1-2021

Assessing the Psychometric Proprieties of the Attitudes Toward Seeking Professional Psychological Help Scale–Short Form (ATSPPH-SF) Among Latino Adults

Lucas Torres

Brooke E. Magnus

Natasha Suhail Najar

Assessing the Psychometric Proprieties of the Attitudes Toward Seeking Professional Psychological Help Scale–Short Form (ATSPPH-SF) Among Latino Adults

Lucas Torres
Psychology Department, Marquette University, Milwaukee, WI
Brooke Magnus
Psychology Department, Marquette University, Milwaukee, WI
Natasha Najar
Psychology Department, Marquette University, Milwaukee, WI

Abstract

The Latino population continues to underutilize mental health services at an alarming rate. The Attitudes Toward Seeking Professional Psychological Help Scale–Short Form (ATSPPH-SF) is one of the most commonly used instruments to assess help-seeking attitudes. The current study sought to evaluate the factor structure and test for the presence of differential item functioning on the ATSPPH-SF with a sample of Latino adult individuals

across nativity status (U.S.- vs. foreign-born), language format (English vs. Spanish), and gender. The analyses revealed two relatively independent factors named Openness to Seeking Treatment and Value and Need in Seeking Treatment. Measurement equivalence and practical implications are discussed in the context of use with Latino individuals.

Keywords

help-seeking, Latino/a, differential item functioning, DIF, measurement equivalence

The general underutilization of mental health care services among racial/ethnic minority individuals, including Latinos, is of growing concern. More specifically, despite prevalence rates that parallel or exceed their White American counterparts, racial/ethnic minorities are less likely to seek professional psychological help and are more often to terminate treatment prematurely (U.S. Department of Health and Human Services, 2001). For instance, among Latinos with a diagnosed mental disorder, only 9.3% reported using a mental health care provider (Vega et al., 2001). Additionally, mental health service use among Latino immigrants was even lower than U.S.-born Latinos, where fewer than 1 in 20 will contact a mental health care provider and less than 1 in 10 will seek services from a general practitioner (U.S. Department of Health and Human Services, 2001). A more recent epidemiological study supported these findings and reported that approximately 5% of Latinos sampled received mental health treatment despite 17% having received a psychiatric diagnosis in the past year (Burnett-Zeigler et al., 2018).

Empirical work has suggested that positive attitudes toward mental health care is associated with help-seeking behaviors (Mojtabai et al., 2002). Still, attitudes toward mental health treatment can vary across cultures, ethnicity, and gender (Gonzalez et al., 2011). Given the key role positive attitudes toward mental health treatment may have in seeking psychological help, it is important to be able to effectively and accurately assess this construct across groups. Unfortunately, minimal work has examined the psychometric properties of instruments measuring attitudes toward seeking mental health treatment among Latino samples. Latinos continue to be one of the fastest growing groups in the United States and, in 2017, were the nation's largest ethnic minority group constituting 18.1% of the total population (U.S. Census Bureau, 2018). The current study sought to evaluate the factor structure and test for the presence of differential item functioning (DIF) on the Attitudes Toward Seeking Professional Psychological Help Scale–Short Form (ATSPPH-SF; Fischer & Farina, 1995) with a sample of Latino adult individuals across nativity status (U.S.- vs. foreign-born), language format (English vs. Spanish), and gender.

ATSPPH-SF

The ATSPPH-SF was adapted from Fischer and Turner's (1970) ATSPPH. Both the original and short-form version of the scale are commonly used measures of help-seeking attitudes (Elhai et al., 2008). The shortened version was developed using 14 of the original 29 items with the highest item-total score correlations (Fischer & Farina, 1995). A factor analysis of the ATSPPH-SF was conducted with 389 college students and revealed a two-factor solution. The authors reported that Factor 1 contained 10 items reflecting a willingness to seek psychological help. Factor 2 contained the remaining four items that reflected a disclosure/interpersonal openness dimension. Due to low internal consistency of the second factor (Cronbach's α = .64), the items were subsequently dropped and the 10 items from Factor 1 were retained to comprise the ATSPPH-SF.

The development of the ATSPPH-SF addressed some of the limitations of the original scale, changed and updated some of the terminology, discarded items with low validity and reliability, and constructed a revised scale with a single unitary dimension. However, some questions regarding the factor structure of the ATSPPH-SF have emerged. For example, Elhai et al. (2008) investigated the factor structure of the ATSPPH-SF with a sample of 296 college students, primarily European Americans. Exploratory factor analysis (EFA) showed evidence for

two factors named Openness to Seeking Treatment for Emotional Problems and Value and Need in Seeking Treatment. A confirmatory factor analysis (CFA) validated the scale's two-factor solution with a large sample of medical patients, most whom were European American.

Some research has examined the factor structure of the ATSPPH-SF with non-European American samples and revealed some inconsistencies. Ang et al. (2007) investigated the factorial structure of the ATSPPH-SF across two Asian samples from Singapore. Using CFA, they found a poor fit with both data samples and the original one-factor solution proposed by Fischer and Farina (1995). Further investigation of each item revealed that the wording of Item #7 ("A person with an emotional problem is not likely to solve it alone; they are likely to solve it with professional help") was double-barreled and was dropped from the model. CFA's performed to test the goodness of fit of the one-factor model with nine items indicated an adequate fit among both Asian samples. Overall, findings from the Ang et al. (2007) study provided strong support for a unidimensional factorial model for the nine-item scale, rather the original 10 items.

However, the single-factor structure of the ATSPPH-SF was found to be a poor fit to the data among a sample of mainland Chinese university students (Fang et al., 2011). Instead, a two-factor model was determined after removing three double-barreled items. The internal consistency of the two-factor structured fell into the low range (below .60) leading the authors to conclude that there was no evidence for the cross-cultural equivalence of the ATSPPH-SF for mainland Chinese individuals. On the other hand, a separate study with Chinese college students found evidence for the two-factor structure of the ATSPPH-SF with acceptable, although low, internal consistency (Chang, 2007). More recent work examined the single- and two-factor models of the ATSPPH-SF among a large sample of individuals from Singapore (Picco et al., 2016). The major findings suggested that a three-factor structure showed the best fit of the data, while poor fit was reported for both the single- (nine item) and two-factor models.

