Measurement Properties of the Center for Epidemiologic Studies Depression Scale (CES-D 10): Findings from HCHS/SOL

Patricia González[1], **Alicia Nuñez**[1], **Erin Merz**[2], **Carrie Brintz**[3], **Orit Weitzman**[3], **Elena Navas**[4], **Alvaro Camacho**[5], **Christina Buelna**[1], **Frank J. Penedo**[6], **Sylvia Wassertheil-Smoller**[7], **Krista Perreira**[8], **Carmen Isasi**[7], **James Choca**[9], **Gregory A. Talavera**[1], and **Linda C. Gallo**[1]

[1]Graduate School of Public Health, San Diego State University

[2]Department of Psychology, California State University, Dominguez Hills

[3]Department of Psychology, University of Miami

[4]School of Public Health, University of Illinois, Chicago

[5]Department of Psychology, University of California San Diego

[6]Department of Medical Social Sciences, Northwestern University

[7]Department of Epidemiology and Public Health, Albert Einstein College of Medicine

[8]Department of Public Policy, University of North Carolina, Chapel Hill

[9]Department of Psychology, Roosevelt University

[10]SDSU/UCSD Joint Doctoral Program in Clinical Psychology

Abstract

The Center for Epidemiologic Studies Depression Scale (CES-D) is a widely used self-report measure of depression symptomatology. This study evaluated the reliability, validity, and measurement invariance of the CES-D 10 in a diverse cohort of Hispanics/Latinos from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). The sample consisted of 16,415 Hispanic/Latino adults recruited from four field centers (Miami, FL; San Diego, CA; Bronx, NY; Chicago, IL). Participants completed interview administered measures in English or Spanish. The CES-D 10 was examined for internal consistency, test-retest reliability, convergent validity, and measurement invariance. The total score for the CES-D 10 displayed acceptable internal consistencies (Cronbach α's = .80 – .86) and test-retest reliability (r's = .41 – .70) across the total sample, language group and ethnic background group. The total CES-D 10 scores correlated in a theoretically consistent manner with the Spielberger State-Trait Anxiety Inventory ($r = .72$, $p < .001$), the Patient Health Questionnaire-9 depression measure ($r = .80$, $p < .001$) the Short Form-12's Mental Component Summary ($r = -.65$, $p < .001$) and Physical Component Summary score ($r = -.25$, $p < .001$). A confirmatory factor analysis showed that a one-factor model fit the CES-D 10 data well (CFI = .986, RMSEA = .047) after correlating one pair of item residual variances. Multiple group analyses showed the one-factor structure to be invariant across

English and Spanish speaking responders and partially invariant across Hispanic/Latino background groups. The total score of the CES-D 10 can be recommended for use with Hispanics/Latinos in English and Spanish.

Keywords

CES-D 10; Hispanics/Latinos; reliability; validity; measurement invariance

---

Depression is widely recognized as a costly and potentially debilitating illness among adults in the United States (US), with nearly 1 in 10 meeting criteria for a depressive disorder (Centers for Disease Control and Prevention (CDC), 2010). Affective functioning, a person's mood, and emotional well-being are important components of overall health. Depression symptoms are associated with worse physical health and can adversely affect risk and outcomes in a variety of chronic diseases, including cardiovascular disease (CVD), cancer, diabetes, and obesity (CDC, 2014). Research suggests that depressive symptoms may differ across Hispanic/Latino background groups. For example, previous research (Wassertheil-Smoller et al., 2014) found persons of Puerto Rican background to have higher levels of depressive symptoms compared to other Hispanic/Latino background groups (e.g., Central American, Cuban, Dominican, Mexican, and South American).

The Center for Epidemiologic Studies Depression Scale (CES-D) is designed to measure self-reported depression symptoms in the general population (Radloff, 1977). The 20 items for the CES-D (hereafter called CES-D 20) were generated from a pool of validated depression measures and were selected to map onto the then current Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 1968) criteria for depression (Radloff, 1977). Major components of depression symptoms assessed include depressed mood, feelings of guilt, hopelessness, loss of appetite and sleep disturbance (Radloff, 1977). The CES-D 20 has been widely used to measure depression symptoms in varied populations (e.g., adolescents, elderly, clinical and non-clinical) and contexts (Perreira, Deeb-Sossa, Harris, & Bollen, 2005) and has been translated into several languages including Spanish. The CES-D 20 has been shown to have high internal consistency, split-half reliability, and moderate test-retest reliability, across a variety of populations and language versions (Eaton, Muntaner, Smith, Tien, & Ybarra, 1999; Knight, Williams, McGee, & Olaman, 1997; Masten, Caldwell-Colbert, Alcala, & Mijares, 1986).

Several shortened versions of the CES-D (Andresen, Malmgren, Carter, & Patrick, 1994; Cole, Rabin, Smith, & Kaufman, 2004; Kohout, Berkman, Evans, & Cornoni-Huntley, 1993) have been developed for contexts in which the full instrument may be too burdensome. In an attempt to reduce administration time and improve clinical utility, Andresen et al. (1994) developed a 10-item version of the CES-D (hereafter called CES-D 10) to screen for depression symptoms in elderly adults. The CES-D 10 was developed from the original version by using item-total correlations and eliminating redundant items from the CES-D 20. Compared to the CES-D 20, which is comprised of both somatic and affective symptoms, the CES-D 10 is primarily comprised of affective symptoms (Cheng & Chan, 2005). Among adult samples, the CES-D 10 has shown comparable reliability (Irwin, Artin, & Oxman,

1999), good predictive accuracy when compared to the full-length CESD-20 (Andresen et al., 1994; Boey, 1999; Dimitrov, 2010; Irwin et al., 1999), and moderate test-retest stability over a 12-month period (Boey, 1999).

While the CES-D 20 has been shown to have good internal consistency reliability, test-retest reliability, convergent validity and measurement validity in general and diverse populations (Eaton et al., 1999; Knight et al., 1997; Masten et al., 1986), fewer studies have examined the psychometric properties of the CES-D 10 (Lee & Chokkanathan, 2008). Further, the majority of the research on the psychometric properties of the CES-D 10 has been conducted primarily with non-Hispanic White samples, despite the fact that the psychometric properties of an instrument are sample dependent. In particular, the applicability of the CES-D 10 across diverse Hispanics/Latinos has been largely unexplored. Indeed, the measurement properties of the Spanish translation of the CES-D 10 among Spanish-speaking Hispanic/Latinos require further examination.

