM, Browne MW, Frierson GM and Andersen BL (2004). Assessing stress in cancer patients: A second-order factor analysis model for the Perceived Stress Scale. *Assessment*, 11, 216–223. doi:10.1177/1073191104267398 [PubMed: 15358877]
- Jiang JM, Seng EK, Zimmerman ME, Sliwinski M., Kim M. and Lipton RB (2017). Evaluation of the reliability, validity, and predictive validity of the subscales of the Perceived Stress Scale in older adults. *Journal of Alzheimer's Disease*, 59(3): 987–996. doi: 10.3233/JAD-170289
- Kleinman M. and Teresi JA (2016). Differential item functioning magnitude and impact measures from item response theory models. *Psychological Test and Assessment Modeling*. 58(1), 79–98. [PubMed: 28706769]

- Lazarus RS and Folkman S. (1984). *Stress, appraisal and coping*. New York, NY: Springer.
- Lord FM (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Luchsinger JA et al. (2018). Comparative effectiveness of 2 interventions for Hispanic caregivers of persons with dementia. *Journal of the American Geriatrics Society*, 66, 1708–1715. doi: 10.1111/jgs.15450 [PubMed: 30084133]
- Luchsinger JA et al. (2016). The Northern Manhattan Hispanic Caregiver intervention Effectiveness Study: Protocol of a pragmatic randomized trial comparing the effectiveness of two established interventions for informal caregivers of persons with dementia. *BMJ Open*, 6(11), e014082. doi:10.1136/bmjopen-201601408
- Luchsinger J. et al. (2015). Characteristics and mental health of Hispanic dementia caregivers in New York City. *American Journal of Alzheimer's Disease and Other Dementias*, 30(6), 584–590. doi:10.1177/1533317514568340
- Maydeu-Olivares A. and Coffman DL (2006). Random intercept item factor analysis. *Psychological Methods*, 11, 344–362. doi: 10.1037/1082-989X.11.4.344 [PubMed: 17154751]
- McDonald RP (1999). *Test theory: a unified treatment*. Mahwah, NJ: L. Erlbaum Associates.
- Meredith W. and Teresi JA (2006). An essay on measurement and factorial invariance. *Medical Care*, 44(Suppl 3), S69–S77. doi:10.1097/01.mlr.0000245438.73837.89 [PubMed: 17060838]
- Mitchell AM, Crane PA and Kim Y. (2008). Perceived stress in survivors of suicide: Psychometric properties of the Perceived Stress Scale. *Research in Nursing and Health*, 31, 576–585. 10.1002/nur.20284 [PubMed: 18449942]
- Muthén LK and Muthén BO (1998-2011). *MPlus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén and Muthén.
- Obbarius N., Fisher F., Obbarius A., Nolte S., Liegl G. and Rose M. (2018). A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample. *Journal of Clinical Epidemiology*, 100, 1–12. doi:10.1016/j.jclinepi.2018.04.004 [PubMed: 29653200]
- Perera MJ et al. (2017). Factor structure of the Perceived Stress Scale-10 (PSS) across English and Spanish language responders in the HCHS/SOL Sociocultural Ancillary Study. *Psychological Assessment*, 29(3), 320–328. doi:10.1037/pas0000336. [PubMed: 27280744]
- Pinquart M. and Sörensen S. (2003). Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. *Psychology and Aging*, 18(2), 250–67. doi: 10.1037/0882-7974.18.2.250 [PubMed: 12825775]
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Raju NS, Fortmann-Johnson KA, Kim W., Morris SB, Nering M., and Oshima LTC, (2009). The item parameter replication method for detecting differential functioning in the DFIT framework. *Applied Measurement in Education*, 33, 133–147. doi: 10.1177/0146621608319514
- Raju NS, van der Linden WJ and Fleer PF (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19, 353–368. doi: 10.1177/014662169501900405.
- Ramírez M. and Hernández R. (2007). Factor structure of the Perceived Stress Scale (PSS) in a sample from Mexico. *The Spanish Journal of Psychology*, 10, 199–206. doi:10.1017/S1138741600006466 [PubMed: 17549893]
- Reis RS, Hino AAF and Añez CRR (2010). Perceived Stress Scale: Reliability and validity study in Brazil. *Journal of Health Psychology*, 15(1), 107–114. 10.1177/1359105309346343 [PubMed: 20064889]
- Reise SP (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. doi:10.1080/00273171.2012.715555 [PubMed: 24049214]
- Reise SP, Moore TM and Haviland MG (2010). Bi-factor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544–559. doi:10.1080/00223891.2010.496477 [PubMed: 20954056]
- Reise S., Morizot J. and Hays R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(Suppl 1), 19–31. doi:10.1007/s11136-007-9183-7 [PubMed: 17479357]

