# Cross-Cultural Confirmatory Factor Analysis of the CES-D in Spanish and Mexican Dementia Caregivers

Andrés Losada[1], Mª de los Ángeles Villareal[2], Roberto Nuevo[3], María Márquez-González[4],
Bertha C. Salazar[5], Rosa Romero-Moreno[1], Ana L. Carrillo[2], and Virginia Fernández-Fernández[1]

[1]Universidad Rey Juan Carlos (Spain)
[2]Universidad Autónoma de Coahuila (Mexico
[3]Hospital Universitario de Getafe (Spain)
[4]Universidad Autónoma de Madrid (Spain)
[5]Universidad Autónoma de Nuevo León (Mexico)

The Center for Epidemiologic Studies-Depression Scale (CES-D) is the most frequently used scale for measuring depressive symptomatology in caregiving research. The aim of this study is to test its construct structure and measurement equivalence between caregivers from two Spanish-speaking countries. Face-to-face interviews were carried out with 595 female dementia caregivers from Madrid, Spain, and from Coahuila, Mexico. The structure of the CES-D was analyzed using exploratory and confirmatory factor analysis (EFA and CFA, respectively). Measurement invariance across samples was analyzed comparing a baseline model with a more restrictive model. Significant differences between means were found for 7 items. The results of the EFA clearly supported a four-factor solution. The CFA for the whole sample with the four factors revealed high and statistically significant loading coefficients for all items (except item number 4). When equality constraints were imposed to test for the invariance between countries, the change in chi-square was significant, indicating that complete invariance could not be assumed. Significant between-countries differences were found for three of the four latent factor mean scores. Although the results provide general support for the original four-factor structure, caution should be exercised on reporting comparisons of depression scores between Spanish-speaking countries.
*Keywords: Alzheimer, caregiving, dementia, depression, Spanish-speaking, Latino.*

La Escala del Centro para Estudios Epidemiológicos (CES-D) es la más utilizada para medir sintomatología depresiva en la investigación sobre cuidadores. El objetivo de este estudio es analizar su estructura dimensional y su equivalencia de medida entre cuidadores de dos países de habla hispana. Se entrevistó a 595 mujeres cuidadoras de familiares con demencia de Madrid, España, y Coahuila, México. La estructura del CES-D se analizó mediante análisis factoriales exploratorios (AFE) y confirmatorios (AFC). Se encontraron diferencias estadísticamente significativas entre las medias de 7 ítems. Los resultados del AFE con la muestra total respaldan claramente una solución de 4 factores. El AFC de la estructura de 4 factores revela cargas elevadas y significativas para todos los ítems (excepto el 4). Al imponer restricciones de igualdad para valorar la equivalencia entre países, el cambio en chi-cuadrado fue significativo, indicando que no se puede asumir una equivalencia completa. Se encontraron diferencias significativas entre países para tres de las cuatro puntuaciones medias en los factores latentes. Aunque los resultados proporcionan un apoyo general a la estructura original de cuatro factores, se debería tener una cierta cautela a la hora de informar sobre comparaciones en puntuaciones en depresión entre países de habla hispana.
*Palabras clave: Alzheimer, cuidadores, demencia, depresión, hispano hablantes, latinos.*

Dementia is a public health priority worldwide. The number of people suffering from dementia is estimated at 35.6 million (Alzheimer's Disease International, 2009). Prevalence rates of dementia for over-65s in Spain range from 7.4% to 17.2% (Berr, Wancata, & Ritchie 2005), while the figure for Mexico is around 8.6% (Llibre et al., 2008). Caring for an elderly relative is associated with high levels of reported distress by caregivers, especially if they are caring for a relative with a diagnosis of dementia (Pinquart & Sörensen, 2003). Depression is one of the principal emotional disorders suffered by caregivers (Schulz, O´Brien, Bookwala, & Fleissner, 1995). A frequently reported finding is that female caregivers have higher depression scores than male caregivers (Pinquart & Sörensen, 2003).

The Center for Epidemiologic Studies-Depression Scale (CES-D) is one of the most frequently used scales in caregiving studies (Pinquart & Sörensen, 2003; Schulz et al., 1995). Although it has been recommended as a measure for psychological outcomes in caregiving research (Brodaty, 2007), further validation of this scale has been required in settings outside North America (Moniz-Cook et al., 2008), and there is a need for assessing measure equivalence across cultural groups so as to advance the study of the influence of culture on the caregiving process (Knight & Sayegh, 2010).

The CES-D was developed by Radloff (1977) with a non-Spanish-speaking sample for assessing depressive symptomatology in the general, non-clinical, population. Through exploratory factor analysis four factors were retained, labelled "depressed affect", "positive affect", "somatic and retarded activity", and "interpersonal". Subsequent studies, using confirmatory factor analyses, have provided support for this four-factor structure (e.g., Knight, Williams, McGee, & Olaman, 1997; Williams et al., 2007). These studies were mostly carried out with North-American or non-European samples.