Cross-cultural researchers have long recognized the importance of ensuring construct comparability in diverse linguistic and cultural groups such as Hispanics/Latinos. Given that 35.3 million Hispanics/Latinos speak a language other than English in the home, Spanish language measures of depression are increasingly necessary in research and clinical contexts. In addition, although translation procedures are followed, careful translation alone does not ensure that multiple language versions of an instrument measure the same construct, in the same way, in different groups (Nair, White, Knight, & Roosa, 2009). Therefore, empirical evaluations of measurement equivalence are necessary across linguistic and cultural groups if group or test score comparisons are planned. In addition, the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2009, 2014) and the International Test Commission (ITC) have developed guidelines for translating and psychometrically evaluating psychological tests (Hambleton, 2001).

It is fundamental to determine whether instruments developed for non-Hispanic Whites can be used effectively in ethnic minority populations (Ramirez, Ford, Stewart, & Teresi, 2005), including Hispanics/Latinos, the fastest growing ethnic minority group in the US (Passel, Cohn, & Lopez, 2011). Unless constructs are measured equivalently across linguistically diverse groups, findings from data pooled across languages and language-group comparisons may produce biased estimates and ultimately misleading conclusions (Chen, 2008; Ramirez et al., 2005). Furthermore, the lack of linguistically and culturally appropriate and/or validated measures limits research conducted in linguistic and ethnic minority groups (Martinez, 2008). Multiple factors, such as social desirability, translation problems, differential responses to positively versus negatively worded items, and response format, can affect the measurement equivalence of the construct and instrument (Hambleton, 2001). Therefore, prior to recommendations concerning the use of an instrument, the psychometric properties including measurement invariance of an instrument must be established. Measurement invariance examines the degree to which the psychometric properties of the observed indicators are generalizable across groups (Vandenberg & Lance, 2000). Measurement invariance is particularly important when linguistic and cultural groups are compared. Establishing measurement invariance allows for the interpretation of differences between groups to be considered accurate and meaningful (Ferro & Speechley, 2013).

The Current Study

Assessment of the psychometric properties of the CES-D 10 in English and Spanish is warranted to determine whether the measure accurately measures depressive symptomatology in Hispanics/Latinos. The present study examined (1) the internal consistency of the CES-D 10, (2) test-retest reliability of the CES-D 10 using a subset of participants who completed a second assessment and a third assessment, (3) convergent validity of the CES-D 10 with measures assessing related constructs (e.g., anxiety, health status), (4) the factor structure of the CES-D 10, and (5) factorial invariance (configural invariance, metric invariance, and scalar/threshold invariance) of the best-fitting model for English and Spanish language groups and for Hispanic/Latino background groups.

We hypothesized that the CES-D 10 would be internally consistent (.70) in the full sample, for language version (English and Spanish), and for diverse Hispanic/Latino background groups (i.e., Dominican, Central American, Cuban, Mexican, Puerto Rican, or South American). For test-retest reliability, we hypothesized that the CES-D 10 would demonstrate moderate to strong test-retest reliability between assessments conducted at baseline (Assessment Time 1), 3–9 months post baseline (Assessment Time 2) and within 1–3 weeks from Time 2 (Assessment Time 3). For convergent construct validity, we hypothesized that the CES-D 10 would be correlated with the Spielberger Trait Anxiety Inventory (STAI), Short Form Health Survey-12 (SF-12) Mental Health Component (MHC) score, SF-12 Physical Health Component (PHC) score and the Patient Health Questionnaire-9 (PHQ-9). Depression and anxiety are known to have high co-morbidity (Clark & Watson, 1991; Mineka, Watson, & Calark, 1998) and as such we expected a moderate correlation between scores from the CES-D 10 and the STAI. We expected the strongest correlation between the CES-D 10 and PHQ-9 as both assess depression symptoms. Based on previous measurement approaches (Björgvinsson, Kertz, Bigda-Peyton, McCoy, & Aderkan, 2013; Carpenter et al., 1998; Yu, Lin, & Hsu, 2013) and given its practical utility for screening purposes, the current study tested the one-factor structure of the CES-D 10 and we hypothesized that the model would demonstrate configural invariance, metric invariance, and scalar/threshold invariance and would provide a good fit to the data across both language groups and Hispanic/Latino background groups.

Methods

Participants and Procedures

Repeated data from three assessment points were used for the current study. Time 1 data were derived from the baseline exam of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a population-based cohort study designed to establish the prevalence, incidence, and risk and protective factors for major chronic diseases among Hispanic/ Latinos from diverse backgrounds. The HCHS/SOL parent study examined 16,415 self-identified Hispanics/Latinos aged 18 to 76 years from randomly selected households from communities surrounding four US field centers (Miami, FL; San Diego, CA; Bronx, NY; Chicago, IL), using a two-stage probability sampling approach. Details of the study sample (Lavange et al., 2010) and approach (Sorlie et al., 2010) have previously been described. Complete data were available for 15,487 individuals who participated in the HCHS/SOL

baseline assessment. For Assessment Time 2, a subsample of the larger study was recruited to participate in the HCHS/SOL Sociocultural Ancillary Study, a separate comprehensive assessment of socioeconomic, cultural, and psychosocial factors. This subsample was comprised of approximately one third of the HCHS/SOL cohort ($n = 5{,}313$), with assessment within 3–9 months of the baseline exam; the methods for this study have previously been described (Gallo et al., 2014). Complete data were available for 4,959 individuals who participated in the HCHS/SOL Sociocultural Ancillary Study. For Assessment Time 3, a subset of participants ($n = 325$) from three (Chicago, IL; Miami, FL; San Diego, CA) of the four field centers, who completed both the HCHS/SOL and HCHS/SOL Sociocultural Ancillary Study exam were recruited to complete a third survey administration within 1–3 weeks of the Sociocultural Ancillary Study with the purpose of providing preliminary evidence of test-retest reliability and convergent validity. Complete data were available for $n = 309$ participants (see Table 1). At each participating field center, Institutional Review Board approval was obtained and all participants provided written informed consent.

Measures

Demographic Variables—Participants completed questions pertaining to socio-demographic characteristics including age, participant sex, marital status, income ($0 = <$ \$20,000, $1 = $ \$20,001–50,000 and $2 = > $ \$50,001), education ($1 = $ less than high school, $2 = $ high school graduate, $3 = $ above high school), number of years living in the US, Hispanic/ Latino background (Central American, Cuban, Dominican, Mexican, Puerto Rican, and South American) and language (i.e., English or Spanish) that each participant selected to complete the interview.