Criticisms of the psychometric properties of the shortened scale including stability of its factor structure and its cross-cultural validity have emerged (Mackenzie et al., 2004). That is, the empirical findings have been inconsistent regarding the factor structure of the ATSPPH-SF when used across cultural groups. Still the ATSPPH-SF has been used widely in the United States with diverse cultural groups and internationally resulting in the translation of the instrument into numerous different languages (Fang et al., 2011). The bulk of ATSPPH-SF psychometric studies have sampled university students with cultural equivalence focusing on Asian groups. To our knowledge, measurement equivalence of the ATSPPH-SF with Latino populations has yet to be reported. Still this measure has been used extensively across Latino groups including university students (Gloria et al., 2010; Miville & Constantine, 2006; Ramos-Sanchez & Atkinson, 2009), Latino men from the community (Davis & Liang, 2015), and older adults (Hartlaub et al., 2014; Jang et al., 2011).

Demographic Differences

As noted, inconsistencies regarding the factor structure of the ATSPPH-SF across ethnically diverse samples suggest the need to better understand the measurement equivalence of this measure, particularly among Latino groups. As such, it is important to assess the equivalence of the ATSPPH-SF across key demographic variables that are thought to influence help-seeking attitudes within the Latino community, namely gender, nativity status, and language preferences.

Gender

Empirical research examining the relationship between gender and attitudes toward help seeking have noted that women tend to report more positive views than men (Gonzalez et al., 2005). This gender difference was found in a meta-analysis utilizing studies with college student samples (Nam et al., 2010). Similarly, the vast majority of Latina women interviewed believed that services would be helpful, whereas men were more likely to

have negative attitudes about the effectiveness of treatment (Bridges et al., 2012). This difference in help-seeking attitudes may help explain the findings that Latina women are three times more likely to seek mental health providers when compared with Latino men (Vega et al., 2001).

Nativity Status

In terms of help seeking, some research has shown that foreign-born Latinos utilize mental health and general medical care at rates lower than their U.S.-born counterparts (Alegría et al., 2007; Vega et al., 2001). Latinos who have lived in the United States for fewer than 5 years have reported less satisfaction with mental health treatment than those in residence for 20 years or greater (Alegría et al., 2007). Additionally, negative attitudes toward treatment have been thought to influence the ability to receive timely care, particularly for Latino immigrant men (Cabassa, 2007).

Language Preference

Empirical research has demonstrated that, among Latino groups, Spanish language preference has been associated with a lower likelihood of utilizing psychological services, particularly targeting anxiety and mood difficulties (Alegría et al., 2007; Berdahl & Stone, 2009; Keyes et al., 2012). Among foreign-born Latinos with a diagnosed disorder, limited English proficiency was related to decreased mental health service use (Kim et al., 2011). Although language preference has been thought to be correlated to an individual's attitudes toward help seeking and beliefs about health, minimal empirical work has sought to examine this link among Latino groups. Given its widespread use, the ATSPPH has been translated into several languages for study across diverse populations (Fang et al., 2011). Still the language equivalence of the ATSSPH-SF has yet to be examined for Spanish-speaking participants.

Purpose of the Current Study

The main objectives of the current study were to examine the factor structure of the ATSPPH-SF scale with a sample of Latino community adults. The two-factor model was tested as well as alternative factor structures. The measurement equivalence of the ATSPPH-SF was also examined by testing for DIF across gender, nativity status, and language format of survey. Because we were most interested in understanding whether there are item-specific characteristics (e.g., vocabulary, phrasing) that may result in DIF, we adopt an item response theory (IRT) approach in testing for DIF rather than the common structural equation modeling (SEM) approach of testing for measurement nonequivalence; however, both methods provide similar information (Thissen, 2017). Thissen (2017) illustrated that for ordinal items with seven plus or minus two response categories, DIF analysis from an IRT framework and multiple groups analysis from an SEM framework yield similar results. The main difference between the two approaches, as Thissen describes, is that IRT DIF analyses focus first on the individual item and then on the parameter type (i.e., discrimination, threshold), whereas SEM multiple group analyses focus first on the parameter type (i.e., factor loadings for all items simultaneously) and then on the item.

Which approach is used largely depends on the research question of interest. From a test theory perspective, it is typically the item that is of interest first, followed by the type of parameter. Some scholars have argued that item-level DIF analysis is necessary in establishing measurement equivalence, particularly with translated instruments, as these strategies provide information that multiple group CFAs at the scale level may overlook (Zumbo, 2003). Meade and Lautenschlager (2004) have also argued that when the equivalence of specific scale items is of primary interest, then the IRT approach may be more desirable. While both approaches have strengths and weaknesses, we adopt the IRT DIF approach in the current application; some of its limitations are addressed in the Discussion section.

Differential Item Functioning

For valid interpretation, test scores should have the same meaning across all groups responding to the questionnaire. An item is said to exhibit DIF when two individuals who are administered the same measure and have equal amounts of the construct that is being measured do not have the same probability of endorsing each response category of the item (Reise & Waller, 2009). In DIF analysis, it is common to consider two individuals from two different groups (e.g., gender, race or ethnicity, etc.). Evaluation of DIF is particularly important when test scores are used to make claims regarding group differences on the construct, as scores computed with DIF-contaminated items may obscure or create misleading differences between groups (Holland & Wainer, 1993). Thus, it is important to ask as follows: Do mean differences reflect real group differences in the latent variable that is being measured (e.g., Openness to Seeking Treatment), or are they merely an artifact of certain items on the measure having different psychometric properties between the two groups?