To our knowledge, although several studies have been carried out with Spanish-speaking samples aimed at analyzing the psychometric properties of the CES-D, neither in Spain (Calvete et al., 2005; Soler et al., 1997), México (Aguilera-Guzmán, Carreño-García, & García, 2004; Golding & Aneshensel, 1999; Reyes-Ortega et al., 2003) nor in other Spanish-speaking countries (Campo-Arias, Díaz-Martínez, Rueda-Jaimes, Cadena-Afanador, & Hernández 2007) have there been any confirmatory factor analyses of the CES-D, except in the McCauley et al. (2006) study, which used a mixed sample (Hispanics, African-American and Euro-American), and the Golding and Aneshensel (1989) study, which compared US-born Mexican-Americans with Mexican-born Mexican-Americans and European Americans. Both studies reported data in support of the four-factor structure of the CES-D. Also, Golding and Aneshensel (1989) failed to find evidence of measurement invariance of the CES-D across groups, and considered the CES-D items according to an interval scale, ignoring their ordinal nature.

Likewise, and even though the CES-D is the most widely used scale in caregiving research, studies aimed at confirming the CES-D factor structure are also few in number. To our knowledge, only two confirmatory studies of this scale have been conducted with caregivers, one in Canada (O'Rourke, 2005) and another in the USA (Roth, Ackerman, Okonkwo, & Burgio, 2008). Despite the fact that the Roth et al. (2008) study includes a group of 208 Hispanic caregivers, it does not differentiate between caregivers of Hispanic origin born in the USA and those of Hispanic origin born outside the USA but currently living there. These two studies confirm the original four-factor structure of the scale, pointing out that the structure is similar for men and women (O'Rourke, 2005) and for different cultural groups such as European American, African-American and Hispanic caregivers (Roth et al., 2008), though in the latter study differences by race were found for 3 of the 4 CES-D latent factors.

The aim of the present work is to analyze the construct structure and measurement equivalence of the CES-D in two Spanish-speaking groups with different cultural backgrounds (i.e., Spain and Mexico). Our hypothesis is that the original structure proposed by Radloff (1977) will be confirmed, and that the CES-D structure will be found to be invariant across countries. We were particularly interested in testing between samples two types of invariance, following Vandenberg and Lance's (2000) method: configural and metric invariance.

Method

*Participants*

The sample was made up of 595 female dementia caregivers (411 from Madrid, Spain, and 184 from Saltillo, Coahuila, a State in Northeastern Mexico). Both samples were recruited in urban settings. Similar procedures for contacting caregivers were used in Spain and in Mexico. Caregivers were recruited through announcements in social and healthcare centres and in the media (newspapers, radio, and television), inviting participation in broader studies aimed at testing the efficacy of psycho-educational interventions (e.g., Losada, Márquez-González, & Romero-Moreno, 2011). The data used in this study came from baseline assessments. In order to participate in the study, caregivers needed to identify themselves as the primary source of help for a relative with Alzheimer's disease or related dementia, and must have been caring for their

Table 1
*Characteristics of the sample*

| | Spain | Mexico | V or *g** | *P* |
|---|---|---|---|---|
| *Caregivers' age* | | | | |
| Mean | 57.18 | 51.98 | | |
| S.D. | 12.17 | 9.67 | 0.45 (.28/.63)** | < .001 |
| Range | 18-98 | 30-82 | | |
| *Educational level (in years)* | | | | |
| Mean | 9.68 | 10.90 | −.25 | |
| S.D. | 5.53 | 3.81 | (−.43/−.08) | .013 |
| Range | 0-26 | 0-20 | | |
| *Time as caregiver (in years)* | | | | |
| Mean | 4.15 | 4.58 | −.13 | |
| S.D. | 3.17 | 3.22 | (−.31/.04) | .196 |
| Range | 0,8-17 | 1-16 | | |
| *Care recipients' age* | | | | |
| Mean | 78.17 | 78.21 | .01 | |
| S.D. | 8.82 | 7.86 | (−.18/.17) | .783 |
| Range | 48-97 | 55-96 | | |
| *Care recipient (%)* | | | | |
| Spouse | 31.1 | 17.4 | | |
| Parent | 58.6 | 74.5 | .154 | .003 |
| Parent-in-law | 4.4 | 4.3 | | |
| Other | 5.8 | 3.8 | | |

*V: Cramer's V for the overall association in the categorical variable; *g*: Hedges' effect size.
**95% Confidence Intervals for Hedge's *g* effect sizes.

relative for at least 3 months. Descriptive data of the sample are shown in Table 1.

*Measures*

The Center for Epidemiologic Studies-Depression Scale (CES-D) (Radloff, 1977; Mexican validation by González, Stewart, Ritter, & Lorig, 1995; Spanish validation by Losada, Peñacoba, Márquez-González, & Cigarán, 2008; see Annex 1) was developed as a measure of depressive symptomatology in general population. This instrument has 20 items selected from the items included in previously validated depression scales. For each item, respondents are required to indicate the frequency of occurrence of the symptom on a Likert-type scale ranging from 0 (rarely or none of the time – less than one day) to 3 (most or all of the time – 5-7 days). Total score ranges from 0 to 60. Through principal components analysis four factors were retained, labelled "depressed affect"; "positive affect"; "somatic and retarded activity"; and "interpersonal".