Center for Epidemiologic Studies Depression Scale—(CES-D 10; Andresen et al., 1994). The CES-D 10 measures frequency of depression symptoms experienced in the past week. Ratings were based on a 4-point response format from 0 (*rarely or none of the time*) to 3 (*most or all of the time*) with positively worded items (items 5 and 8) reverse scored, and total scores ranging from 0 to 30. The CES-D 10 has demonstrated good internal consistency reliability in the general population, in older adults and in multiethnic populations (Cheng & Chan, 2005; Irwin et al., 1999). The CES-D 10 has also demonstrated acceptable to good sensitivity and specificity in detecting a depression diagnosis (Björgvinsson, Kertz, Bigda-Peyton, McCoy, & Aderka, 2013; Zhang et al., 2012). The Spanish version of the CES-D 10 was translated by the HCHS/SOL following recommended translation guidelines (Van de Vivier & Hambleton, 1996). Briefly, translation followed a four step process that included: (1) Creation of two independent translations; (2) Comparison and review of translations by a committee comprised of bilingual/bicultural members from each of the primary Hispanic/Latino background groups; (3) Pilot-testing of the approved version via focus groups comprised of bilingual and monolingual representatives; and (4) Validation of translated instruments with a group of bilingual representatives. Internal consistency for the current sample was acceptable ($\alpha_{full sample} = .80$; α English $= .82$; α Spanish $= .82$).

Spielberger Trait Anxiety Inventory—(STAI; Spielberger, Gorsuch, & Lushene, 1970). Trait anxiety was measured using the 10-item version of the 20-item Trait Anxiety Scale from the STAI. The short form version correlates highly with the full version ($r = .96$, unpublished work). Respondents rate how they generally feel (e.g., nervous and restless) on a 4-point scale from 1 (*almost never*) to 4 (*almost always*). Total scores range from 10 to 40 with higher scores reflecting greater endorsement of anxious feelings. This abbreviated measure has demonstrated good psychometric properties including good test-retest reliability (Bromberger & Matthews, 1996; Matthews, Kelsey, Meilahn, Kuller, & Wing, 1989). The Spanish translation of the STAI (Salman, 1998) is available from Mind Garden (http://www.mingarden.com/products/staisad.htm). Internal consistency was high for the current sample (Cronbach's α full sample = .93; α English = .92; α Spanish = .94).

Short-Form 12 Health Survey—(SF-12; Ware, Kosinski, & Keller, 1995). The SF-12 is a general health-related quality of life instrument. It was originally developed as an alternative to the widely used Short-Form 36 Health Survey (SF-36), for use with studies in which the SF-36 may have been too lengthy. The measure yields two scores: the Physical Component Summary (PCS) and the Mental Health Component Summary (MCS) score. Three of the five items for the MCS reflect symptoms from the diagnostic criteria for depression and anxiety, such as feeling depressed and feeling restless (Vilagut et al., 2013). Scores are standardized to population norms, with the mean set at 50 ($SD = 10$) and where a zero score indicates the lowest level of health. The SF-12 has demonstrated good test-retest reliability (Amir, Lewin-Epstein, Becker, & Buskila, 2002; Gandek et al., 1998). The Spanish SF-12 is available for public use from Quality Metric at http://www.qualitymetric.com.

Patient Health Questionnaire-9—(PHQ-9; Spitzer, Kroenke, & Williams, 1999). The PHQ-9 is a self-report assessment of recent (past 2 weeks) depression symptoms on a 4-point scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 27. It contains nine items that parallel the diagnostic criteria for depression outlined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition-Text Revision (DSM-IV-TR; American Psychiatric Association, 2000). The PHQ-9 has demonstrated good internal consistency reliability (Merz, Malcarne, Roesch, Riley, & Sadler, 2011; Spitzer, Kroenke, & Williams, 1999) and excellent test-retest reliability (Kroenke, Spitzer, & Williams, 2001; Patten & Schopflocher, 2009; Pinto-Meza, Serrano-Blanco, Penarrubia, Blanco, & Haro, 2005). The Spanish PHQ-9 is provided by the authors of the measure and is available for public use. Internal consistency was good in the current sample (α full sample = .85; α English = .82; α Spanish = .86).

Procedure and Statistical Analyses

Analyses were performed using IBM Statistical Package for the Social Sciences (SPSS) Version 20 (IBM Corp. Armonk, NY), Statistical Analysis System Version 9.4 (SAS. Cary, North Carolina) and M*plus* Version 7.0 (Muthén & Muthén, 2006). Alpha was set at 0.05. Descriptive statistics for the sample were conducted in SAS using complex sampling procedures to account for the stratified multi-stage area probability study design of household addresses and were weighted relative to the 2010 census to adjust for sampling

probability and nonresponse (Lavange et al., 2010). Confirmatory factor analyses (CFA) and multiple group CFAs across language and ethnic background group were conducted in M*plus* 7.0.

Internal Consistency, Test-retest Reliability and Convergent Validity

Cronbach's alpha (α) was calculated for parent study data to evaluate the internal consistency reliability of the CES D-10 across the full sample, language (English and Spanish), and Hispanic/Latino background groups (Dominican, Central American, Cuban, Mexican, Puerto Rican, or South American), with reliability coefficients exceeding .70 considered adequate. Test-retest reliability was estimated by correlating CES-D 10 total scores between Assessment Time 1 with Assessment Time 2 and Assessment Time 3. Convergent validity was calculated using Pearson's correlations between the CES-D 10 with STAI, SF-12 MHC, SF-12 PHC and PHQ-9.

Confirmatory Factor Analysis (CFA) and Multiple-Group CFA

CFA was used to examine the validity of the one-factor structure of the CES-D 10 based on data from Assessment Time 1. CFA and multiple group CFA models were tested for data fit using weighted least squares mean- and variance adjusted (WLSMV) estimation with THETA parameterization, which is appropriate for modeling categorical or ordinal data (Beauducel & Herzberg, 2006). Evaluation of the overall model fit was based on the following goodness-of-fit indices: (a) the model Chi-square (χ^2; Hu & Bentler, 1999), (b) the Comparative Fit Index (CFI; Bentler, 1990); and (c) the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990). Evidence of statistical model fit is provided when the χ^2 value is not statistically significant; however, the χ^2 index tends to falsely reject adequate model fit with large sample sizes, such as ours (Hoyle, 2000). We followed accepted guidelines described by Hu and Bentler (1999) and determined that models with a CFI greater than or equal to .95, and an RMSEA less than or equal to .06, fit reasonably well.

