Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Measuring Depressive Symptoms Among Latinos in the US: A Psychometric Evaluation of the CES-D Boston Form

Sumeyra Sahbaz (sahbaz@utexas.edu)
University of Texas at Austin
Pablo Montero-Zamora
University of Texas at Austin
Aigerim Alpysbekova
University of Texas at Austin
Christopher P. Salas-Wright
Boston College
Augusto Pérez-Gómez
Corporación Nuevos Rumbos
Juliana Mejía-Trujillo
Corporación Nuevos Rumbos
Saskia R. Vos
University of Miami
Carolina Scaramutti
University of Miami
Eric C. Brown
University of Miami
Mildred M. Maldonado-Molina
University of Florida
Melissa M. Bates
University of Florida
Maria Fernanda Garcia
Boston College
Maria Duque
Boston College
María Piñeros-Leaño
Boston College
Seth J. Schwartz
University of Texas at Austin

Research Article

Keywords: Depression, CES-D, Measurement, Latino, CFA, Psychometrics

Posted Date: January 29th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3891054/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose: We examined the extent to which depressive symptomatology measures operate across different Latino subgroups as there is inconsistency regarding its performance across Latinos, a large and rapidly growing cultural group in the United States.

Methods: We evaluated the reliability and structural validity of the scores generated by the Center for Epidemiologic Studies Depression Boston Form (CES-D-B) using four distinct Latino samples residing in US: Mexicans, Venezuelans, Cubans, and “other Latinos” (total *N*=1033). To further explore structural validity of CES-D-B scores, we conducted measurement invariance analyses across different countries of origin, gender groups, educational levels, and languages of assessment (English, Spanish).

Results: For all four samples, CES-D-B scores were highly reliable, and the factor structure had a good to excellent fit to the data. While measurement invariance analyses for different educational levels indicated scalar invariance across all samples, the same level of measurement equivalency was achieved only for Mexicans and Venezuelans with varying gender and languages of assessment.

Conclusions: The findings indicated that CES-D-B scores are internally consistent, possess a strong four-factor structure, and have somewhat equivalent psychometric properties across diverse Latino groups. Findings from this study highlight the importance of considering gender and languages of assessment when assessing depressive symptoms of various Latino subgroups.

Introduction

According to the World Health Organization, depression is projected to emerge as the second most significant cause of illness, disability, and mortality worldwide by the year 2030, thereby becoming the primary contributor to disease burden in high-income nations [1, 2]. In the United States, depression is a prevalent issue within primary care medical settings [3, 4], as well as in the general population, where an estimated 4.92% of individuals are believed to experience clinical depression [5]. Furthermore, research has indicated that Latinos residing in the United States (US) are at increased risk for depression compared to non-Latino Whites [6]. The additional burdens that many minority individuals face in the US, such as encounters with discrimination, low-income levels, limited educational opportunities, anxieties related to legal status, as well as high rates of unemployment and underemployment, may contribute to a more rapid deterioration of their mental health [7]. Although the Latino population in the US continues to experience rapid growth and now represents approximately 18% of the total US population [8], there is still a limited understanding of the mental health disparities present within this group, despite the evidence indicating higher rates of depressive symptomatology among Latinos compared to their non-Latino White counterparts [6]. It is, therefore, imperative to proactively screen and identify individuals with elevated depressive symptoms to mitigate the impact of this global public health concern [9]. In order to adequately address the mental health concerns of minority populations and evaluate how well

interventions work, it is crucial for researchers and practitioners to give priority to establishing measures that provide valid scores of the construct being assessed [10].

Although, self-report screening measures have gained popularity as brief and easily administered tools in clinical and research settings, an expanding body of literature suggests that the utility of these measures cannot be assumed without considering respondents' cultural and linguistic context. Various aspects of cultural context, including regional variations in language and dialect [11], differences in beliefs and values [12], variations in how questions and response options are perceived and understood [13], as well as levels of acculturation [14] can influence psychometric properties of measures of depressive symptoms and related constructs. Importantly, failure to adequately assess and consider cultural context can introduce systematic bias into research findings [15]. Therefore, it is crucial to evaluate the cross-cultural validity of the scores gained from psychosocial measures that have been validated using primarily non-Latino White samples [16-18]. Careful cross-cultural measurement research will allow us to account for the inherent cultural differences among individuals from different genders, country of origin, language, and educational level groups. To this end, in the present study we evaluated the validity of the scores gained from the Center for Epidemiologic Studies Depression Boston Form (CES-D-B), a widely used measure of depressive symptoms, among four distinct samples of Latino adults (Mexicans, Venezuelans, Cubans, and a fourth group composed of Latinos from diverse Latin American countries) that vary by gender, national origin, educational level, and preferred language.