*Data analysis*

Comparisons between the two samples were performed in the main socio-demographic variables, employing t-tests (and effect sizes; Hedge's *g*) for continuous variables (age, educational level) and chi-square tests (with Cramer's V as effect size) for categorical variables.

Before the different analyses of the CES-D scores, tests were carried out for the detection of outliers. Next, descriptive analysis and statistical comparisons (*chi*-square) were carried out in order to test for differences across groups in the specific items of the CES-D. Differences between each of the CES-D items were calculated by means of t-tests for independent samples. Unbiased effect sizes (Hedges' *g*) for the differences were then calculated.

Analysis of the CES-D structure in terms of configural invariance was undertaken by first performing an exploratory factor analysis of the estimated polychoric correlation matrix (correlation index for categorical variables), separately for each country and then for the whole sample. We used the

robust weighted least square means and variance adjusted estimator (WLSMV) and a geomin oblique rotation (Muthén & Muthén, 1998-2006). Fit indices (according to explained values) were used to assess the adequacy of models and data for factor analysis.

We then tested both the configural and metric equivalence of the scale through multi-group CFA for categorical outcomes, using the WLSMV estimator and polychoric correlations. Tests for normality were not carried out because no assumptions are required about the distribution of the indicators on using the WLMSV estimator in MPlus (Flora & Curran, 2004; Muthén & Asparouhov, 2002). The model suggested in the exploratory analyses was assumed for the CFA in order to analyze for structure invariance across samples, and a baseline model (configural invariance) was then compared with a model in which constrictions were applied (metric invariance).

Thus, in a first baseline model, item thresholds and factor loadings were freely estimated in both groups. Scale factors were fixed to 1, and factor means were fixed to 0. A second model was estimated in which factor loadings and item thresholds were constrained to be equal across groups. In this case, scale factors were set to 1 in the Spanish group and freely estimated in the Mexican group, and factor means were fixed to 0 in the Spanish group and freely estimated in the Mexican group. Note that factor loadings and thresholds have to been taken in tandem because the item characteristic curve is influenced by both parameters, and scale factors have to be included in the model as they take into account possible differences in variances across groups (Muthén & Asparouhov, 2002; Muthén & Christoffersson,1981). See Muthén & Muthén (1998-2006) for an explanation of this procedure in the Mplus frame. This second model is nested in the previous unconstrained model, and therefore the change in fit allows for a statistical test of invariance across groups using the DIFFTEST procedure implemented in the Mplus software (Muthén & Muthén, 1998-2006).

According to the usual recommendations (Hu & Bentler, 1999; Schreiber, Stage, King, Nora, & Barlow, 2006; Yu, 2002; or see Ximénez & Revuelta, 2010 for a recent example of research on measurement invariance), and specifically considering the values proposed by Schreiber et al. (2006) for assessing fit in structural equation modeling (SEM) with categorical outcomes, several fit indices were used: a) lack of significance of $\chi^2$; b) comparative fit index (CFI; >.96); c) Tucker-Lewis index (TLI; >.96); d) root mean square error of approximation (RMSEA; < .05); and e) weighted root mean square residual (WRMR; < 1.0).

All factor analysis modeling was carried out using Mplus Version 4.21 (Muthén & Muthén, 1998-2006). Preliminary data summaries and preliminary examination of differences between countries on the individual items were carried out using Stata 11.0 (Stata Statistical Software Release, 2009).

## Results

### Sample descriptive data

Spanish caregivers are older than Mexican caregivers ($t = -5.40$; $p < .001$) and have lower educational level ($t = 2.68$; $p = .013$). Significant differences between countries were also found for the relationship with the care recipient ($\chi^2 = 14.1$; $p = .003$). More spouse caregivers are found in the Spanish sample, while in the Mexican sample more caregivers take care of their parents. There were no statistical significant differences between samples in time as caregiver ($p = .196$), or in age of the care recipient ($p = .783$). Effect sizes for these comparisons were in general low, indicating that the samples are comparable in terms of their main characteristics. See Table 1 for specific descriptive data and effect size.

The presence of potential outliers was tested according to Cook's distance, considering as potential cases values of more than 1 (Cook & Weisberg, 1983). According to this criterion, no case was considered to be an outlier.

### Differences across groups in item means

Differences between individual items across samples are shown in Table 2. After Bonferroni adjustment for multiple comparisons in order to control for family-wise Type I error (significance level was adjusted to .05 / 21 = .0024), there were statistically significant differences in 7 items, but not for the total score. The effect sizes were in general small, the maximum ranging from .01 to .70, with only three comparisons with effect sizes higher than .50. The average effect size for these comparisons was .24, indicating that there were no large or relevant differences between samples in item mean scores.