We then evaluated the one-factor structure of the CES-D 10 for group invariance across language and Hispanic/Latino background, respectively. Following recommendations by Vanderberg and Lance (2000; see Dimitrov, 2010), we used a sequential model comparison approach of nested models (between increasingly restrictive models) that addressed configural invariance (i.e., equivalent factor structure), metric invariance (i.e. equivalent factor loadings), and scalar/threshold invariance (i.e., equivalent item thresholds in the case of ordinal data) across groups. Because the χ^2 is heavily influenced by sample size, we determined the factorial invariance between nested models by examining changes in descriptive indices of model fit as main criteria. Specifically, decreases in CFI values of less than or equal to .01 and increases in RMSEA values of less than or equal to .015 indicated invariance at each step (Chen, 2007; Cheung & Rensvold, 2002; Dimitrov, 2010). When evaluating language invariance between English and Spanish language groups, configural invariance was tested by comparing whether the factor structure was equivalent across languages with no equality constraints imposed on factor loadings or thresholds. Equality constraints across groups were then imposed sequentially; first on factor loadings and subsequently on both factor loadings and item thresholds to test for metric and scalar/

threshold invariance, respectively. If the change in either of the fit indices between a less restrictive and a more restrictive model was greater than the specified cut-off criteria, we tested for partial metric or scalar/threshold invariance by releasing the equality constraints of parameters (i.e., loadings or thresholds) for items with the largest modification indices one at a time until the change criteria were sufficiently met. Furthermore, we conducted multiple-group analyses to test for invariance of the one-factor structure CES-D 10 across the six Hispanic/Latino background groups following the same procedure used to establish measurement invariance across languages.

Results

The age of the target population ranged from 18 to 76, with an average of 41.3 years (SE = .25). As shown in Table 2, nearly half (46.5%) of the target population had annual household incomes of less than $20,000, two-thirds (67.1%) had a high school education or greater, and half were married or cohabitating (49.7%). Within the sample, the majority of participants chose to complete the interview in Spanish (76.2%).

Internal Consistency, Test-retest Reliability and Convergent Validity

The means and standard deviations for the CES-D 10 items are shown in Table 4. As shown in Table 4, internal consistency reliabilities for the full sample, language and background groups were acceptable, with Cronbach's alphas .70. With regard to test-retest reliability, among participants who completed the CES-D 10 a second time within the 3–9 month interval (Assessment Time 2) the temporal stability coefficient was moderate, r = 53, p < .001 and among participants who completed the CES-D 10 within the 1–3 week interval (Assessment Time 3) the temporal stability coefficient was strong, r = .70, p < .001. Test-retest reliability of the CES-D 10 total score across language group and background group was also acceptable, with Pearson's correlations ranging from r = .42 to r = .78. However, within the third assessment time (n = 325), small sample sizes for Dominican As shown in Table 5, the CES-D 10 positively correlated with the STAI, r = .72, p < .001, negatively with the SF-12 MHC, r = −.65, p < .001 and the SF-12 PHC, r = −.25, p < .001. The CES-D 10 correlated positively and strongly with the PHQ-9, r = .80, p < .01.

Confirmatory Factor Analysis (CFA)

Table 6 shows the model fit indices for the one-factor CES-D 10 model for the total sample. Based on the results of the CFI and RMSEA fit indices, this model did not yield an acceptable fit to the data (χ^2 = 2637.35, df = 35, p < .001; CFI = .968; RMSEA = .069), as indicated by the RMSEA. The modification indices suggested correlating the residual variances of the two reverse-worded items of the CES-D 10 (i.e., item 5 "*I felt hopeful about the future*" and item 8 "*I was happy*") to improve model fit. After making this modification, the one-factor model yielded acceptable fit to the data (χ^2 = 1197.46, df = 34, p < .001; CFI = .986; RMSEA = .047). All unstandardized factor loadings were statistically significant (values ranged from .22 to 2.01, ps < .001; standard errors ranged from .01 to .05 [data not shown]).

Multiple-Group CFA across Language

We conducted multiple-group analyses to test for measurement invariance of the one-factor structure of the CES-D 10 across English- and Spanish-language groups, retaining the better-fitting model in which the residual variances of the two reverse-worded items were correlated. Fit statistics for the configural, metric and scalar variance models, and their sequential model comparison between nested models are shown in Table 6.

Configural invariance—A multiple-group CFA was evaluated, with the overall factor structure constrained equal and factor loadings and item thresholds estimated freely across language groups. The model yielded acceptable fit (CFI = .985; RMSEA = .049), suggesting that the one-factor structure of the CES-D 10 is equivalent across language groups. As shown in Table 7, all unstandardized factor loadings were statistically significant in the English (.18 to 1.79, ps < .001) and Spanish-language group (.17 to 1.59, ps < .001).

Metric invariance—We tested metric invariance by constraining all factor loadings equal and estimating thresholds freely across language groups. As seen in Table 6, the metric invariance model fit well according to the descriptive fit indices (CFI = .986; RMSEA = .045). The sequential model comparison between nested models (metric vs. configural model) showed that the change in CFI was less than or equal to .01 (∆CFI = .001, and the change in RMSEA was less than or equal to .015 (∆RMSEA = −.004). Therefore, results indicate that the one-factor structure of the CES-D 10 has measurement invariance; factor loadings are equivalent across English- and Spanish-language groups.

Scalar invariance—A subsequent model constrained both factor loadings and item thresholds equal across language groups. As seen in Table 6, the scalar/threshold invariance model exhibited acceptable fit (CFI = .976; RMSEA = .049). In the sequential model comparison of nested models (scalar model vs. metric model), no salient differences in descriptive fit were noted (∆CFI = −.01; ∆RMSEA = .004). These results indicate that item thresholds were invariant across English- and Spanish-language groups.

Multiple-group CFA across the Six Hispanic/Latino Background Groups

Configural invariance—Following the same procedure used to establish configural invariance between language groups, we examined the fit of the one-factor solution structure of the CES-D 10 across six Hispanic/Latinos background groups. Factor loadings and item threshold were estimated freely across groups. The configural invariance model exhibited adequate descriptive fit (CFI = .985; RMSEA = .049; see Table 7). As shown in Table 8, all unstandardized factor loadings were statistically significant across the six Hispanic/Latino background groups (ps < .001).

Metric invariance—Subsequently, a metric invariance model was tested by constraining factor loadings equal and estimating item thresholds freely across Hispanic/Latino background groups. As seen in Table 7, the metric invariance model fit well descriptively (CFI = .987; RMSEA = .041). Moreover, the change in fit indices (∆CFI = .002; ∆RMSEA = −.008) between increasingly restrictive nested models (metric vs. configural model) suggested that the factor loadings were invariant across background groups.