Background

The Center for Epidemiologic Studies Depression Scale (CES-D) is a widely used self-report questionnaire specifically developed to measure the frequency and severity of depressive symptoms in the general population [19]. The CES-D scale consists of 20 items covering four dimensions of depressive symptoms: depressed affect, positive affect, somatic and retarded activity, and interpersonal. The scale assesses the frequency of these symptoms over the week prior to assessment using a four-point Likert scale ranging from 0 (rarely or none of the time) to 3 (most or all the time). Total score can range from 0 to 60, with higher scores indicating more severe depressive symptoms. CES-D scale has been found to serve as a reliable screening tool for assessing and monitoring changes in depressive symptoms for both clinical and non-clinical populations.

CES-D was first translated into Spanish by Roberts [20], who tested the factor structure of the 20-item scale among White American, African Americans, and Mexican Americans and found similar internal consistency and factor structure for the different samples. Subsequentially, various short forms of the CES-D were developed to include nine, ten, or eleven items [21–24]. Kohout and colleagues [21] developed two short forms. One of these was the Boston Form, which consisted of 10 items (two of which were worded positively) and covering the original four dimensions. The CES-D Boston Form (CES-D-B) was validated with adults from different U.S states [21], various general and clinical female populations [25], and older adults [26]. The Spanish version of the CES-D-B was initially utilized and validated by Grzywacz et al. [27] to assess depressive symptoms among Mexican immigrants residing in various U.S states.

Another study conducted by Grzywacz et al. [28] evaluated the psychometric properties specifically for use with Latino farmworkers. This Spanish version of the 10-item scale was validated more recently with Bolivian patients [29]. Although the developer of the CES-D-B [21] proposed a four-factor structure for the scale (depressed affect, positive affect, somatic and retarded activity, and interpersonal problems), other researchers have argued that the factor structure may differ across distinct groups of people. More specifically, some researchers (e.g., [25]) have identified a unidimensional structure with various female samples, whereas others (e.g., [26]) have found a three-factor structure when assessing depressive symptoms among older adults. Moreover, Grzywacz et al. [27] suggested a three-factor structure with Mexican immigrants, whereas Yu et al. [30] found a two-factor structure when assessing depressive symptoms among Chinese clinical and general samples, and Schantz et al. [29] also proposed a two-factor structure when validating with the scale with patients in Bolivia. The fact that different factor structures have been proposed is cumbersome for researchers interested in studying depressive symptoms among populations of interest and making meaningful comparisons among these different populations. Given the differences in factor structure obtained with different cultural groups, there appears to be a need to account for cultural influences when assessing validity of CES-D-B scores. In response to this need, we conducted a comprehensive evaluation of the psychometric properties of this scale with four distinct samples of Latino adults that varying across national origin, gender, educational level, and language of assessment.

The Present Study

The purpose of the present study was twofold. First, we evaluated the psychometric properties, such as reliability and validity of scores generated by the CES-D-B assessing depressive symptoms among different groups. Next, we tested the model fit of the proposed four-factor structure for all four Latino samples using confirmatory factor analysis (CFA). To establish measurement invariance, we examined equivalence of factor structure (configural invariance), factor loadings (metric invariance), and item intercepts (scalar invariance) using multi-group confirmatory factor analysis (MGCFA). Specifically, we compared the structure of CES-D-B scores (a) among Mexicans, Venezuelans, Cubans, and “other Latinos” residing in the US and (b) across gender, educational level, and language of assessment. These comparisons were intended to allow us to evaluate the extent to which the factor structure of CES-D-B depressive symptoms scores would be equivalent among diverse groups of US Latinos.

Methods

Participants and Procedures

In the present study, we investigated factor structure and psychometric properties of CES-D-B scores in a sample including Latinos (N = 1033) residing in the US by analyzing survey responses from four samples. Sample 1 consisted of Venezuelans (*n* = 433), sample 2 was made up of Mexicans (*n* = 288), sample 3 consisted of Cubans (*n* = 125), and the remaining participants from other Latin American countries comprised sample 4 (*n* = 236), which will be referred to as “other Latinos”.

The present study utilized data from two studies of Latinos residing in the US. The Venezuelans in sample 1 were US-based participants in the Colombia and Miami's Newest Arrivals (CAMINAR) study, which collected data from newly arrived Venezuelan migrant parents in South Florida during the months of October and November of 2017. The remaining samples consisted of participants in a study investigating cultural stress, psychological distress, family dynamics, and alcohol use among Latino parents residing in the US with children between the ages of 8 and 16 years. This second data collection took place between June 2022 and February 2023. For both studies, eligible participants completed a 30-minute online survey (in English or Spanish, depending on participants' preferences) covering demographic information, cultural stress, and mental health measures. Participants received compensation upon completing the survey. The survey was developed using Qualtrics Survey Software. Approval for these studies was obtained from the associated universities' Institutional Review Boards.