Within an IRT framework, DIF is often classified as one of two types: uniform or nonuniform. An item exhibits uniform (or threshold) DIF when after controlling for levels of the latent variable, one group is consistently more likely to endorse the item than the other group. Uniform DIF is shown graphically with item response functions that are shifted horizontally for one group relative to the other. In contrast, an item exhibits nonuniform (or discrimination) DIF when it is more salient for one group than another; thus, the relationship between the item and the construct differs between groups (Orlando & Marshall, 2002). Nonuniform DIF tends to manifest as the item being more easily endorsed by one group only at certain levels of the latent variable. Nonuniform DIF is shown graphically with item response functions that cross, as the item slopes are unequal.

Method

Participants

The current study included 437 adult individuals who self-identified as Latina/o, Hispanic, or of Spanish origin (i.e., Mexican, Puerto Rican, South American, Cuban, or other). The sample averaged approximately 43 years of age (SD = 16.59, range = 18-84) and included mostly women (67%, n = 287). The vast majority of participants identified as Mexican/Mexican American or Chicano (90%, n = 392). For nativity status, most individuals indicated having been born in the United States (60%, n = 259). In terms of annual household income, approximately 30% of the sample (n = 127) reported earning less than $20,000, 45% ($n$ = 199) indicated earning between $20,000 and $50,000, while about 21% (n = 92) stated over $75,000.

Measures

The ATSPPH-SF (Fischer & Farina, 1995) is a 10-item measure used to assess attitudes toward seeking psychological help. This unidimensional scale was developed from the original 29-item multidimensional measure (Fischer & Turner, 1970). Participants are asked to rate each statement on a 4-point Likert-type scale (0 = *disagree*, 1 = *somewhat disagree*, 2 = *somewhat agree*, 3 = *agree*). Total obtained score can range from 0 to 30, with higher scores indicating more positive attitudes toward treatment. The proportion of missing data at the item level ranged from less than 0.01 to 0.04.

Procedure

Participants were recruited, as part of a larger study examining Latino mental health, from local community/church fairs or neighborhood events celebrating Latino culture in a moderately sized Midwestern city. Interested participants were informed about the nature of the study, which involved answering questions about their daily lives as Latinos living in the United States, confidentiality, and anonymity. Participants who gave verbal consent were asked to complete a series of questionnaires that included the ATSPPH-SF. On completion, participants were compensated with a $10 gift card. All procedures were approved by the host institution's internal review board.

Questionnaires were made available in either English (n = 302) or Spanish (n = 135). The Spanish ATSPPH-SF was created for the current study using a double translation procedure; the gold standard in developing equivalent versions of instruments (Knight et al., 2009; Marín & Marín, 1991). A bilingual professional, not related to the project, first translated the English ATSPPH-SF into Spanish. Then, the Spanish version was back-translated into English by bilingual, bicultural research assistants. The back-translated and original English versions were then checked for accuracy and found to be similar in content. Translation of measures is a necessary step for inclusion of representative samples in research and it has been argued that including only English-speaking Latinos limits research to a select group of Latinos (Knight et al., 2009). Furthermore, scholars contend that it is reasonable to assess measurement equivalence when participants self-select into language administration groups so long as the construct being evaluated is not directly influenced by language fluency or cultural practices (Schwartz et al., 2014), neither of which apply to the ATSPPH-SF. Analysis of internal consistency for the current study of the ATSPPH-SF revealed a Cronbach's α of .70 for English and .70 for Spanish.

Analytic Plan

Scale Dimensionality

To evaluate the dimensionality of the measure, unidimensional and multidimensional graded response models (GRMs) were fit to the data using IRT patient-reported outcomes (IRTPRO; Cai et al., 2011). Two- and three-factor EFAs were also conducted. The GRM is a type of item response model designed for the analysis of ordered categorical responses (e.g., Never, Rarely, Sometimes, Often, Always). For a unidimensional GRM, each item is characterized by a discrimination parameter (a_j) that describes the strength of the relationship between item j and the latent variable θ, as well as a set of category-specific threshold parameters (b_{jk}). The larger the discrimination parameter a_j, the stronger the relationship between the item j and the latent variable θ. Threshold parameter b_{jk} describes the location on θ where, for item j, a respondent has a 0.5 probability of endorsing category k or higher. For an item with K response categories, a total of $K − 1$ threshold parameters are estimated (the probability of responding in the first category or higher is 1).

The multidimensional model found in the ATSPPH literature comprises two latent variables, with five items loading on each latent variable. We also considered a unidimensional model in which all 10 items load on a single latent variable, as well as two independent unidimensional models with five items each. For model identification, the mean and variance of each latent variable were fixed to 0 and 1, respectively. For multidimensional models, the correlation between the latent variables was estimated. Models were fit to the full sample, as well as to each subgroup: those taking the questionnaire in English, those taking the questionnaire in Spanish, U.S.-born respondents, foreign-born respondents, men, and women. To evaluate absolute model fit, models were first estimated using weighted least square mean and variance adjusted (WLSMV) as implemented in M*plus* version 8, which accounts for the ordinal nature of the data. After obtaining model fit information, all model parameters were estimated using maximum likelihood in IRTPRO (Cai et al., 2011). For each model, the correlation between the latent variables (if relevant), model fit, and item parameter estimates (discriminations and thresholds) were evaluated. We considered multiple goodness-of-fit statistics, including the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the standardized root mean square residual (SRMR). General guidelines exist for "poor," "acceptable," and "good" levels of model fit (e.g., Browne & Cudeck, 1993; Hu & Bentler, 1999); however, these guidelines were developed and tested with maximum likelihood under the assumption of multivariate normality, which deviates from the current study. Zhao (2014) conducted a large-scale simulation study examining the performance of these fit statistics with ordinal data and the WLSMV estimator, and cautioned against using strict cutoff values because so much variability exists among models (e.g., sample size, model complexity, etc.). For these reasons, rather than using stringent cutoffs, we evaluated model fit with the general goal of obtaining RMSEA values less than 0.08 (Browne & Cudeck, 1993), CFI and TLI values greater than 0.95 (Hu & Bentler,

1999), and SRMR values less than 0.08 (Hu & Bentler, 1999). We identified a single best-fitting model to be used for the DIF analyses.