Scalar invariance—The descriptive fit indices showed that the scalar/threshold invariance model, in which both factor loadings and item thresholds were constrained equal across Hispanic/Latinos background groups, fit adequately (CFI = .969; RMSEA = .050; [see Table 7]). When comparing nested models (scalar vs. metric model), the change in RMSEA was acceptable (RMSEA = .009), but the change in CFI was above the recommended cut-off (CFI = −.018), suggesting that there was not full scalar/threshold invariance (i.e., equal thresholds for all items) across Hispanic/Latino background groups (see Table 7). Consequently, we tested for partial scalar/threshold invariance by releasing the equality constraints of item thresholds with the largest modification indices one at a time until CFI provided evidence of invariance. As seen in Table 7, we released equality constraints (i.e., freed item thresholds) one at a time until a negative CFI was lower than −.01. We had to release equality constraints for the following six item thresholds: Item 5 (thresholds 1 and 2), item 8 (thresholds 1 and 2), item 2 (threshold 1), and item 7 (threshold 3) in the Mexican and Cuban background group. No further modifications of item thresholds were needed to establish acceptable CFI caused by scalar/threshold invariance constraints across Hispanic/ Latino background groups. Therefore, results indicate partial scalar/threshold invariance across Hispanic/Latino background groups. Item thresholds were equivalent across Hispanic/Latino background groups (i.e., Dominicans, Central and South Americans, and Puerto Ricans) with the exception of the aforementioned items for Mexican and Cuban background groups.

Discussion

The present study is the first to comprehensively examine the psychometric properties of the CES-D 10 in a large and diverse sample of Hispanics/Latinos. Specifically, this study evaluated the internal consistency, test-retest reliability, convergent validity, factor structure and measurement invariance of the CES-D 10 in a heterogeneous sample of English- and Spanish-speaking Hispanics/Latinos.

Consistent with previous studies (Björgvinsson, Kertz, Bigda-Peyton, McCoy, & Aderka, 2013; Boey, 1999; Carpenter et al., 1998; Radloff, 1977; Yu et al., 2013), the CES-D 10 scores had acceptable internal consistency reliability in the current cohort. In alignment with previous research (Boey, 1999; Radloff, 1977) among non-Hispanic White samples, test-retest reliability was also good and was strong within a 1–3 week timeframe, and moderate across a longer timeframe of 3–9 months. Convergent construct validity was established by the observed theoretically consistent patterns and magnitudes of correlations with other relevant self-report measures.

CFA results from the current study support the unidimensionality of the CES-D 10. These findings align with results from previous studies (Björgvinsson, Kertz, Bigda-Peyton, McCoy, & Aderka, 2013; Carpenter et al., 1998; Yu et al., 2013), which reported an adequate fit for a one-factor model (i.e., all items loading on one factor). In addition, our findings demonstrate strong measurement invariance across language groups and acceptable partial invariance across Hispanic/Latino background groups. In other words, the CES-D 10 measured the same underlying construct for both English- and Spanish-speaking participants and Hispanic/Latino background groups, these results have practical implications for

researchers and clinicians. For example, since the CES-D 10 demonstrated no substantive differences in reliability or factor structure across language or background groups, future research that identifies differences in total depression scores across such groups can be more confident that they are attributable to true differences in depression symptoms, as opposed to differences stemming from measurement error artifacts.

Invariance of factor loadings across groups is required for valid comparisons of scale scores. However, strong and strict invariance may be less important in the context of basic research in which variation among groups may reflect differences that are relevant to the scientific investigation (Meredith & Teresi, 2006). The failure of any level of factorial invariance may arise because of the measure, population, or both. In other words, both the measure and the differences between the groups can contribute to the lack of measurement invariance. A perfect invariant measure is an elusive goal (Cheung & Reinsvold, 1999) due in part to the different interpretation of scale items across cultural groups due to diverse backgrounds. Consequently, it is not alarming that full measurement invariance was not established across background groups as strict factorial invariance is often difficult to establish particularly in cross-cultural research. In the present study, the same CES-D items (i.e., items 2, 5, 7, and 8) may have been viewed differently among the background groups (Mexican and Cuban groups, respectively). We recommend that future research examine measurement invariance at the item level using differential item functioning.

The present study provides evidence to support the use of the CES-D 10 in linguistically and ethnically diverse Hispanics/Latinos. Findings highlight that the CES-D 10 has strong psychometric properties and thus may maintain advantages over the original scale validated among non-Hispanic White respondents (Van Dam & Earleywine, 2011), which suggests that the CES-D10 can be used in lieu of the CES-D 20. Findings have important implications as the CES-D 10 may be more practical for clinical settings, and also in large studies that mandate a brief assessment of depression symptoms (Irwin et al., 1999; Lee & Chokkanathan, 2008).

There are some limitations to the current study. First, the current study may not be nationally representative of the general U.S. Hispanic/Latino population, given that participants were recruited from Miami, San Diego, the Bronx, New York City and Chicago. Second, we did not examine the relationship between the CES-D 10 with a gold standard assessment of depression (e.g., structured psychiatric interview). As such, the current study does not provide information regarding the utility of the proposed cut scores for the CES-D10. Future research would benefit from examining other forms of validity. Another study limitation is that invariance was not established across other demographic groups such as age, sex, and nativity. Despite these limitations, the present study extends the literature on the CES-D 10 in a heterogeneous, population-based sample of Hispanics/Latinos and several theoretical and clinical implications can be drawn from the findings.

Conclusions

In light of the growing diversity of Hispanic/Latinos in the U.S. population and the increasingly recognized importance of screening for depression in health settings, there is

need for an effective depression screening measure that can be used in English- and Spanish-speaking Hispanic/Latinos of diverse backgrounds. The current findings indicate that the CES-D 10 has strong psychometric properties, making it a viable measurement tool for assessing depression symptoms among Hispanics/Latinos in research or clinical settings.

Acknowledgments

References

American Educational Research Association, American Psychological Association. Standards for educational and psychological testing. Washington, DC: AERA; 1999, 2014. National Council on Measurement in Education, Joint Committee on Standards for Educational & Psychological Testing (US).

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 2nd. Washington, DC: American Psychiatric Association; 1968.

Amir M, Lewin-Epstein N, Becker G, Buskila D. Psychometric Properties of the SF-12 (Hebrew Version) in a Primary Care Population in Israel. Medical Care. 2002; 40(10):918–928. [PubMed: 12395025]

Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: Evaluation of a short form of the CES-D. American Journal of Preventive Medicine. 1994; 10:77–84. [PubMed: 8037935]

Beauducel A, Herzberg PY. On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Structural Equation Modeling. 2006; 13:186–203.

Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]

Björgvinsson T, Kertz SJ, Bigda-Peyton JS, McCoy KL, Aderka IA. Psychometric Properties of the CES-D-10 in a Psychiatric Sample. Assessment. 2013; 20(4):429–436. [PubMed: 23513010]

Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. International Journal of Geriatric Psychiatry. 1999; 14:608–617. [PubMed: 10489651]

Bromberger JT, Matthews KA. A "feminine" model of vulnerability to depressive symptoms: a longitudinal investigation of middle-aged women. Journal of Personality & Social Psychology. 1996; 70(3):591–598. [PubMed: 8851743]

Carpenter JS, Andrykowski MA, Wilson J, Hall L, Rayens MK, Sachs B, Cunningham LL. Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. Issues Mental Health Nursing. 1998; 19:481–494.

MMWR. Centers for Disease Control and Prevention (CDC). Current Depression Among Adults - United States, 2006 and 2008. 2010; 59:1229–1235.

Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling. 2007; 14:464–504.

Chen FF. What Happens if we Compare Chopsticks with Forks? The Impact of Making Inappropriate Comparisons in Cross-Cultural Research. Journal off Personality and Social Psychology. 2008; 95(5):1005–1018.

Cheng S-T, Chan ACM. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. International Journal of Geriatric Psychiatry. 2005; 20:465–470. [PubMed: 15852439]

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

Clark LA, Watson D. Tripartite Model of Anxiety and Depression: Psychometric Evidence and Taxonomic Implications. Journal of Abnormal Psychology. 1991; 100(3):316–336. [PubMed: 1918611]

Cole JC, Rabin AS, Smith TL, Kaufman AS. Develpment and validation of a Rasch-derived CES-D short form. Psychological Assessment. 2004; 16:360–372. [PubMed: 15584795]

Dimitrov DM. Testing for factorial invariance in the context of construct validation. Measurement and Evaluation in Counseling and Development. 2010; 43:121–149.

Eaton, WW., Muntaner, C., Smith, C., Tien, A., Ybarra, M. Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESDR). In: Masten, J., editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. Mahwah, NJ: Lawrence Elrbaum; 1999.

Ferro MA, Speechley KN. Factor structure and longitudinal invariance of the Center for Epidemiological Studies Depression Scale (CES-D) in adult women: application in a population-based sample of mothers of children with epilepsy. Arch Womens Ment Health. 2013; 16:159–166. [PubMed: 23420273]

Gallo LC, Penedo FJ, Carnethon M, Isasi C, Sotres-Alvarez S, Malcarne VL, Talavera GT. The Hispanic Community Health Study/Study of Latinos Sociocultural Ancillary Study: Sample, Design, and Procedures. Ethnicity and Disease. 2014 Winter;24:77–83. [PubMed: 24620452]

Gandek B, Ware JE, Aaronson NK, Apolene G, Bjorner JB, Brazier JE, Sullivan M. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine counties: Results from the IQOLA Project. Journal of Clinical Epidemiology. 1998; 11:1171–1178.

Hambleton RK. The next generation of the ITC Test Translation and Adaptation Guidelines. European Journal of Psychological Assessment. 2001; 17(3):164–172.

Hoyle, RH. Confirmatory factor analysis. In: Tinsley, HEA., Brown, SD., editors. Handbook of applied multivariate statistics and mathematical modeling. San Diego, CA: Academic Press; 2000. p. 465-497.

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

Irwin M, Artin KH, Oxman MN. Screening for Depression in the Older Adult: Criterion Validity of the 10-Item Center for Epidemiological Studies Depression Scale (CES-D). Archives of Internal Medicine. 1999; 159:1701–1704. [PubMed: 10448771]

Knight RG, Williams S, McGee R, Olaman S. Psychometric properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. Behavioral Research Therapy. 1997; 35(4):373–380.

Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D depression symptoms index. Journal of Aging and Health. 1993; 5:179–193. [PubMed: 10125443]

Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001; 16:606–613. [PubMed: 11556941]

Lavange LM, Kalsbeek WD, Sorlie PD, Aviles-Santa LM, Kaplan RC, Barnhart J, Elder JP. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Annals of Epidemiology. 2010; 20:642–649. [PubMed: 20609344]

Lee AE, Chokkanathan S. Factor structure of the 10-item CES-D scale among community dwelling older adults in Singapore. International Journal of Geriatric Psychiatry. 2008; 23:592–597. [PubMed: 18023070]

Martinez SM, Ainsworth BE, Elder JP. A review of physical activity measures used among US Latinos: Guidelines for developing culturally appropriate measures. Annals of Behavioral Medicine. 2008; 36:195–207. [PubMed: 18855091]

Masten WG, Caldwell-Colbert AT, Alcala SJ, Mijares BE. Reliability and validity of the Center for Epidemiological Studies-Depression Scale. Hispanic Journal of Behavioral Sciences. 1986; 8:77–84.

Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR. Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. American Journal of Epidemiology. 1989; 129(6):1132–1144. [PubMed: 2729252]

Meredith W, Teresi JA. An essay on measurement and factorial invariance. Med Care. 2006; 44(11 Suppl 3):S69–S77. [PubMed: 17060838]

Merz E, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. Cultural Diversity and Ethnic Minority Psychology. 2011; 17(3):309–316. [PubMed: 21787063]

Mineka S, Watson D, Calark LA. Comorbidity of anxiety and unipolar mood disorders. Annual Review of Psychology. 1998; 49:377–412.

Muthén, L., Muthén, B. Mplus User's Guide. 4th. Los Angeles, CA: 2006.

Nair RL, White RMB, Knight RG, Roosa MW. Cross-Language Measurement Equivalence of Parenting Measures for use with Mexican American Populations. Journal of Family Psychology. 2009; 23(5):680–689. [PubMed: 19803604]

Passel, JS., Cohn, DV., Lopez, MH. Census 2010: 50 Million Latinos (Pew Hispanic Center report). Washington, D.C: Pew Hispanic Center; 2011.

Patten SB, Schopflocher D. Longitudinal epidemiology of major depression as assessed by the Brief Patient Health Questionnaire (PHQ-9). Comprehensive Psychiatry. 2009; 50:26–33. [PubMed: 19059510]

Perreira KM, Deeb-Sossa N, Harris KM, Bollen K. What are we measuring? An evalution of the CES-D across race/ethnicity and immigrant generation. Social Forces. 2005; 83(4):1567–1602.