### Exploratory factor analysis

The results of the Exploratory Factor Analysis clearly supported a four-factor solution for both countries in separate analyses. In both cases, the first four eigenvalues were over 1, and far higher than the rest. A scree plot produced an 'elbow' after the fourth factor. Specifically, in the Spanish sample, the first four factors explained 63.11% of the variance with the following eigenvalues: 8.67 (43.4%); 1.66 (8.3%); 1.27 (6.4%); and 1.01 (5.1%), and after fitting a four-factor model all the fit indices showed adequate fit: $\chi^2(116) = 138.7$, $p = .074$; CFI = .997; TLI = .995; RMSEA = .022. Likewise, in the Mexican sample, the first four factors explained 64.14% of the variance with the following eigenvalues: 8.57 (42.9%); 1.60 (8.0%); 1.41 (7.0%); and 1.25 (6.3%), and after fitting a four-factor model all the fit indices showed adequate fit: $\chi^2(116) = 210.6$, $p < .001$; CFI = .966; TLI = .945; RMSEA = .067. Three- or five-factor models

Table 2

*Descriptive data for CES-D items in Mexican and Spanish caregivers, with comparison tests and effect size estimates*

| Items | Mexico | | Spain | | $t$ | $p$ | $g$ |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | | |
| 1. Bothered by things | 0.69 | 0.81 | 0.53 | 0.84 | 2.2 | .027 | .19 |
| 2. Appetite was poor | 0.57 | 0.86 | 0.49 | 0.91 | 0.9 | .353 | .09 |
| 3. Can't shake the blues | 0.94 | 0.94 | 1.01 | 1.10 | −0.7 | .431 | −.07 |
| 4. Just as good as others | 0.96 | 1.18 | 1.36 | 1.28 | −3.7 | <.001 | −.33 |
| 5. Trouble concentrating | 0.98 | 0.94 | 0.93 | 1.03 | 0.5 | .592 | .05 |
| 6. Felt depressed | 1.19 | 1.05 | 1.17 | 1.13 | 0.2 | .860 | .02 |
| 7. Everything was an effort | 0.84 | 0.97 | 0.95 | 1.07 | −1.3 | .203 | −.11 |
| 8. Hopeful about the future | 1.20 | 1.16 | 1.98 | 1.08 | −8.0 | <.001 | −.70 |
| 9. Life has been a failure | 0.31 | 0.69 | 0.41 | 0.79 | −1.4 | .150 | −.14 |
| 10. Fearful | 0.51 | 0.87 | 0.74 | 0.96 | −2.7 | .005 | −.26 |
| 11. Sleep was restless | 1.17 | 1.06 | 1.16 | 1.18 | 0.1 | .898 | .01 |
| 12. Happy | 1.05 | 1.00 | 1.58 | 0.95 | −6.2 | <.001 | −.55 |
| 13. Talked less than usual | 0.65 | 0.92 | 0.58 | 0.93 | 0.9 | .368 | .08 |
| 14. Lonely | 0.82 | 1.00 | 0.91 | 1.11 | −1.0 | .285 | −.08 |
| 15. People were unfriendly | 0.46 | 0.86 | 0.18 | 0.50 | 5.1 | <.001 | .44 |
| 16. Enjoyed life | 0.93 | 1.07 | 1.63 | 1.08 | −7.3 | <.001 | −.65 |
| 17. Crying spells | 0.76 | 0.91 | 0.98 | 1.06 | −2.4 | .010 | −.22 |
| 18. Sad | 1.01 | 0.96 | 1.36 | 1.09 | −3.7 | <.001 | −.33 |
| 19. People disliked me | 0.51 | 0.88 | 0.27 | 0.66 | 3.6 | <.001 | .33 |
| 20. Could not get going | 0.86 | 1.00 | 0.92 | 1.05 | −0.6 | .538 | −.06 |
| TOTAL CES-D | 16.5 | 10.8 | 19.4 | 11.5 | −2.8 | .005 | −.26 |

*g*: Hedges' *g*, unbiased estimate of the effect size

showed adequate but poorer fit in both countries. Overall, items loaded in the factor suggested in the previous literature. Finally, we tested the combined sample, obtaining very similar results: 62.06% of variance explained for the first four factors, and after fitting a four-dimensional model, all the fit indices showed adequate fit: $c^2(116) = 145.8$, $p = 0.032$; CFI = .998; TLI = .997; RMSEA = .025. We conclude that a four-dimensional model provided an adequate summary of the data. The pattern of loadings mostly reproduced previous findings of a four-factor structure for the CES-D (Radloff, 1977). Therefore, we assumed this structure for further confirmatory analyses and tests of invariance across groups.