- Remor E. (2006). Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). *The Spanish Journal of Psychology*, 9, 86–93. doi:10.1017/S1138741600006004 [PubMed: 16673626]
- Revelle W. (2015). Psych: package Psych. Retrieved from <http://cran.r-project.org/package=psych>
- Rizopoulos D. (2009). ltm: Latent Trait Models under IRT. <http://cran.r-project.org/web/packages/ltm/index.html>.
- Samejima F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 100–114. doi: 10.1007/BF02290599
- Schulz R. and Beach SR (1999). Caregiving as a risk factor for mortality: The Caregiver Health Effects Study. *Journal of the American Medical Association*, 282(23), 2215–2219. doi: 10.1001/jama.282.23.2215 [PubMed: 10605972]
- Sharp LK, Kimmel LG, Kee R., Saltoun C. and Chang C. (2007). Assessing the Perceived Stress Scale for African American adults with asthma and low literacy. *Journal of Asthma*, 44, 311–316. doi:10.1080/02770900701344165
- Taylor JM (2015). Psychometric analysis of the ten-item Perceived Stress Scale. *Psychological Assessment*, 27, 90–101. doi:10.1037/a0038100 [PubMed: 25346996]
- Teresi JA et al. (2017). Methodological issues in measuring subjective well-being and quality-of-life: Applications to assessment of affect in older, chronically and cognitively impaired, ethnically diverse groups using the Feeling Tone Questionnaire. *Applied Research in Quality of Life*, 12(2), 251–288. doi:10.1007/s11482-017-9516-9 [PubMed: 30220935]
- Wu SM and Amtmann D. (2013). Psychometric evaluation of the Perceived Stress Scale in multiple sclerosis. *ISRN Rehabilitation*, 2013. 10.1155/2013/608356.
- Zumbo BD, Gadermann AM and Zeisser C. (2007). Ordinal versions of coefficient alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6, 21–29. Available at: <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/4>

Table 1.

Demographic Characteristics of the Caregiver Sample

Language of the interview			
	English (<i>n</i> = 154; 34%)	Spanish (<i>n</i> = 299; 66%)	Total (<i>n</i> = 453)
Gender			
Female	118 (77%)	263 (88%)	381 (84%)
Male	36 (23%)	36 (12%)	72 (16%)
Age			
Age 19 to 59	110 (71%)	145 (49%)	255 (56%)
Age 60 to 92	44 (29%)	153 (51%)	197 (44%)
Mean (SD)	54.8 (10.0)	60.3 (11.4)	58.4 (11.2)
Missing	0	1	1
Education			
High school or less	56 (37%)	188 (64%)	244 (55%)
Post high school	95 (63%)	106 (36%)	201 (45%)
Mean number of years (SD)	14.3 (3.0)	11.7 (3.9)	12.6 (3.8)
Missing	3	5	8

Table 2.Item Loadings from the Bifactor Model CFA Solution (MPlus) Total Sample ($N = 453$)

Item name	Item description	Bifactor solution		
		FG λ (SE)	F1 λ (SE)	F2 λ (SE)
PSS1	In the last month, how often have you been upset because of something that happened unexpectedly?	0.58 (0.05)	(0.84 (0.16))	
PSS2	In the last month, how often have you felt that you were unable to control the important things in your life?	0.75 (0.03)	(0.19 (0.06))	
PSS3	In the last month, how often have you felt nervous and stressed?	0.80 (0.03)	(0.22 (0.07))	
PSS4R [†]	In the last month, how often have you felt confident about your ability to handle your personal problems?	0.51 (0.04)		0.44 (0.05)
PSS5R	In the last month, how often have you felt that things were going your way?	0.59 (0.04)		0.28 (0.05)
PSS6	In the last month, how often have you felt that you could not cope with all the things you had to do?	0.70 (0.03)	0.05 (0.06)	
PSS7R	In the last month, how often have you been able to control irritations in your life?	0.53 (0.04)		0.41 (0.05)
PSS8R	In the last month, how often have you felt that you were on top of things?	0.65 (0.03)		0.48 (0.05)
PSS9	In the last month, how often have you been angered because of things that happened that were outside of your control?	0.68 (0.04)	0.34 (0.09)	
PSS10	In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	0.82 (0.02)	0.13 (0.06)	

[†]R in the item name indicates the reversal of the original response categories to align their direction with the rest of the items measuring perceived stress

Model fit statistics:

Root Mean Error of Approximation (RMSEA) = 0.044; Comparative Fit Index (CFI) = 0.996