The majority of participants in all four samples were women, accounting for 65%, 60%, 77%, and 75% of the respective samples. Furthermore, in each sample more than 50% of participants were 40 years of age or younger. For the Venezuelan sample, 67% of the participants had at least a college degree or more and 89% of them took the survey in Spanish. The percentage of individuals with a college degree was 35% for the Mexican sample. However, most of the Mexican sample preferred to take the survey in English. For the Cuban sample, 40% had obtained at least a college degree and about 60% completed the survey in Spanish. Within the "other Latinos" sample, 40% of the participants had obtained at least a college degree and 63% took the survey in English. In the sample as a whole, 67% of participants were women, about 50% had obtained at least a college degree, slightly more than 50% were 40 years of age or younger, and 57% completed the survey in Spanish.

Measures

Depressive Symptoms. The Center for Epidemiologic Studies Depression Boston Form (CES-D-B: [21]) was used to measure symptoms of depression, and participants were asked to indicate how often within the week prior to assessment they had experienced each of the ten symptoms. Example items include "*I felt depressed"*, and "*I could not get going"*. Responses were recorded on a 4-point Likert scale ranging from 0 (Rarely or none of the time) to 3 (Most or all of the time). Two positively worded items, "*I was happy"*, and "*I enjoyed life"*, were reverse coded prior analyzing the data so that higher scores indicate higher depressive symptomatology. The scale is internally consistent as indicated by Cronbach's alpha coefficients ranging from .83 to .88 for all four samples.

Analytic Plan

Preliminary Analyses. To gain a basic understanding of how CES-D-B scores were distributed, descriptive statistics such as mean and standard deviations were examined for all four samples (Venezuelans, Mexicans, Cubans, and "other Latinos"). To identify any potential redundancy among the scale items, we examined the pattern of inter-item correlations. Additionally, we assessed internal consistency before proceeding with subsequent analyses. Finally, for all samples we examined associations of the CES-D-B scores with measures of anxiety and stress to determine concurrent validity.

Confirmatory Factor Analysis (CFA). The fit of the CES-D-B items to the proposed four-factor solution was evaluated using confirmatory factor analysis (CFA). Absolute fit indices, including the chi-square test (χ^2), root mean square error of approximation (RMSEA), and standardized root mean residual (SRMR), were utilized, along with relative fit indices such as the Tucker-Lewis Index (TLI) and comparative fit index (CFI). Acceptable model fit was indicated by SRMR values below .05, CFI and TLI values above .90, and RMSEA values below .08, as recommended by Hu and Bentler [31].

Measurement Invariance. To offer further validity evidence for the CES-D-B scores, we evaluated whether the construct measured was equivalent across different groups of individuals. To assess measurement invariance, which is essential when making group comparisons, we utilized multi-group confirmatory factor analysis modeling as it represents the most powerful and versatile approach to evaluating measurement invariance [32]. Measurement invariance ensures that a construct can be meaningfully assessed within each group and holds a consistent meaning across different groups [33, 34]. Consequently, we ran between-group and within-group analyses and investigated whether the proposed model for the CES-D-B remained consistent across various groupings, including country of origin, gender, educational level, and languages of assessment. The results of the measurement invariance tests were used to determine whether the same construct is measured across groups, and whether group differences emerging from subsequent analyses were likely reflect true group differences rather than differences in measurement structure.

To investigate measurement invariance for all four samples, we followed the stepwise approach, which allows us to identify the point at which invariance is no longer achieved between groups [35]. In this approach, we progressively imposed equality constrains on factor loadings and item intercepts, evaluating whether each more constrained model differed from the less constrained mode that preceded it.

We began with configural invariance, allowing factor loadings and item intercepts to be freely estimated while examining consistency in the number of factors and in the patterns of indicator-factor relationships across groups. Additionally, we assessed the degree to which the factor structure fit all groups (different gender, educational level, and languages of assessment) equally well. Provided that configural invariance was supported across groups, we proceeded to the next level of measurement invariance. At the metric or weak factorial invariance level, we constrained factor loadings equal across groups and evaluated the decrease in model fit associated with this constraint. Metric invariance evaluates the extent to which individual item responses exhibited similar relationships with the underlying construct (depressive symptoms) across groups. Metric invariance is supported if model fit does not degrade significantly relative to the fit of the configural invariance model. Provided that metric invariance is supported across groups, then the scalar invariance would be examined. For scalar or strong factorial invariance, item intercepts (in addition to factor loadings) are constrained equal across groups, and the resulting decrease in model fit is examined. Scalar invariance allowed us to ascertain whether individuals in different groups, and who share the same levels of the underlying latent construct (depression), would receive the same raw scores. The absence of scalar invariance, known as differential item functioning [36], would

indicate that individuals in one group provided different raw scores than individuals in another group, despite having the same level of the underlying trait. We expect that the suggested four-factor structure will fit adequately to the CES-D-B data collected from Mexicans, Venezuelans, Cubans, and "other Latinos" and that within-group measurement invariance tests will indicate measurement equivalence for groups of different gender, educational level, and language of assessment.

Results

Preliminary Analyses

On average, Mexicans reported the highest depressive symptoms scores (19.20), whereas Cubans reported the lowest score (17.10). For Venezuelans and Mexicans, the item with the highest mean response was "*Everything was an effort".* On the other hand, among Cubans and "other Latinos", the item with the highest mean response was "*I Was happy"*. For all four samples, inter-item correlations among the CES-D Boston Form items did not exceed ($r > .70$), suggesting that redundancy among the items was not a concern.