## Confirmatory Factor Analysis (CFA)

The CFA for the whole sample with the four factors revealed high and statistically significant loading coefficients for all items, except for item 4 ("just as good as others": standardized coefficient of .136). For the baseline multi-group CFA analysis (without equality constraints in factor loadings and thresholds), the model presented a moderate-to-good fit: $\chi^2(126) = 310.4$, $p < .001$; CFI = .952; TLI = .980; RMSEA = .070; WRMR = 1.321. All regression coefficients in both groups were positive, statistically significant and moderate-to-high, except in the case of item

4. These results are shown in Table 3 (unstandardized coefficients are not reported, but are available from the corresponding author on request). This result permits us to consider the configural invariance between the two samples as acceptable, with the exception of item 4, which is working inadequately in both samples.

When factor loadings and thresholds were constrained to be equal in order to test for the metric invariance between Spanish and Mexican caregivers, the change in chi-square was significant ($\Delta\chi^2(26) = 98.2$, $p < .001$), indicating that complete invariance could not be assumed. Moderate fit indices were found for this final model: CFI = .952; TLI = .973; RMSEA = .080; WRMR = 1.794. Spanish caregivers scored significantly higher on the latent mean of Depressed Affect (the mean difference obtained by the SEM model attains 0.179 points in favor of the Spanish group; $SE = .104$; $p = .028$), and on the latent mean of Positive Affect, with 0.124 points in favor of the Spanish group ($SE = .045$; $p = .006$). Likewise, the Mexican group scored significantly higher on the latent mean of Interpersonal, with .421 points in their favor ($SE = .093$; $p < .001$).

Given the negative results obtained for item 4 in both samples, the fit of a model with that item removed was analyzed. The deterioration in fit between unconstrained and constrained models was in this case only marginally

Table 3

*Unconstrained and constrained standardized factor loadings for the CES-D four-factor model in Mexican and Spanish caregivers*

| Dimension | Items | Unconstrained | | Constrained |
|---|---|---|---|---|
| | | Mexico | Spain | |
| Somatic and Retarded Activity | 1. Bothered by things | .490 (.072) | .472 (.049) | .455 (.042) |
| | 2. Appetite was poor | .606 (.075) | .505 (.051) | .514 (.044) |
| | 5. Trouble concentrating | .671 (.057) | .514 (.044) | .539 (.038) |
| | 7. Everything was an effort | .625 (.061) | .712 (.033) | .657 (.038) |
| | 11. Sleep was restless | .548 (.062) | .561 (.041) | .532 (.038) |
| | 13. Talked less than usual | .498 (.069) | .689 (.041) | .604 (.040) |
| | 20. Could not get going | .695 (.058) | .862 (.025) | .785 (.041) |
| Depressed Affect | 3. Can't shake the blues | .733 (.039) | .862 (.020) | .848 (.023) |
| | 6. Felt depressed | .853 (.026) | .881 (.018) | .892 (.022) |
| | 9. Life has been a failure | .869 (.042) | .656 (.045) | .749 (.035) |
| | 10. Fearful | .735 (.055) | .628 (.038) | .677 (.035) |
| | 14. Lonely | .847 (.033) | .761 (.029) | .804 (.025) |
| | 17. Crying spells | .828 (.032) | .736 (.027) | .776 (.024) |
| | 18. Sad | .910 (.020) | .930 (.012) | .949 (.021) |
| Positive Affect | 4. Just as good as others | .254* (.085) | .039** (.066) | .138* (.049) |
| | 8. Hopeful about the future | .541 (.065) | .613 (.045) | .598 (.039) |
| | 12. Happy | .782 (.047) | .868 (.028) | .824 (.035) |
| | 16. Enjoyed life | .916 (.043) | .818 (.032) | .845 (.034) |
| Interpersonal | 15. People were unfriendly | .709 (.084) | .939 (.085) | .667 (.044) |
| | 19. People disliked me | .794 (.078) | .593 (.069) | .844 (.070) |

Values in brackets are standard errors.

All factor loadings significant with $p < .001$, except *, $p = .005$, and **, $p = .556$.

significant: $(\Delta\chi^2(21) = 35.1, p = .028)$, with moderate-to-adequate goodness-of-fit indices: CFI = .955; TLI = .976; RMSEA = .079; WRMR = 1.741. This model of partial metric equivalence was therefore considered adequate.

## Discussion

The aim of this study was to test the CES-D factor structure in Spanish-speaking caregivers from different cultural backgrounds (Spain and Mexico). The results provide general support for the original four-factor structure of the CES-D, as proposed by Radloff (1977), and as reported in other studies carried out with caregiver samples (Roth et al., 2008). Thus, the results of this study provide support for the consideration of the same dimensions or factors of depression in studies carried out with Spanish-speaking caregivers from different cultural backgrounds.