Differential Item Functioning

After determining scale dimensionality, each item was tested for language-based (English vs. Spanish), nativity-based (U.S.-born vs. foreign-born), and gender-based (men vs. women) DIF, using all other items on the respective subscale as anchors that are presumed to be DIF-free. For example, in testing the first item on a five-item subscale for DIF, the other four items were used as anchors. For each candidate item, Wald χ^2 tests were conducted to test for both uniform and nonuniform DIF. To control the false discovery rate, the Benjamini and Hochberg (1995) procedure was used to adjust p values. This procedure is commonly used when examining several items for DIF on a single measure.

An item that exhibits statistically significant DIF does not necessarily indicate practical or clinical significance; thus, we used graphical and statistical methods to evaluate the magnitude of the DIF. Expected score plots, which are a function of the item discrimination and threshold parameters, trace a respondent's expected response (e.g., strongly disagree) as a function of the latent variable. For an item without DIF, the expected score plots should not differ between groups: Two individuals having the same level of the construct are expected to endorse the same response category, even if the individuals belong to two different groups. If an item exhibits uniform or nonuniform DIF, the expected score plots are no longer equivalent between groups. Larger discrepancies in the expected score plots suggest greater impact of DIF. The noncompensatory DIF (NCDIF) index, which is the average squared difference in the expected item scores between groups, was used to quantify the discrepancy between groups in their expected responses. Items with an NCDIF > 0.006 may be considered as showing practically significant DIF (Raju et al., 1995). Items that were determined to have practically significant DIF were removed from the measure. A final set of item parameters was estimated for the full sample using the revised measure. Item and test information functions were also examined to compare the relative performance of each item and measure.

Group Differences

To examine group differences on the latent variables after correcting for DIF, item parameters were fixed to those from the full sample calibration and IRT-based scores (i.e., response pattern expected a posteriori scores) were calculated for each subscale, as observed scores are commonly used in practice. Independent samples t tests were conducted on the IRT-based scores to determine whether there are significant differences in scores between (a) those who took the measure in English and Spanish, (b) U.S.-born and foreign-born respondents, and (c) men and women. Confidence intervals (CIs) and Cohen's d were also calculated for the mean differences. All analyses of group differences were done in R.

Results

Scale Dimensionality

To establish the dimensionality of the scale, we tested several models, including a one-factor IRT model, two-factor IRT models with correlated and uncorrelated factors, and two- and three-factor EFAs. Model fit statistics of all tested models for the full sample can be found in Table 1. The one-factor IRT model exhibits very poor fit, suggesting that at least two factors are necessary. While the three-factor EFA exhibits the best fit, interpretability is lost, as no items load significantly on the third factor. The IRT model with two correlated latent variables showed marginal to adequate fit in the full sample (RMSEA = 0.095, CFI = 0.94, TLI = 0.92, SRMR = 0.06), as well as in each of the six subgroups (RMSEA = 0.06-0.13, CFI = 0.90-0.97, TLI = 0.87-0.96, SRMR = 0.049-0.094); discrimination parameter estimates for the full sample are shown in Table 2. The correlation between the two latent variables was not significantly different from zero for any of the six subgroups (r = −0.14-0.04);

furthermore, estimating a two-factor model in which the correlation between the two factors is fixed to zero results in the improvement of most model fit statistics (RMSEA = 0.072, CFI = 0.96, TLI = 0.95, SRMR = 0.06). This finding suggests that the multidimensional scale really comprises two unidimensional scales (Liu et al., 2018). Thus, we proceeded to treat the measure as two unidimensional measures, entitled "Openness to Seeking Treatment" and "Value and Need in Seeking Treatment."

Table 1. Fit Statistics and Factor Correlations for All Tested Models.

Model	RMSEA [90% CI]	CFI	TLI	SRMR	Factor correlation(s) (*SE*)
One factor	0.258 [0.245, 0.272]	0.510	0.370	0.143	—
Two correlated factors	0.095 [0.081, 0.110]	0.936	0.915	0.062	0.072 (0.059)
Two uncorrelated factors	0.072 [0.057, 0.087]	0.962	0.952	0.064	—
Two-factor EFA	0.065 [0.048, 0.083]	0.977	0.960	0.040	−0.081 (0.070)
Three-factor EFA	0.047 [0.023, 0.070]	0.992	0.979	0.027	−0.085 (0.120)
					0.231 (0.123)
					0.139 (0.150)
Openness (unidimensional)	0.082 [0.046, 0.122]	0.991	0.983	0.019	—
Value (unidimensional)	0.085 [0.049, 0.125]	0.977	0.955	0.028	—

Note. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardized root mean square residual; SE = standard error; EFA = exploratory factors analysis.

Table 2. Discrimination Parameter Estimates (SEs) From the Multidimensional IRT Model When Fit to the Full Sample.

Item	Openness to Seeking Treatment	Value and Need in Seeking Treatment
If I believe I was having a mental breakdown, my first inclination would be to get professional attention.	2.12 (0.23)	—
If I were experiencing a serious emotional crisis at this point in my life, I would be confidence that I could find relief in psychotherapy.	2.80 (0.31)	—
I would want to get psychological help if I were worried or upset for a long period of time.	3.37 (0.43)	—
I might want to have psychological counseling in the future.	1.37 (0.15)	—
A person with an emotional problem is not likely to solve it alone; he or she is likely to solve it with professional help.	1.44 (0.15)	—
The idea of talking about problems with a psychologist strikes me as a poor way to get rid of emotional conflict.	—	1.36 (0.18)
There is something admirable in the attitude of a person who is willing to cope with his or her conflicts and fears without resorting to professional help.	—	1.03 (0.14)
Considering the time and expense involved in psychotherapy, it would have doubtful value for a person like me.	—	1.46 (0.18)
A person should work out their own problems; getting psychological counseling would be a last resort.	—	1.92 (0.25)
Personal and emotional troubles, like many things, tend to work out by themselves.	—	1.74 (0.23)

| | | r = −.05 (0.07) | |

Note. SE = standard error; IRT = item response theory.