Our analyses provided support for the concurrent validity of CES-D-B scores within our sample, as evidenced by significant positive correlations between the scale and scores related to stress and anxiety. For all four samples, correlations between the CES-D-B and stress scores were statistically significant and ranged between $r = .16$ and $r = .40$. Correlations between the CES-D-B and anxiety varied between $r = .67$ and $r = .82$.

Factor Structure of CES-D Boston Form Scores

A confirmatory factor model was estimated for the CES-D-B across all four samples. Following the scoring algorithm for the CES-D-B [21], ten items were specified as loading on four factors. More specifically, items "*Felt depressed"*, "*Felt lonely"*, and "*Felt sad"* load on the factor Depressed Affect, items "*I was happy"* and "*Enjoyed life"* load on the factor Positive Affect, items "*Everything was effort"*, *"Sleep was restless"*, and "*Could not get going"* load on factor Somatic and Retarded Activity, and items "*People unfriendly"* and "*People disliked me"* load on factor Interpersonal Problems.

For all samples the four-factor structure was supported by the data (see Table 1). More specifically, for Venezuelans, the four-factorial structure provided an excellent fit to the data, CFI = .99, TLI = .99, RMSEA = .02 (90% CI = [.00, .05]), SRMR = .02, $\chi^2(29, N = 433) = 31.26$, $p > .05$. For Mexicans, the model provided a good fit to the data, comparative fit index (CFI) = .98, Tucker-Lewis index (TLI) = .96, root mean square error of approximation (RMSEA) = .06 (90% CI = [.03, .08]), standardized root mean square residual (SRMR) = .04, $\chi^2(29, N = 288) = 55.33$, $p > .05$. For Cubans, the factorial solution was excellent, comparative fit index (CFI) = .99, Tucker-Lewis index (TLI) = .99, root mean square error of approximation (RMSEA) = .02 (90% CI = [.00, .07]), standardized root mean square residual (SRMR) = .04, $\chi^2(29, N = 125) = 31.03$, $p > .05$. And finally for "other Latinos" sample, the four-factorial had an acceptable fit to the data with comparative fit index (CFI) = .96, Tucker–Lewis index (TLI) = .93, root mean square error of

approximation (RMSEA) = .07 (90% CI = [.05, .09]), standardized root mean square residual (SRMR) = .04, $\chi^2(29, N = 236) = 364.73$, $p < .05$.

Table 1
Descriptive Statistics and Confirmatory Factor Analysis Parameter Estimates

CES-D Boston Form	Mean				Estimate/Factor Loadings				Standard Errors			
Items	**Ven**	**Mex**	**Cub**	**Oth**	**Ven**	**Mex**	**Cub**	**Oth**	**Ven**	**Mex**	**Cub**	**Oth**
Felt depressed	1.76	1.77	1.58	1.59	.87	.83	.71	.74	.02	.03	.06	.05
Everything was effort	2.18	2.22	1.86	1.92	.52	.58	.63	.62	.05	.05	.08	.05
Sleep was restless	1.93	2.09	1.76	1.98	.71	.74	.69	.64	.04	.03	.07	.05
Was happy.	1.96	2.03	2.00	2.03	.87	.68	.81	.75	.07	.07	.08	.07
Felt lonely.	1.94	1.88	1.59	1.73	.80	.80	.71	.77	.03	.03	.07	.05
People unfriendly.	1.55	1.75	1.47	1.51	.71	.71	.50	.64	.05	.05	.12	.07
Enjoyed life.	1.94	1.98	1.88	1.80	.74	.89	.83	.90	.07	.07	.09	.05
Felt sad.	1.92	1.90	1.74	1.84	.80	.82	.78	.82	.03	.03	.06	.03
People disliked me.	1.44	1.76	1.42	1.54	.81	.84	.80	.91	.04	.03	.11	.05
Could not get going.	1.72	1.91	1.74	1.93	.80	.74	.80	.76	.03	.04	.06	.04

CFA Model Fit	Venezuelans	Mexicans	Cubans	"Other Latinos"
$\chi 2$ (df)	31.26 (29)	55.33 (29)	31.03 (29)	64.73* (29)
CFI/TLI	.99/.99	.98/.96	.99/.99	.96/.93
RMSEA	.02	.06	.02	.07
RMSEA 90% CI	.00 .05	.03 .08	.00 .07	.05 .09
SRMR	.02	.04	.04	.04
AIC	6338.62	6132.76	2620.30	4765.87

Note. Prompt of the CES-D Boston Form is: Below is a list of ways you might have felt or behaved. Please indicate how often you have felt this way during the past week. * $p < .05$. For all factor loadings, the p-value is less than .001. Items 1, 5, and 8 load on factor depressed affect, 4 and 7 on factor positive affect, 2, 3, and 10 on factor somatic complaints, and 6 and 9 on interpersonal problems.