However, the results of this study also suggest a need for caution on reporting comparisons in depression scores between Spanish-speaking groups. Several between-countries differences were found in this study. First, significant albeit small differences were found in 7 item mean scores. In addition, our results do not provide support for full metric invariance of the CES-D factor structure across countries, and differences in latent mean scores were also found in three of the four CES-D dimensions (no differences were found in the Somatic factor). These results suggest, in a similar way to the findings from previous studies (e.g., Knight & Sayegh, 2010), the importance of further analyzing the influence of culture in the caregiving process, and also provide support for the notion that even though Spanish-speaking subgroups may share nuclear cultural values, important differences may exist between them (Aranda & Knight, 1996). Although the psychometric analyses conducted for this study do not provide clues as to why certain items function differently across the groups analyzed, several differences may exist between caregivers from different (although culturally similar) backgrounds such as Spain and Mexico. Caregivers may attribute different meanings or senses to certain words or items, or "higher order" cultural norms or values (e.g., familism)

may be introducing biases in their responses. Regarding the use of the CES-D with Spanish-speaking populations, our results also suggest that both in the Spanish sample and in the Mexican sample item number 4 does not contribute well to the fit of the factor structure, so that its deletion might be considered. A similar result was found in a comparison of Hispanic samples with arthritis in US, in which the suppression of the item 4 was proposed (González et al., 1995). In any case, even when some differences were found in paired comparisons between items, the model of partial metric equivalence in which only one item had been considered as not invariant suggests that the interpretation of the meanings of items was very similar across samples, and the scales are comparable.

This study has several limitations that may influence the generalization of the results. First, different versions from the CES-D validated to Spanish were used in each country. Although each of the versions presented good psychometric properties (González et al., 1995; Losada et al., 2008), the discrepancy in findings may be due to differences in the wording of the items. Second, although no between-sex differences in the CES-D factor structure have been found in previous studies (O'Rourke, 2005), this study was carried out with only female Spanish-speaking dementia caregivers, so that the results may not be generalizable to male caregivers or caregivers of relatives with non-dementia problems. Also, although the Spanish sample size can be considered adequate for the type of analysis carried out, the Mexican sample size could be considered small. Furthermore, neither sample was recruited from the general caregiver population through random procedures, so that caregivers who do not seek help or participate in research studies such as the one reported here might be under-represented. Finally, although similar procedures for contacting caregivers were used in the two countries, non-controlled site differences (for example, cultural factors that may influence caregivers' intention to participate in studies) may have introduced sample biases that could affect the results.

Taken together, the results of this study are of interest, given that, to our knowledge, this is the first study in which the CES-D factor structure has been compared using two Spanish-speaking samples from different countries. Spanish-speakers have usually been considered as a homogeneous group (for example, in the only available study with caregivers from the USA, in which the CES-D structure was analyzed using CFA) (Roth et al., 2008), despite research suggesting that there may be significant differences between Spanish-speaking subgroups (Hilgeman et al., 2009; Losada et al., 2006). Treating Spanish-speakers as a single group, rather than as subgroups that might appraise the caregiving situation differently, may limit the generalization capacity of studies (Hilgeman et al., 2009). Our study points in this direction, suggesting the need to analyze in more depth the possible differences between Spanish-speaking

subgroups with regard to how culture may influence emotions such as sadness (e.g., how they are experienced and verbalized). Although this study contributes to the important research objective of providing culturally appropriate instruments that permit the comparison of studies across cultures, further research is needed that can provide information regarding the reasons why differences between item meanings may be found between groups with similar cultural backgrounds. The use of focus groups or the implementation of research specifically designed to analyze the invariance of assessment instruments may be helpful for achieving this aim.

References

Aguilera-Guzmán, R. M., Carreño-García, M. S., & García, F. J. (2004). Características psicométricas de la CES-D en una muestra de adolescentes rurales mexicanos de zonas con alta tradición migratoria [Psychometric properties of the CES-D in a sample of Mexican rural adolescents from areas with strong migratory tradition]. *Salud Mental, 27*, 57–66.
Alzheimer's Disease International. (2009). *Alzheimer's Disease International. World Alzheimer Report*. London, England: Author Retrieved from http://www.alz.co.uk/research/files/WorldAlzheimerReport.pdf
Aranda, M. P., & Knight, B. G. (1996). The influence of ethnicity and culture on the caregiver stress and coping process: A sociocultural review and analysis. *The Gerontologist, 37*, 342–354. http://dx.doi.org/10.1093/geront/37.3.342
Berr, C., Wancata, J., & Ritchie, K. (2005). Prevalence of dementia in the elderly in Europe. *European Neuropsychopharmacology, 15*, 463–471. http://dx.doi.org/10.1016/j.euroneuro.2005.04.003
Brodaty, H. (2007). Meaning and measurement or caregiver outcomes. *International Psychogeriatrics, 19*, 363–381. http://dx.doi.org/10.1017/S1041610207005182
Calvete, E., Estévez, A., Landín., C., Martínez, Y., Cardeñoso, O., Villardón, L., & Villa A. (2005). Self talk and affective problems in college students: Valence of thinking and cognitive content specificity. *The Spanish Journal of Psychology, 8*, 56–67.
Campo-Arias, A., Díaz-Martínez, L. A., Rueda-Jaimes, G. E., Cadena-Afanador, L. P., & Hernández, N. L. (2007). Psychometric properties of the CES-D Scale among Colombian adults from the general population. *Revista Colombiana de Psiquiatría, 36*, 664–674.
Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika, 70*, 1–10. http://dx.doi.org/10.1093/biomet/70.1.1
Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491. http://dx.doi.org/10.1037/1082-989X.9.4.466