Openness to Seeking Treatment

The unidimensional Openness to Seeking Treatment scale exhibited good fit in the full sample (RMSEA = 0.08, CFI = 0.99, TLI = 0.98). Item parameter estimates for this subscale are shown in Table 3. Discrimination parameters range from 1.37 to 3.37, with "I would want to get psychological help if I were worried or upset for a long period of time (Item #5)" being the most discriminating.

Table 3. Item Discrimination (*a*) and Threshold (*b*) Parameter Estimates From the Two Unidimensional Models When Fit to the Full Sample.

	a (SE)	b1 (SE)	b2 (SE)	b3 (SE)
Openness to Seeking Treatment				
If I believe I was having a mental breakdown, my first inclination would be to get professional attention.	2.12 (0.23)	−1.31 (0.11)	−0.76 (0.09)	−0.02 (0.08)
If I were experiencing a serious emotional crisis at this point in my life, I would be confidence that I could find relief in psychotherapy.	2.80 (0.31)	−1.39 (0.11)	0.67 (0.08)	0.33 (0.07)
I would want to get psychological help if I were worried or upset for a long period of time.	3.37 (0.43)	−1.38 (0.10)	−0.76 (0.07)	0.23 (0.07)
I might want to have psychological counseling in the future.	1.37 (0.15)	−1.19 (0.14)	−0.30 (0.10)	1.11 (0.14)
A person with an emotional problem is not likely to solve it alone; they are likely to solve it with professional help.	1.44 (0.15)	−1.64 (0.16)	−0.51 (0.10)	0.75 (0.12)
Value and Need in Seeking Treatment				
The idea of talking about problems with a psychologist strikes me as a poor way to get rid of emotional conflict.	1.36 (0.18)	−0.21 (0.09)	0.75 (0.11)	1.94 (0.21)
There is something admirable in the attitude of a person who is willing to cope with their conflicts and fears without resorting to professional help.	1.02 (0.14)	−1.42 (0.20)	−0.13 (0.11)	1.38 (0.19)
Considering the time and expense involved in psychotherapy, it would have doubtful value for a person like me.	1.44 (0.18)	−0.85 (0.12)	0.28 (0.09)	1.67 (0.18)
A person should work out their own problems; getting psychological counseling would be a last resort.	1.92 (0.25)	−0.58 (0.09)	0.30 (0.08)	1.33 (0.14)
Personal and emotional troubles, like many things, tend to work out by themselves.	1.76 (0.22)	−0.52 (0.09)	0.39 (0.09)	1.78 (0.17)

Differential Item Functioning

Results of the DIF analysis for the Openness to Seeking Treatment subscale are shown Table 4. All *p* values have been adjusted for multiple comparisons according to the Benjamini–Hochberg procedure. The item "A person with an emotional problem is not likely to solve it alone; they are likely to solve it with professional help (Item

#7)" exhibited significant nonuniform DIF, χ^2 (1) = 16.1, p < .001, such that the item is much more effective at measuring individual differences for those who took the questionnaire in English (a = 1.90) rather than Spanish (a = 0.69). The upper panel of Figure 1 shows the expected response as a function of Openness to Seeking Treatment. English speakers are represented with the solid black line; Spanish speakers are represented with the dashed blue line. Levels of the latent variable along the x-axis (θ) can be interpreted as standard scores, where θ = 0 indicates someone who is average on the construct. For both groups, those with higher levels of Openness to Seeking Treatment are more likely to express agreement with the item; however, the nonuniform nature of the DIF is shown with (a) the overall flatter slope for the Spanish-speaking group and (b) the intersecting slopes. For those who are below (approximately) average on Openness to Seeking Treatment (θ < −0.5), people taking the questionnaire in Spanish are more likely to endorse agreement with the item than those taking the questionnaire in English; for those who are at or above average, the opposite is true. The discrimination parameter is also considerably weaker for those who took the questionnaire in Spanish, reflected in the flatter slope. This suggests that the item is not as useful for measuring individual differences in those completing the questionnaire in Spanish. The NCDIF index for this item is 0.20, which is well above the practical significance cutoff value of 0.006 that is suggested by Raju et al. (1995).

Table 4. Discrimination (Nonuniform) and Threshold (Uniform) Differential Item Functioning Statistics for the Openness to Seeking Treatment Scale.

	English vs. Spanish				U.S.-born vs. foreign-born				Men vs. women			
	Discrimination		Threshold		Discrimination		Threshold		Discrimination		Threshold	
	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p
If I believe I was having a mental breakdown, my first inclination would be to get professional attention.	2.0 (1)	.16	3.6 (3)	.31	4.3 (1)	.04	4.4 (3)	.23	1.6 (1)	.21	4.1 (3)	.25
If I were experiencing a serious emotional crisis at this point in my life, I would be confidence that I could find relief in psychotherapy.	0.0 (1)	.86	2.1 (3)	.56	2.5 (1)	.12	1.6 (3)	.66	1.5 (1)	.23	1.2 (3)	.76
I would want to get psychological help if I were worried or upset for a long period of time.	0.0 (1)	.84	1.8 (3)	.62	0.2 (1)	.65	0.9 (3)	.84	0.0 (1)	.88	1.0 (3)	.79
I might want to have	0.3 (1)	.56	7.9 (3)	.05	0.9 (1)	.34	3.5 (3)	.33	0.0 (1)	1.0	7.8 (3)	.05