Measurement Invariance of the CES-D Boston Form

Within-group measurement invariance tests were conducted to study equivalence among individuals of different gender, educational level, and languages of assessment. We used the stepwise approach (configural, metric, and scalar) to examine the equivalence of CES-D-B scores across gender, educational levels, and languages of assessment (English and Spanish) for Venezuelans, Mexicans, Cubans, a fourth group composed of Latinos from diverse Latin American countries (“other Latinos”) as well as for the overall sample. The results of the within-group measurement invariance tests across gender, educational levels, and languages of assessment for all four samples are presented in the following section of the study.

Measurement Invariance across Gender Groups

When we conducted measurement invariance tests for different gender groups within each sample, we have found scalar invariance for the samples consisting of Venezuelans and Mexicans. However, for the samples consisting of Cubans, the fourth group composed of Latinos from diverse Latin American countries as well as for the overall sample we found metric invariance across gender groups (see Table 2). More specifically, for Venezuelans and Mexicans, the number of factors, the patterns of item-factor relationships, as well as intercepts are equivalent for males and females. However, for the remaining samples, we obtained metric invariance indicating that we have partial measurement equivalence for the CES-D-B when used with men and women from Cuba, and other Latin American countries.

Table 2
CES-D Boston Form Gender Invariance Tests

CES-D Boston Form	χ^2 (df), p-value	CFI/TLI	RMSEA	p-value (χ^2diff test)
Venezuelans				
Configural Invariance	73.5 (58), p > .05	.99/.99	.02	-
Metric Invariance	80.01 (64), p > .05	.99/.98	.02	.34
Scalar Invariance	85.7 (70), p > .05	.99/.98	.02	.42
Mexicans				
Configural Invariance	117.5 (58), p < .00	.96/.93	.06	-
Metric Invariance	123.8 (64), p < .00	.96/.94	.06	.42
Scalar Invariance	136.19 (70), p < .00	.95/.94	.06	.06
Cubans				
Configural Invariance	82.11 (58), p < .05	.95/.92	.06	-
Metric Invariance	98.25 (64), p < .01	.92/.90	.06	.01
Scalar Invariance	107.20 (70), p < .01	.92/.90	.06	.01
“Other Latinos”				
Configural Invariance	98.60 (58), p < .01	.96/.94	.05	-
Metric Invariance	105.63 (64), p < .01	.96/.95	.05	.32
Scalar Invariance	122.05 (70), p < .01	.95/.94	.06	.01
Overall sample				
Configural Invariance	157.53 (58), p < .01	.98/.97	.04	-
Metric Invariance	159.68 (64), p < .01	.98/.97	.04	.90
Scalar Invariance	194.45 (70), p < .01	.97/.96	.04	.01

Notes. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation; SRMR: standardized root mean square residual. For Cubans sample, 23% were male and for “other Latinos” sample, we had 25% male only.

Measurement Invariance across Educational Levels

To assess measurement invariance for individuals of different educational levels (college graduate vs. less), we evaluated the model fit of the increasingly constrained models. For all samples, tests indicated no significant differences in model fit after constraining the number of factors, factor loadings, or item intercepts, indicating scalar invariance across different educational levels. This finding suggests that

comparisons between participants with and without a college degree are appropriate for all samples (see Table 3).

Table 3
CES-D Boston Form Educational Level Invariance Tests

CES-D Boston Form	χ^2 (df), p-value	CFI/TLI	RMSEA	p-value (χ^2diff test)
Venezuelans				
Configural Invariance	68.80 (58), p > .05	.99/.99	.02	-
Metric Invariance	72.63 (64), p > .05	.99/.99	.02	.54
Scalar Invariance	82.20 (70), p > .05	.99/.99	.02	.12
Mexican				
Configural Invariance	100.39 (58), p < .01	.97/.95	.05	-
Metric Invariance	108.66 (64), p < .01	.97/.95	.05	.23
Scalar Invariance	111.62 (70), p < .01	.97/.95	.05	.80
Cubans				
Configural Invariance	78.54 (58), p > .05	.95/.93	.05	-
Metric Invariance	84.67 (64), p = .04	.95/.93	.05	.42
Scalar Invariance	96.81 (70), p < .05	.94/.92	.06	.06
"Other Latinos"				
Configural Invariance	112.93 (58), p < .05	.95/.92	.06	-
Metric Invariance	122.63 (60), p < .05	.95/.92	.06	.12
Scalar Invariance	132.31 (70), p < .05	.94/.92	.06	.12
Overall sample				
Configural Invariance	144.49 (58), p < .05	.98/.97	.04	-
Metric Invariance	145.91 (64), p < .05	.98/.97	.04	.98
Scalar Invariance	153.32 (70), p < .05	.98/.97	.04	.23

Notes. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation; SRMR: standardized root mean square residual.