Table 3.Classical Test Reliability Analysis Using Cronbach's Alpha. Total sample ($N = 453$)

Item name	Item description	Mean (SD)	Corrected item-total correlation	Alpha if item deleted
PSS1	In the last month, how often have you been upset because of something that happened unexpectedly?	2.88 (1.23)	0.61	0.87
PSS2	In the last month, how often have you felt that you were unable to control the important things in your life?	2.64 (1.28)	0.67	0.87
PSS3	In the last month, how often have you felt nervous and stressed?	3.17 (1.26)	0.73	0.86
PSS4R	In the last month, how often have you felt confident about your ability to handle your personal problems?	2.19 (1.08)	0.48	0.88
PSS5R	In the last month, how often have you felt that things were going your way?	2.81 (1.15)	0.53	0.88
PSS6	In the last month, how often have you felt that you could not cope with all the things you had to do?	2.76 (1.19)	0.60	0.87
PSS7R	In the last month, how often have you been able to control irritations in your life?	2.34 (1.06)	0.50	0.88
PSS8R	In the last month, how often have you felt that you were on top of things?	2.47 (1.11)	0.62	0.87
PSS9	In the last month, how often have you been angered because of things that happened that were outside of your control?	2.9 (1.15)	0.64	0.87
PSS10	In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	2.58 (1.25)	0.71	0.86
Reliability: Coefficient alpha unstandardized and (standardized)		0.883 (0.881)		

Table 4.

Item Response Theory (IRT) Reliability Estimates at Varying Levels of the Attribute (Theta) Estimate Based on Results of the IRT Analysis (IRTPRO) and Fit Statistics

IRT Reliability							
PSS (Theta)	Total sample (n = 452)	Age 19 – 59 (n = 255)	Age 60 – 92 (n = 197)	Education 0 years to HS (n = 244)	Education beyond HS (n = 201)	English interview (n = 154)	Spanish interview (n = 299)
-2.4	0.77	0.78	0.77	0.75	0.79	0.82	0.75
-2.0	0.84	0.85	0.83	0.83	0.85	0.88	0.82
-1.6	0.89	0.90	0.87	0.88	0.89	0.91	0.87
-1.2	0.91	0.92	0.90	0.91	0.91	0.93	0.90
-0.8	0.92	0.93	0.91	0.93	0.92	0.93	0.92
-0.4	0.92	0.93	0.91	0.93	0.92	0.93	0.92
0.0	0.92	0.93	0.91	0.92	0.92	0.93	0.92
0.4	0.92	0.93	0.91	0.92	0.91	0.93	0.92
0.8	0.92	0.93	0.91	0.93	0.92	0.93	0.92
1.2	0.92	0.93	0.91	0.93	0.92	0.93	0.92
1.6	0.92	0.93	0.90	0.92	0.92	0.91	0.92
2.0	0.90	0.91	0.89	0.89	0.90	0.89	0.90
2.4	0.87	0.87	0.86	0.86	0.88	0.84	0.87
2.8	0.83	0.83	0.82	0.81	0.84	0.81	0.84
Overall (average)	0.89	0.90	0.88	0.89	0.89	0.90	0.88
IRT RMSEA *	0.06	0.05	0.05	0.05	0.06	<0.01	0.05

Note: Reliability estimates were calculated for theta levels for which there were respondents. HS = high school.

*RMSEA = Root Mean Error of Approximation based on M₂ statistics (on full marginal tables).

Table 5.

Differential Item Functioning (DIF) Summary

Item name	Item description	IRT Wald Test			lordif results		
		Age	Education	Interview language	Age	Education	Interview language
PSS1	In the last month, how often have you been upset because of something that happened unexpectedly?						
PSS2	In the last month, how often have you felt that you were unable to control the important things in your life?						
PSS3	In the last month, how often have you felt nervous and stressed?						
PSS4R	In the last month, how often have you felt confident about your ability to handle your personal problems?		NU				
PSS5R	In the last month, how often have you felt that things were going your way?						NU X
PSS6	In the last month, how often have you felt that you could not cope with all the things you had to do?						
PSS7R	In the last month, how often have you been able to control irritations in your life?						
PSS8R	In the last month, how often have you felt that you were on top of things?						U X NCDIF= 0.0621
PSS9	In the last month, how often have you been angered because of things that happened that were outside of your control?						
PSS10	In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?						U X

U = Uniform DIF; NU=Non-Uniform DIF; no items showed DIF after the Bonferroni correction

X = Identified using Chi-Square criterion (< 0.01)

NCDIF = Non-compensatory DIF index; NCDIF threshold for 5 response categories = 0.0960