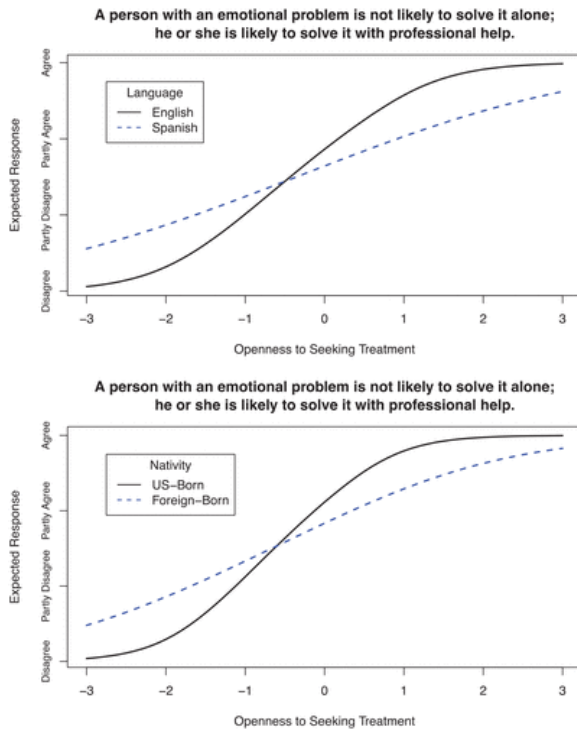|												
psychological counseling in the future.												
A person with an emotional problem is not likely to solve it alone; he or she is likely to solve it with professional help.	*16.1 (1)	<.01	5.8 (3)	.12	*13.8 (1)	<.01	2.5 (3)	.48	0.0 (1)	.86	1.8 (3)	.61

Figure 1. Upper panel: Expected score plot (item response function) for an item with language-based (upper panel) and nativity-based (lower panel) DIF.
Note. DIF = differential item functioning.

This item also exhibited significant nativity-based DIF in its discrimination parameter, χ^2 (1) = 13.8, $p < .001$, such that the item is better able to measure individual differences for those born in the United States (a = 2.24) than those who are foreign-born (a = 0.91). The magnitude of the DIF is more easily observed in the lower panel of Figure 1; US-born respondents are shown with the solid black line, and foreign-born respondents are shown with the dashed blue line. Similar to what was observed when comparing the item between languages, the relationship between Openness to Seeking Treatment and the expected response is noticeably weaker for foreign-born individuals that U.S.-born individuals. Furthermore, for those who are below average on Openness to Seeking Treatment, foreign-born individuals are more likely to endorse agreement with the item than U.S.-born individuals; for those who are at or above average, the reverse holds. The magnitude of the DIF for this item is also well above the suggested cutoff for practical significance, with an NCDIF of 0.16. No items on the Openness to Seeking Treatment scale exhibited significant gender-based DIF (Table 4). This suggests that the discrimination and threshold parameters can be considered equivalent for men and women. None of the items on the Openness scale exhibited significant uniform DIF.

Value and Need in Seeking Treatment

Differential Item Functioning
None of the items on the Value and Need in Seeking Treatment scale exhibited significant uniform of nonuniform DIF in comparing across language, nativity, and gender (Table 5). Thus, the item parameters and response functions can be viewed as equivalent across all subgroups and no revisions are necessary.

Table 5. Discrimination (Nonuniform) and Threshold (Uniform) Differential Item Functioning Statistics for the Value and Need in Seeking Treatment Scale.

| | English vs. Spanish | | | | U.S.-born vs. Foreign-born | | | | Men vs. Women | | | |
| | Discrimination | | Threshold | | Discrimination | | Threshold | | Discrimination | | Threshold | |
	χ2 (df)	p	χ2 (df)	p	χ2 (df)	p	χ2 (df)	p	χ2 (df)	p	χ2 (df)	p
The idea of talking about problems with a psychologist strikes me as a poor way to get rid of emotional conflicts.	0.8 (1)	.38	4.4 (3)	.23	1.1 (1)	.30	3.4 (3)	.34	0.1 (1)	.80	4.0 (3)	.26
There is something admirable in the attitude of a person who is willing to cope with his or her conflicts and fears without resorting to professional help.	1.4 (1)	.24	2.2 (3)	.54	5.0 (1)	.03	0.4 (3)	.94	0.9 (1)	.33	0.5 (3)	.93
Considering the time and expense involved in psychotherapy, it would have doubtful value for a person like me.	0.0 (1)	.92	4.2 (3)	.24	0.0 (1)	.99	4.5 (3)	.22	0.8 (1)	.36	4.0 (3)	.26
A person should work out his or her own problems; getting psychological counseling would be a last resort.	0.4 (1)	.54	0.7 (3)	.88	1.5 (1)	.22	0.9 (3)	.82	0.1 (1)	.75	0.6 (3)	.90
Personal and emotional troubles, like many things, tend to work out by themselves.	1.9 (1)	.17	3.6 (3)	.31	1.9 (1)	.16	4.8 (3)	.19	0.4 (1)	0.52	1.0 (3)	.80

Note. df = degrees of freedom.

Scale Revisions

The results of the DIF analyses suggest that the item "A person with an emotional problem is not likely to solve it alone; he or she is likely to solve it with professional help (Item #7)" does not perform similarly across language and nativity subgroups. When an item exhibits DIF, the researcher must decide whether and how to account for the DIF. If the DIF is small in effect size, one option is to retain the item and ignore the DIF because the effect is negligible (i.e., scores computed with and without correcting for DIF are the same; Orlando & Marshall, 2002). If the magnitude of the DIF is more substantial, a second option is to retain the item but allow the discrimination and/or threshold parameters for that item to vary across groups; this approach ensures that any group differences on the latent variable are real differences rather than an artifact of DIF (Orlando & Marshall, 2002). A third option is to remove the item from the scale (Teresi et al., 2007). We chose this last approach for three reasons. First, because of the overlap in group membership (e.g., a U.S.-born Spanish-speaking man), it is difficult to determine the appropriate DIF adjustment that takes into account all group memberships simultaneously. Second, the magnitude of the DIF is so large in this case that the quality of the item is questionable, particularly for individuals who are foreign-born or Spanish speakers. Third, across the entire sample as well as all subgroups, this item is less discriminating than most of the other items on the scale, possibly due to its double-barreled structure; thus, the scale does not lose much in psychometric value when this item is dropped. For subsequent analyses, the Openness to Seeking Treatment scale comprises four items, and the Value in Seeking Treatment scale retains all five original items.