Golding, J. M., & Aneshensel, C. S. (1989). Factor structure of the Center for Epidemiologic Studies Depression Scale among Mexican Americans and Non-Hispanic whites. *Psychological Assessment, 1,* 163–168. http://dx.doi.org/10.1037//1040-3590.1.3.163

González, V., Stewart, A., Ritter, P., & Lorig, K. (1995). Translation and validation of arthritis outcome measures into Spanish. *Arthritis and Rheumatism, 38*, 1429–1446. http://dx.doi.org/10.1002/art.1780381010

Hilgeman, M. M., Durkin, D. W., Sun, F., DeCoster, J., Allen, R. S, Gallagher-Thompson, D., & Burgio, L. D. (2009). Testing a theoretical model of the stress process in Alzheimer's caregivers with race as a moderator. *The Gerontologist, 49*, 248–261. http://dx.doi.org/10.1093/geront/gnp015

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling, 6*, 1–55. http://dx.doi.org/10.1080/10705519909540118

Knight, B. G., & Sayegh, P. (2010). Cultural values and caregiving: Updated sociocultural stress and coping model. *Journals of Gerontology: Psychological Sciences, 65B*, 5–13. http://dx.doi.org/10.1093/geronb/gbp096

Knight, R. G., Williams, S., McGee, R., & Olaman, S. (1997). Psychometric properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behaviour Research and Therapy, 35*, 373–380. http://dx.doi.org/10.1016/S0005-7967(96)00107-6

Llibre Rodriguez, J. J., Ferri, C. P., Acosta, D., Guerra, M., Huang, Y., Jacob, K. S., ... 10/66 Dementia Research Group (2008). Prevalence of dementia in Latin America, India, and China: A population-based cross-sectional survey. *Lancet, 372,* 464–474. http://dx.doi.org/10.1016/S0140-6736(08)61002-8

Losada, A., Márquez-González, M., & Romero-Moreno, R. (2011). Mechanisms of action of a psychological intervention for dementia caregivers: Effects of behavioral activation and modification of dysfunctional thoughts. *International Journal of Geriatric Psychiatry, 26,* 1119–1127. http://dx.doi.org/10.1002/gps.2648

Losada, A., Peñacoba, C., Márquez-González, M., & Cigarán, M. (2008). *Cuidar Cuidándose. Evaluación e intervención interdisciplinar con cuidadores familiares de personas con demencia* [Caregiving caring of yourself. Assessment and multidisciplinary intervention for family caregivers of people with dementia]. Madrid, Spain: Obra Social Caja Madrid.

Losada, A., Robinson Shurgot, G., Knight, B. G., Márquez, M., Montorio, I., Izal, M., & Ruiz, M.A. (2006). Cross-cultural study comparing the association of familism with burden and depressive symptoms in two samples of Hispanic dementia caregivers. *Aging and Mental Health, 10*, 69–76. http://dx.doi.org/10.1080/13607860500307647

McCauley, S. R., Pedroza, C., Brown, S. A., Boake, C., Levin, H. S., Goodman, H. S., & Merritt, S. G. (2006). Confirmatory factor structure of the Center for Epidemiologic Studies-Depression Scale (CES-D) in mild-to-moderate traumatic brain injury. *Brain Injury, 20*, 519–527. http://dx.doi.org/10.1080/02699050600676651

Moniz-Cook, E., Vernooij-Dassen, M., Woods, R., Verhey, F., Chattat, R., de Vugt, ... INTERDEM group (2008). A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging and Mental Health,* 12, 14–29. http://dx.doi.org/10.1080/13607860801919850

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 4.* Retrieved from http://www.statmodel.com/download/webnotes/CatMGLong.pdf

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419. http://dx.doi.org/10.1007/BF02293798

Muthén, L., & Muthén, B. (1998-2006). *Mplus user's guide* (4th Ed.). Los Angeles, CA: Author.

O'Rourke, N. (2005). Factor structure of the Center for Epidemiologic Studies-Depression Scale (CES-D) among older men and women who provide care to persons with dementia. *International Journal of Testing, 5*, 265–277. http://dx.doi.org/10.1207/s15327574ijt0503_5

Pinquart, M., & Sörensen, S. (2003). Differences between caregivers and noncaregivers in psychological health and physical health: A meta-analysis. *Psychology and Aging, 18*, 250–267. http://dx.doi.org/10.1037/0882-7974.18.2.250

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. http://dx.doi.org/10.1177/014662167700100306

Reyes-Ortega, M., Soto-Hernández, A. L., Milla-Kegel, J. G., García-Ramírez, A., Hubard-Vignau, L., Mendoza-Sánchez, H., ... Wagner-Echegaray, F. A. (2003). Actualización de la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D). Estudio piloto en una muestra geriátrica Mexicana [Updating of the Center for Epidemiologic Studies-Depression Scale (CES-D). Pilot study in a geriatric Mexican sample]. *Salud Mental, 26*, 59–68.