Pinto-Meza A, Serrano-Blanco A, Penarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: Can it be carried out over the telephone? Journal of General and Internal Medicine. 2005; 20:738–742.

Radloff LS. The CES-D scale: A scale. 1977

Ramirez M, Ford M, Stewart AL, Teresi JA. Measurement issues in health disparities research. Health Services Research. 2005; 40(5):1640–1657. [PubMed: 16179000]

Salman, E. Research Edition Spanish Translation: State-Trait Anxiety Inventory (Form Y). Redwood City, CA: Mind Garden; 1998.

Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello A, Heiss G. Design and Implementation of the Hispanic Community Health Study/Study of Latinos. Annals of Epidemiology. 2010; 20:629–641. [PubMed: 20609343]

Spielberger, CD., Gorsuch, RL., Lushene, RE. The State-Trait Anxiety Inventory: Test Manual. Palo alto, CA: Consulting Psychological Press; 1970.

Spitzer R, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999; 282:1737–1744. [PubMed: 10568646]

Steiger JS. Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research. 1990; 25:173–180. [PubMed: 26794479]

Van Dam NT, Earleywine M. Validation of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R): Pragmatic depression assessment in the general population. Psychiatry Research. 2011; 186:128–132. [PubMed: 20843557]

Van de Vijver FJR, Hambleton RK. Translating Tests: Some practical guidelines. European Psychologist. 1996; 1(2):89–99.

Vilagut G, Forero CG, Pinto-Meza A, Haro JM, de Graaf R, Bruffaerts R, Alonso J. The Mental Component of the Short-Form 12 Health Survey (SF-12) as a Measure of Depressive Disorders in the General Population: Results with Three Alternative Scoring Methods. Value in Health. 2013; 16:564–573. [PubMed: 23796290]

Ware JE, Kosinski M, Keller SD. A 12-item short form health survey. Construction of scales and preliminary tests of reliability and validity. Medical Care. 1995; 34:220–233.

Wassertheil-Smoller S, Arredondo EM, Cai J, Castaneda S, Choca JP, Gallo LC, Zee PC. Depression, anxiety, antidepressant use, and cardiovascular disease among Hispanic men and women of different national backgrounds: results from the hispanic Community Health Study/Study of Latinos (HCHS/SOL). Annals of Epidemiology. 2014; 24(11):822–830. [PubMed: 25439033]

Yu S, Lin Y-H, Hsu W-H. Applying structural equation modeling to report psychometric properties of Chinese version 10-item CES-D depression scale. Qual Quant. 2013; 47:1511–1518.

Zhang W, O'Brien N, Forrest JI, Salters KA, Patterson TL, Montaner JSG, Lima VD. Validating a shortened depression scale (10 item CES-D) among HIV-positive people in British Columbia, Canada. PLOS ONE. 2012; 7(7):1–5.

Table 1

Measures Administered across Assessment Times

Instrument	Assessment Time 1 (HCHS/SOL Baseline)	Assessment Time 2 Sociocultural Ancillary Study	Assessment Time 3
Demographics	X		
CES-D 10	X	X	X
STAI	X	X	X
SF-12-MHC	X		X
SF-12 PHC	X		X
PHQ-9			X

Note. CES-D 10 = Center for Epidemiological Studies of Depression (10 items); STAI = Spielberger Trait Anxiety Inventory; SF-12 MHC = Short-Form 12 Health Survey Mental Health Component; SF-12 PHC = Short-Form Health Survey Physical Health Component; PHQ-9 = Patient Health Questionnaire-9. An "**X**" indicates that an assessment was completed at the specified time.

Table 2

Weighted Sample Characteristics for the Total Sample and by Hispanic/Latino Background Group (N =15,487)

	Overall (N = 15,487)	Dominican (n = 1,400)	Central American (n = 1,704)	Cuban (n = 2,302)	Mexican American (n = 6,412)	Puerto Rican (n = 2,629)	South American (n = 1,040)
Age[2]	41.32 (0.25)	39.55 (0.70)	39.67 (0.48)	46.46 (0.53)	38.52 (0.38)	42.99 (0.52)	42.35 (0.77)
Women[1]	52.07 (0.54)	60.44 (1.93)	52.46 (1.67)	47.55 (1.07)	53.30 (0.97)	48.81 (1.41)	54.66 (2.00)
Married or Cohabitating[1]	49.72 (0.78)	37.59 (1.84)	46.49 (1.61)	51.64 (1.51)	59.61 (1.22)	32.58 (1.46)	49.58 (2.16)
Income[1]							
< $20,000	46.50 (1.02)	54.53 (2.34)	53.16 (2.16)	54.24 (1.64)	38.92 (1.63)	49.38 (1.90)	43.40 (2.12)
$20,001–$50,000	40.75 (0.70)	37.20 (2.13)	38.87 (1.97)	36.69 (1.51)	45.25 (1.14)	36.08 (1.87)	44.73 (2.01)
> $50,001	12.75 (0.82)	8.27 (1.28)	7.97 (1.14)	9.06 (1.12)	15.84 (1.52)	14.54 (1.21)	11.97 (1.48)
Education[1]							
< High School/GED	32.89 (0.74)	36.79 (1.82)	37.99 (1.65)	22.25 (1.07)	36.42 (1.33)	36.45 (1.61)	22.14 (1.97)
High School	28.58 (0.58)	23.13 (1.89)	26.33 (1.49)	30.09 (1.41)	30.04 (1.02)	27.84 (1.26)	27.84 (1.90)
> High School/GED	38.53 (0.83)	40.08 (1.80)	35.68 (1.64)	47.66 (1.53)	33.54 (1.57)	35.71 (1.64)	50.02 (2.17)
Years in US[1]							
< 10	28.32 (0.97)	24.48 (1.91)	37.30 (2.10)	49.13 (1.83)	23.95 (1.19)	6.44 (0.88)	41.20 (2.40)
10	71.68 (0.97)	75.52 (1.91)	62.70 (2.10)	50.87 (1.83)	76.05 (1.18)	93.56 (0.88)	58.80 (2.40)
US Born[2]	21.36 (0.79)	16.31 (2.04)	7.32 (1.14)	7.28 (0.94)	23.77 (1.04)	48.14 (1.57)	5.45 (0.95)
Spanish language[1]	76.21 (0.90)	76.59 (2.43)	88.27 (1.48)	92.65 (0.86)	77.90 (1.09)	41.23 (1.81)	89.50 (1.42)
CES-D 10[2]	6.99 (0.08)	7.16 (0.22)	6.59 (0.17)	6.95 (0.17)	6.34 (0.13)	8.83 (0.22)	6.49 (0.23)
STAI[2]	17.02 (0.35)	17.23 (0.26)	16.54 (0.16)	16.31 (0.16)	16.97 (0.13)	18.36 (0.22)	16.25 (0.21)
SF-12 MHS[2]	49.18 (0.15)	48.51 (0.52)	50.35 (0.33)	49.80 (0.32)	49.49 (0.27)	47.20 (0.42)	50.14 (0.46)

Note.