The revised Openness to Seeking Treatment and Value in Seeking Treatment scales demonstrate good or excellent fit for the full sample (Openness to Seeking Treatment: RMSEA = 0.06, CFI = 1.00, TLI = 0.99; Value in Seeking Treatment: RMSEA = 0.08, CFI = 0.98, TLI = 0.96). Item parameter estimates for the revised scales are shown in Table 6. Figure 2 displays the item information curves for both scales. The Openness to Seeking Treatment scale provides the most information about individual differences for those who are within approximately 1 standard deviation of the mean, with "If I were experiencing a serious emotional crisis at this point in my life, I would be confident that I could find relief in psychotherapy (Item #3)" and "I would want to get psychological help if I were worried or upset for a long period of time (Item #5)" being the best-performing items. Notably, the scale does a very poor job of measuring individual differences at high ends of the latent variable continuum ($\theta > 2$). For the Value and Need in Seeking Treatment scale, the item providing the most information is "A person should work out his or her own problems; getting psychological counseling would be a last resort (Item #9)." The range of maximal information is wider for this scale, spanning from approximately 1 standard deviation below the mean to 2 standard deviations above the mean; however, the Openness to Seeking Treatment scale provides considerably more information on average, which can be observed in comparing the two solid black lines. Overall, the Value and Need in Seeking Treatment scale is quite weak in its ability to discriminate among individuals across the latent variable continuum.

Table 6. Item Discrimination (*a*) and Threshold (*b*) Parameter Estimates From the Two Revised Unidimensional Models When Fit to the Full Sample.

|	a (SE)	b_1 (SE)	b_2 (SE)	b_3 (SE)
Openness to Seeking Treatment				
If I believe I was having a mental breakdown, my first inclination would be to get professional attention.	2.18 (0.24)	−1.30 (0.11)	−0.76 (0.09)	−0.02 (0.08)
If I were experiencing a serious emotional crisis at this point in my life, I would be confidence that I could find relief in psychotherapy.	3.05 (0.37)	−1.36 (0.11)	−0.67 (0.08)	0.33 (0.07)
I would want to get psychological help if I were worried or upset for a long period of time.	3.03 (0.37)	−1.42 (0.11)	−0.79 (0.08)	0.23 (0.07)
I might want to have psychological counseling in the future.	1.31 (0.15)	−1.22 (0.14)	−0.30 (0.10)	1.13 (0.15)
Value and Need in Seeking Treatment				
The idea of talking about problems with a psychologist strikes me as a poor way to get rid of emotional conflict.	1.36 (0.18)	−0.21 (0.09)	0.75 (0.11)	1.94 (0.21)
There is something admirable in the attitude of a person who is willing to cope with his or her conflicts and fears without resorting to professional help.	1.02 (0.14)	−1.42 (0.20)	−0.13 (0.11)	1.38 (0.19)
Considering the time and expense involved in psychotherapy, it would have doubtful value for a person like me.	1.44 (0.18)	−0.85 (0.12)	0.28 (0.09)	1.67 (0.18)
A person should work out his or her own problems; getting psychological counseling would be a last resort.	1.92 (0.25)	−0.58 (0.09)	0.30 (0.08)	1.33 (0.14)
Personal and emotional troubles, like many things, tend to work out by themselves.	1.76 (0.22)	−0.52 (0.09)	0.39 (0.09)	1.78 (0.17)

Note. SE = standard error.
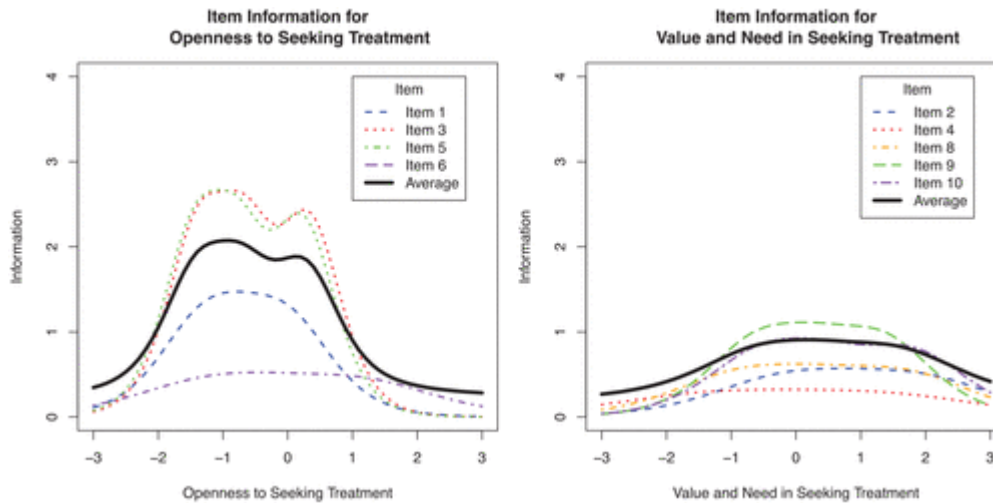
Figure 2. Item information curves for openness to seeking treatment and value and need in seeking treatment scales.

Note. The solid black lines represent the average item information for each scale.