Roth, D. L., Ackerman, M. L., Okonkwo, O. C., & Burgio, L. D. (2008). The four-factor model of depressive symptoms in dementia caregivers: A structural equation model of ethnic differences. *Psychology and Aging, 23*, 567–576. http://dx.doi.org/10.1037/a0013287

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting Structural Equation Modelling and Confirmatory Factor Analysis results: A review. *The Journal of educational research, 99*, 323–338. http://dx.doi.org/10.3200/JOER.99.6.323-338

Schulz, R., O´Brien, A. T., Bookwala, J., & Fleissner, K. (1995). Psychiatric and physical morbidity effects of dementia caregiving: Prevalence, correlates, and causes. The *Gerontologist, 35*, 771–791. http://dx.doi.org/ 10.1093/geront/35.6.771

Soler, J., Pérez-Sola, V., Puigdemont, D., Pérez-Blanco, J., Figueres, M., & Alvarez, E. (1997). Validation study of the Center for Epidemiologic Studies-Depression (CES-D) in affective disorders in Spanish population. *Actas Españolas de Psiquiatría, 25*, 243–249.

Stata Statistical Software Release 11.0. (2009). Stata Statistical Software Release 11.0. College Station, TX: StataCorp.

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

Williams, C. D., Taylor, T. R., Makambi, K., Harrell, J., Palmer, J. R., Rosenberg, L., & Adams-Campbell, L. L. (2007). CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Research, 150*, 173–180. http://dx.doi.org/10.1016/j.psychres.2006.02.007

Ximénez, C., & Revuelta, J. (2010). Factorial invariance in a repeated measures design: an application to the study of person-organization fit. *The Spanish Journal of Psychology, 13*, 485–493.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models withbinary and continuous outcomes*. Los Angeles, CA: University of California.

# APPENDIX

## Original, Spanish and Mexican versions of the CES-D

| Original version | Spanish version | Mexican version |
|---|---|---|
| 1. I was bothered by things that usually don't bother me | 1. Me molestaron cosas que habitualmente no me molestan | 1. Me molestaron cosas que normalmente no me molestan |
| 2. I did not feel like eating; my appetite was poor | 2. No tuve hambre; tenía poco apetito | 2. No me sentía con ganas de comer; no tenía apetito |
| 3. I felt that I could not shake off the blues even with help from my family or friends | 3. Sentía que no podía librarme de la tristeza incluso con la ayuda de mi familia o amigos | 3. Me sentía que no podía quitarme de encima la tristeza aún con la ayuda de mi familia |
| 4. I felt that I was just as good as other people | 4. Sentí que era, al menos, tan bueno como otras personas | 4. Sentía que yo era tan bueno/a como cualquier otra persona |
| 5. I had trouble keeping my mind on what I was doing | 5. Tuve problemas para concentrarme en lo que hacía | 5. Tenía dificultad en mantener mi mente en lo que hacía |
| 6. I felt depressed | 6. Me sentí deprimido | 6. Me sentía deprimido |
| 7. I felt that everything I did was an effort | 7. Sentí que todo lo que hacía era un esfuerzo | 7. Sentía que todo lo que hacía era un esfuerzo |
| 8. I felt hopeful about the future | 8. Me sentí optimista sobre el futuro | 8. Me sentía optimista sobre el futuro |
| 9. I thought my life had been a failure | 9. Pensé que mi vida había sido un fracaso | 9. Pensé que mi vida había sido un fracaso |
| 10. I felt fearful | 10. Me sentí temeroso | 10. Me sentía con miedo |
| 11. My sleep was restless | 11. Mi sueño era inquieto, no descansaba | 11. No podía dormir bien |
| 12. I was happy | 12. Estaba contento | 12. Estaba contento/a |
| 13. I talked less than usual | 13. Hablaba menos de lo habitual | 13. Hablé menos de lo usual |
| 14. I felt lonely | 14. Me sentí solo/a | 14. Me sentía solo/a |
| 15. People were unfriendly | 15. La gente me resultaba antipática | 15. Pensaba que la gente no era amistosa |
| 16. I enjoyed life | 16. Disfruté la vida | 16. Disfruté de la vida |
| 17. I had crying spells | 17. Lloré en ocasiones | 17. Pasé ratos llorando |
| 18. I felt sad | 18. Me sentí triste | 18. Me sentía triste |
| 19. I felt that people dislike me | 19. Sentí que no le gustaba a la gente | 19. Sentía que yo no le caía bien (gustaba) a la gente |
| 20. I could not get "going" | 20. No tenía ganas de nada | 20. No tenía ganas de hacer nada |