Measurement Invariance across Languages of Assessment

When we ran within-group measurement invariance tests for different languages of assessment, we found scalar invariance for Venezuelans, Mexicans, and "other Latinos". However, for Cubans, we

obtained configural invariance only which indicates that we have same factor structure for individuals who took the survey in English and Spanish (see Table 4). However, factor loadings and item intercepts for Cuban participants were not equivalent across participants completing the CES-D-B in English versus Spanish. Overall, for Venezuelans, Mexicans, and “other Latinos” we can make meaningful comparison of the individuals who took the survey in different languages.

Table 4
CES-D Boston Form Language Invariance Tests

CES-D Boston Form	χ^2 (df), p-value	CFI/TLI	RMSEA	p-value (χ^2diff test)
Venezuelans				
Configural Invariance	87.32 (58), p < .01	.98/.98	.03	-
Metric Invariance	92.20 (64), p < .01	.98/.98	.03	.54
Scalar Invariance	104.31 (70), p < .01	.98/.98	.03	.06
Mexican				
Configural Invariance	123.68 (58), p < .01	.95/.92	.06	-
Metric Invariance	129.96 (64), p < .01	.95/.93	.06	.32
Scalar Invariance	138.56 (70), p < .01	.95/.93	.06	.17
Cubans				
Configural Invariance	101.01 (58), p < .01	.91/.86	.08	-
Metric Invariance	116.77 (64), p < .01	.90/.84	.08	.01
Scalar Invariance	118.11 (70), p < .01	.90/.87	.08	.01
“Other Latinos”				
Configural Invariance	113.59 (58), p < .01	.95/.92	.06	-
Metric Invariance	118.62 (64), p < .01	.95/.93	.06	.54
Scalar Invariance	122.58 (70), p < .01	.95/.93	.06	.67
Overall sample				
Configural Invariance	153.63 (58), p < .01	.98/.97	.04	-
Metric Invariance	157.31(64), p < .01	.98/.97	.04	.68
Scalar Invariance	170.14 (70), p < .01	.98/.97	.04	.04

Notes. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation; SRMR: standardized root mean square residual.

Overall, results of the within-group measurement invariance tests revealed scalar invariance across gender groups in samples comprising Venezuelans and Mexicans. However, for Cubans and the overall sample, metric invariance was found, indicating partial measurement equivalence for CES-D-B when used with men and women from Cuba and other Latin American countries. Measurement invariance tests across different educational levels, including college graduates and those with less education, revealed scalar invariance indicating consistent model fit without significant differences in factors, factor loadings, or item intercepts across all samples. This implies that comparisons between participants with and without a college degree are deemed appropriate for all groups. And finally, within-group measurement invariance tests for different languages of assessment revealed scalar invariance for Venezuelans, Mexicans, and "other Latinos", allowing meaningful comparisons across individuals who took the survey in different languages. However, for Cubans, only configural invariance was observed, indicating a similar factor structure for those who completed the survey in English and Spanish, but non-equivalence in factor loadings and item intercepts between the two language groups (see Table 5 for a summary).

Table 5
Within-group Measurement Invariance Tests for Gender, Educational Levels, and Languages of Assessment

	Venezuelans	Mexicans	Cubans	"Other Latinos"	Total Sample
Gender Groups	Scalar Invariance	Scalar Invariance	Configural Invariance	Metric Invariance	Metric Invariance
Educational Levels	Scalar Invariance	Scalar Invariance	Scalar Invariance	Scalar Invariance	Scalar Invariance
Languages of Assessment	Scalar Invariance	Scalar Invariance	Configural Invariance	Scalar Invariance	Metric Invariance

Discussion

The Center for Epidemiologic Studies Depression Boston Form (CES-D-B) scale is a commonly used measure assessing depressive symptoms of general populations. Although, the measure was used with various with Latinos residing in the U.S [27, 28, 37], there is inconsistency regarding the psychometric properties found in different studies. The aim of the current project was to study the extent to which CES-D-B measures depressive symptomatology across different Latino subgroups by systematically examining reliability, validity, and measurement invariance across gender, educational level, and language of assessment.

In the present study, we evaluated the psychometric properties and measurement equivalence of the CES-D-B scores using four samples of Latino subgroups residing in the United States. First, we examined the internal consistency of the measure and its concurrent validity using its correlation with other mental health measures (anxiety and stress). Second, we evaluated the between-group measurement

equivalence by providing evidence for structural validity across the four samples. And finally, we investigated within-group measurement equivalence for the four samples separately using measurement invariance tests across gender groups, educational levels, and languages of assessment.