[1] n (%),

[2] M (SE); CES-D 10 = Center for the Epidemiological Studies of Depression (10 items); STAI = Spielberger Trait Anxiety Inventory; SF-12 MHS = Short-Form Health Survey Mental Health Component.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

CES-D 10 Item-Level Descriptive Statistics (N = 15,487)

CES-D 10		*M*	*SD*
Item 1	I was bothered by things that usually don't bother me.	0.55	0.84
Item 2	I had trouble keeping my mind on what I was doing.	0.69	0.94
Item 3	I felt depressed.	0.79	0.99
Item 4	I felt that everything I did was an effort.	0.75	1.03
Item 5▪	I felt hopeful about the future.	0.86	1.10
Item 6	I felt fearful.	0.48	0.84
Item 7	My sleep was restless.	0.92	1.08
Item 8▪	I was happy.	0.83	1.00
Item 9	I felt lonely.	0.72	1.03
Item 10	I could not "get going".	0.70	0.94

▪*Note*. Items 5 and 8 are reversed coded items.

Table 4

Descriptive Statistics, Internal Consistency and Test-Retest for the CES-D 10 Total Score for the Total Sample, Language Groups, and Hispanic/Latino Background Groups

	Range	M	SE	α	Test-Retest (1–3 weeks) Pearson's r (N = 325)	Test-Retest (3–9 months) Pearson's r (N = 5,313)
Overall	0–30	7.30	.05	.82	.70**	.53**
Language						
English	0–30	7.88	.11	.82	.69**	.57**
Spanish	0–30	7.17	.05	.82	.71**	.52**
Background Group						
Dominican	0–30	7.33	.16	.82	NA	.48**
Central American	0–30	7.09	.15	.82	.59**	.41**
Cuban	0–30	7.46	.14	.86	.62**	.52**
Mexican	0–30	6.65	.07	.80	.44**	.54**
Puerto Rican	0–30	9.12	.13	.82	.59**	.55**
South American	0–30	6.69	.18	.82	NA	.54**

Note. NA = Assessment Time 3 had small sample size composition among the Dominican (n = 3) and South American (n = 21) background groups and therefore results are not presented for these groups.

Table 5

Pearson's Correlations between CES-D 10 Scores at Assessment Time 1 and Time 3 and Validity Measures across Language Groups and Background Groups

	Overall	English	Spanish	Dominican	Central American	Cuban	Mexican	Puerto Rican	South American
STAI Time 1	.72**	.77**	.71**	.71**	.70**	.75**	.71**	.77**	.71**
SF-12 MHC Time 1	−.65**	−.68**	−.64**	−.63**	−.65**	−.71**	−.60**	−.68**	−.64**
SF-12 PHC Time 1	−.25**	−.27**	−.24**	−.24**	−.24**	−.24**	−.20**	−.29**	−.15**
PHQ-9 Time 3 (N =325)	.80**	.81**	.80**	NA	.62**	.83**	.77**	.86**	.87**

Note. CES-D 10 = Center for the Epidemiological Studies of Depression. STAI = Spielberger Trait Anxiety Inventory; SF-12 = Short-Form 12. MHC = Mental Health Component score. PHC = Physical Health Component score.

**
$p < .01$.

Table 6

Goodness of Fit Statistics for the One-Factor Model of the CES-D 10 and Tests of Factorial Invariance across Language

Model	S-Bχ^2	Df	P	CFI	RMSEA	Model Comparison	CFI	RMSEA
M1. One-factor baseline	2637.35	35	<.001	.968	.069	--	--	--
M1a. One-factor baseline (modified)	1197.455	34	<.001	.986	.047	--	--	--
M2. Configural	1348.891	68	<.001	.985	.049	--	--	--
M3. Metric	1260.756	77	<.001	.986	.045	M3-M2	.001	–.004
M4. Scalar	2077.007	106	<.001	.976	.049	M4-M3	–.001	.004

Note. M1 = One-factor baseline model (free model). M1a = Modified baseline model (one-factor model with correlated residual variances of the two reverse-worded items, i.e., item 5 and 8). CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

Table 7

Goodness of Fit Statistics for the One-Factor Model of the CES-D 10 and Tests of Factorial Invariance across Hispanic/Latino Background Group

Model	S-Bχ²	Df	P	CFI	RMSEA	Model Comparison	CFI	RMSEA
M1. Configural	1467.557	204	<.001	.985	.049	--	--	--
M2. Metric	1303.701	249	<.001	.987	.041	M2-M1	.002	-.008
M3. Scalar	2974.421	394	<.001	.969	.050	M3-M2	-.018	-.009
M3a. Scalar (modified)	2312.386	385	<.001	.977	.044	M3a-M2	-.010	-.008

Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation. M3 = Partially invariant scalar model (no modifications). M3a = Partially invariant scalar model (Item 5 [free thresholds 1 and 2], item 8 [free thresholds 1 and 2], item 2 [free threshold 1], and item 7 [free threshold 3]) in the Mexican and Cuban background group).

Table 8

Unstandardized Factor Loadings and Descriptive Statistics from Baseline Models of the CES-D 10 by Language Groups (N = 15,487)

Item	Spanish Loading	Spanish M (SD)	English Loading	English M (SD)
Item 1	1.00	.54 (.83)	1.00	.58 (.86)
Item 2	1.17	.67 (.93)	1.17	.82 (.96)
Item 3	2.05	.80 (.99)	2.05	.75 (1.00)
Item 4	1.58	.70 (1.01)	1.58	.98 (1.11)
Item 5	0.28	.84 (1.10)	0.28	.96 (1.09)
Item 6	1.37	.47 (.83)	1.37	.51 (.87)
Item 7	1.07	.89 (1.07)	1.07	1.06 (1.13)
Item 8	0.96	.82 (1.01)	0.96	.87 (.95)
Item 9	1.47	.72 (1.03)	1.47	.73 (1.00)
Item 10	1.26	.72 (.95)	1.26	.62 (.91)

Note. The factor loading for the first item was fixed to 1 to set the metric for the latent variable; all ps < .001.