The first key finding involves supporting the four-factor structure in the original CES-D scale which was indicated by good to great fit for all four samples. The number of factors (Depressed Affect, Positive Affect, Somatic and Retarded Activity, Interpersonal Problems) as well as the item-factor constellations were equivalent for Mexicans, Venezuelans, Cubans, and “other Latinos” group. This finding aligned with our hypothesis prior running the confirmatory factor analyses. The second key finding arises from within-group measurement invariance results, which indicated scalar invariance for individuals with differing educational levels among Mexicans, Venezuelans, Cubans, and “other Latinos. Thus, our hypothesis was supported by the results, and obtaining scalar invariance means that the same factor structure fits the data well, the meaning of the latent construct is equivalent, and it is appropriate to compare observed means and run analyses such as ANOVAs and t-tests across all study samples. The third key finding involves supporting scalar invariance for languages of assessment among Mexicans, Venezuelans, and “other Latinos” which supports our hypothesis and indicates that the analyses mentioned above are appropriate for these groups. The fourth key finding arises from measurement invariance test for gender groups which indicates scalar invariance for Mexicans and Venezuelans, metric invariance for “other Latinos” and configural invariance for Cubans. The lack of scalar invariance for languages of assessment and gender groups in the Cuban sample could be attributed to the small size as well as the unequal representation of the two groups.

The present study contributes to the literature by providing considerable evidence that the CES-D-B captures the same meaning across various Latin American groups of differing educational levels and languages of assessment. Studies such as ours are crucial if researchers and practitioners are to have maximal confidence in using the CES-D-B across Latino groups with different characteristics. Without evidence of measurement equivalence across groups, prevention and treatment intervention may be judged as effective when they are not, (or may not be judged to be effective when they are). Both erroneous conclusions carry, risk of harming participants and of wasting valuable resources and time. Further, culturally valid and reliable instruments are needed to ensure that we can accurately assess the burden of depressive symptomatology within and across Latino subgroups.

Limitations

The present findings, vis-à-vis examining the CES-D-B for use with various Latino groups residing in the US, should be interpreted considering several limitations. First, Latino migrants are a heterogenous population. Although our study was conducted using four different Latino samples with relatively large sizes, the size of the Cuban sample was relatively small. A larger sample of Cubans would help researchers gain a more in-depth understanding of the extent of measurement invariance within that group. Second, we had to group individuals from various Latin American countries into the “other Latino” group and assumed homogeneity within this group. Future research could benefit from using a larger

sample that would allow further investigation of the other Latino group (including validity, reliability and measurement invariance). Third, our study was cross-sectional, and as such, studies tracking patterns of interindividual variability and intraindividual change using longitudinal methods may represent an important future advance using the CES-D-B with Latino migrant samples.

In conclusion, the scale CES-D-B provides a robust assessment of Latinos' depressive symptoms and allows meaningful comparisons of depression among groups of differing countries of origin, educational level, and preferred language. CES-D-B is brief, simple to administer, and is appears to be appropriate for use with Latino adults. We hope that the present results will inspire more work in this direction.

Declarations

>

Author Contribution

S. S. wrote the main manuscript and ran statistical analyses to develop the methods and results sections of the manuscript. P. M.Z. provided part of the data analyzed in the manuscript and assisted with the statistical analyses.A. A. conducted literature review and contributed to the introduction and literature review sections of the manuscript.C. P. S. W. provided part of the data analyzed in the manuscript and contributed to the development of the manuscript. S. J. S. provided part of the data analyzed in the manuscript and contributed to the development of the manuscript.All other others have either helped with either research design, data collection, data management, or data cleaning in addition to contributing to the development of the manuscript.

References

1. Yu B, Zhang X, Wang C, Sun M, Jin L, Liu X (2020) Trends in depression among Adults in the United States, NHANES 2005–2016. J Affect Disord 263:609–620. https://doi.org/10.1016/j.jad.2019.11.036
2. Institute of Health Metrics and Evaluation (IHME) (2019) Global Health Data Exchange (GHDx). https://vizhub.healthdata.org/gbd-results/ (Accessed 4 March 2023)
3. Hasin DS, Grant BF (2015) The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. Soc Psychiatry Psychiatr Epidemiol 50:1609–1640. https://doi.org/10.1007/s00127-015-1088-0
4. Vos, T, Abajobir, A A, Abate, K H, Abbafati, C, Abbas, K M, Abd-Allah, F, … Criqui, M H (2017) Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100): 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2
5. Institute for Health Metrics and Evaluation (IHME) (2020) GBD compare data visualization. http://vizhub.healthdata.org/gbd-compare. Accessed 10 Jan 2023

6. Mendelson T, Rehkopf DH, Kubzansky LD (2008) Depression among Latinos in the United States: A meta-analytic review. J Consult Clin Psychol 76(3):355–366. http://dx.doi.org/10.1037/0022-006X.76.3.355
7. Schwartz S, Meyer IH (2010) Mental health disparities research: The impact of within and between group analyses on tests of social stress hypotheses. Soc Sci Med 70(8):1111–1118
8. United States Census Bureau (2021) *2020 Census Demographic Data Tables: Hispanic or Latino, and not Hispanic or Latino by Race*. https://data.census.gov/table?q=Latinos&g=010XX00US&tid=DECENNIALPL2020.P2
9. Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR (2011) A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English-and Spanish-speaking Latinas. Cult Divers ethnic Minor Psychol 17(3):309–316. https://doi.org/10.1037/a0023883
10. Arafat S, Chowdhury HR, Qusar M, Hafez M (2016) Cross cultural adaptation & psychometric validation of research instruments: A methodological review. J Beh Health 5(3):129–136
11. Onodera N, Smith T, Harkness J, Mohler P (2005) Methods for Assessing and Calibrating Response Scales Across Countries and Languages. Comp Sociol 4(3–4):365–415
12. Benítez I, He J, Van de Vijver FJ, Padilla JL (2016) Linking extreme response style to response processes: A cross-cultural mixed methods approach. Int J Psychology 51(6):464–473
13. Davis RE, Resnicow K, Couper MP (2011) Survey response styles, acculturation, and culture among a sample of Mexican American adults. J Cross-cultural Psychology 42(7):1219–1236
14. Tan YW, Burgess GH, Green RJ (2021) The effects of acculturation on neuropsychological test performance: A systematic literature review. Clin Neuropsychol 35(3):541–571
15. Choi BC, Pak AW (2005) A catalog of biases in questionnaires. Prev Chronic Dis 2(1):A13
16. Allen J, Walsh JA (2000) A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research. In: Dana RH (ed) Handbook of cross-cultural and multicultural personality assessment. Personality and clinical psychology series. Erlbaum, Mahwah, NJ, pp 63–85
17. Bravo M (2003) Instrument development: Cultural adaptations for ethnic minority research. In: Bernal G, Trimble JE, Burlew AK, Leong FT (eds) Handbook of racial and ethnic minority psychology. Sage, Thousand Oaks, CA, pp 220–236
18. Padilla AM, Medina A (1996) Cross-cultural sensitivity in assessment. Using tests in culturally appropriate ways. In: Suzuki LA, Meller PJ, Ponterotto JG (eds) Handbook of multicultural assessment: Clinical, psychological, and educational applications. Jossey-Bass, San Francisco, pp 3–28
19. Radloff LS (1977) The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1(3):385–401
20. Roberts RE (1980) Reliability of the CES-D scale in different ethnic contexts. Psychiatry Res 2(2):125–134

21. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J (1993) Two shorter forms of the CES-D depression symptoms index. J Aging Health 5(2):179–193
22. Andresen EM, Malmgren JA, Carter WB, Patrick DL (1994) Screening for depression in well older adults: Evaluation of a short form of the CES-D. Am J Prev Med 10(2):77–84
23. Santor DA, Coyne JC (1997) Shortening the CES–D to improve its ability to detect cases of depression. Psychol Assess 9(3):233
24. Cole JC, Rabin AS, Smith TL, Kaufman AS (2004) Development and validation of a Rasch-derived CES-D short form. Psychol Assess 16(4):360–372. https://doi.org/10.1037/1040-3590.16.4.360
25. Carpenter J, Andrykowski M, Wilson J, Hall L, Kay Rayens M, Sachs B, Cunningham L (1998) Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. Issues Ment Health Nurs 19(5):481–494
26. Irwin M, Artin KH, Oxman MN (1999) Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). Archives of Internal Med 159(15):1701–1704
27. Grzywacz JG, Hovey JD, Seligman LD, Arcury TA, Quandt SA (2006) Evaluating short-form versions of the CES-D for measuring depressive symptoms among immigrants from Mexico. Hispanic J Behav Sci 28(3):404–424
28. Grzywacz JG, Quandt SA, Chen H, Isom S, Kiang L, Vallejos Q, Arcury TA (2010) Depressive symptoms among Latino farmworkers across the agricultural season: Structural and situational influences. Cult Divers Ethnic Minor Psychol 16(3):335–343. https://doi.org/10.1037/a0019722
29. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, Piette JD (2017) Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients. Int J Psychiatry Med 52(4–6):315–327
30. Yu SC, Lin YH, Hsu WH (2013) Applying structural equation modeling to report psychometric properties of Chinese version 10-item CES-D depression scale. Qual Quant 47:1511–1518. https://doi.org/10.1007/s11135-011-9604-0
31. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equation Modeling: Multidisciplinary J 6(1):1–55
32. Jöreskog KG, Moustaki I (2001) Factor analysis of ordinal variables: A comparison of three approaches. Multivar Behav Res 36(3):347–387
33. Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 9(2):233–255
34. Putnick DL, Bornstein MH (2016) Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Dev Rev 41:71–90. 10.1016/j.dr.2016.06.004
35. Brown TA (2015) Confirmatory factor analysis for applied research. Guilford
36. McDonald R (1999) Test theory: A unified treatment. Psychology Press

37. Schwartz SJ, Salas-Wright CP, Pérez-Gómez A, Mejía-Trujillo J, Brown EC, Montero-Zamora P, Meca A, Scaramutti C, Soares MH, Vos SR, Javakhishvili N, Dickson-Gomez J (2018) Cultural stress and psychological symptoms in recent Venezuelan immigrants to the United States and Colombia. Int J Intercultural Relations 67:25–34. https://doi.org/10.1016/j.ijintrel.2018.09.001