Group Differences

The item parameters in Table 6 were used to estimate IRT scale scores for the entire sample for both subscales. Mean differences were then examined for each pair of groups. People taking the questionnaire in Spanish (*M* = 0.21, *SD* = 0.93) are significantly more open to seeking treatment than those taking the questionnaire in English (*M* = −0.09, *SD* = 0.87) by more than one third of a standard deviation, *t*(425) = 3.24, *p* = .001, 95% CI = [0.12, 0.48], *d* = 0.38. Those who are foreign-born (*M* = 0.16, *SD* = 0.95) are also significantly more open to seeking treatment that those who are U.S.-born (*M* = −0.10, *SD* = 0.85), though the magnitude of the difference is slightly smaller, *t*(422) = 2.94, *p* = .003, 95% CI = [0.02, 0.26], *d* = 0.29. Finally, women (*M* = 0.09, *SD* = 0.88) are significantly more open to seeking treatment than men (*M* = −0.17, *SD* = 0.91) by just under one third of a standard deviation, *t*(417) = 2.78, *p* = .006, 95% CI = [−0.03, 0.25], *d* = 0.29.

Individuals who took the questionnaire in Spanish (*M* = 0.14, *SD* = 0.89) see significantly more Value and Need in Seeking Treatment that those who took the questionnaire in English (*M* = −0.06, *SD* = 0.94) by approximately one quarter of a standard deviation, *t*(425) = 2.28, *p* = .023, 95% CI = [0.03, 0.38], *d* = 0.23. There was no significant difference between foreign-born (*M* = −0.06, *SD* = 0.83) and U.S.-born (*M* = 0.08, *SD* = 0.91) respondents in Value and Need in Seeking Treatment, *t*(422) = 1.59, *p* = .112. Last, men (*M* = 0.05, *SD* = 0.80) and women (*M* = −0.04, *SD* = 0.88) did not exhibit a significant difference in Value and Need in Seeking Treatment, *t*(417) = 0.91, *p* = .36.

Discussion

The current study sought to examine the factor structure and measurement equivalence of the ATSPPH-SF among a group of Latino participants. The main findings revealed that this measure's factor structure was best conceptualized as multiple unidimensional measures. That is, a two-factor structure in which the latent factors were uncorrelated provided the best fit for the data. This factor structure is in line with past work (Elhai et al., 2008), although the current results suggest that they are relatively independent factors, labeled Openness to Seeking Treatment and Value and Need in Seeking Treatment.

The main analyses revealed that the item "A person with an emotional problem is not likely to solve it alone; he or she is likely to solve it with professional help (Item #7)" did not perform similarly across language and nativity groups. In their validation study of the ATSPPH-SF with Asian samples, Ang et al. (2007) found that this item showed poor factor loadings partly because of its double-barreled wording. That is, participants could agree with one part of the item but disagree with the other. Spanish-speaking and foreign-born participants might be particularly conflicted about the two halves of this item or have more difficulty understanding which portion of the item to answer. After removing the problematic item, analyses revealed a good to excellent fit.

In reviewing the two-factor structure, it appears that Openness to Seeking Treatment is an effective measure of this construct especially for those who score in the middle ranges. That is, this subscale provides more precision and better discrimination in measuring a Latino individual's openness to seeking psychological services. However, these items may be less effective at differentiating those who score in the high range. Still these findings suggest that Openness to Seeking Treatment, with Item #7 removed, can be used as an adequate measure in future research with Latino populations. The Values and Need in Seeking Treatment subscale, on the other hand, was shown to be less able to differentiate individuals across ranges of the construct. As such, use of this subscale with Latino populations should be used cautiously.

After removing the DIF-contaminated item, participants who took the ATSPPH-SF in Spanish showed higher Openness to Seeking Treatment and Value and Need in Seeking Treatment than those who preferred English. Similarly, foreign-born participants reported more Openness to Seeking Treatment than their U.S.-born counterparts. These findings are in contrast with past research which has reported that Spanish-speaking and foreign-born Latinos are less likely to utilize psychological services and show less favorable attitudes regarding help seeking. Consistent with previous work, Latina women showed higher Openness to Seeking Treatment when compared with Latino men. No gender differences were observed for Value and Need in Seeking Treatment, thus, providing further evidence for considering these subscales as independent constructs.

It is important to note that the ATSPPH-SF assesses an individual's attitudes or opinions toward mental health services. Like other ethnic minority individuals, Latinos often experience a multitude of institutional barriers, including underinsurance and underemployment, to name a few, which may interfere with their ability to use services despite having more favorable opinions to doing so. Still the ATSPPH-SF has been criticized for its item content and ambiguous choice of language (Mackenzie et al., 2004). For example, the scale appears to use the terms "professional attention" and "psychological counseling" interchangeably. Confusion may arise given that "professional attention" could be interpreted as general medical services rather that psychological treatment. This notion could be particularly salient among Latinos who are more likely to visit a general practitioner than use specialized mental health services (Alegría et al., 2007; Vega et al., 2001). Given these and other difficulties, more recent instruments, including the Inventory of Attitudes Toward Seeking Mental Health Services (Mackenzie et al., 2004) and the Mental Help Seeking Attitudes Scale (Hammer et al., 2018) have been developed within a theoretical framework and have shown promising findings when compared with the ATSPPH-SF (Hammer et al., 2018; Wei et al., 2017). Still the psychometric properties of these measures have yet to be examined with the Latino population and the ATSPPH-SF continues to be the most used scale assessing help-seeking attitudes suggesting that further research is needed.

Limitations and Future Directions

We adopted the IRT DIF approach in evaluating the equivalence of the measure between groups. Because we were primarily interested in understanding whether certain item characteristics might take on different cultural meanings, we believe that the IRT DIF approach could most directly address our research questions. However, some limitations should be noted. First, the IRT DIF approach does not offer a formal test of configural invariance (i.e., whether the factor structure differs between groups). In the present case, as well as in most

applications of IRT DIF analyses, it is assumed that each subscale is unidimensional. Future research should consider evaluating measurement equivalence from an SEM multiple-group approach, which allows a formal test of configural invariance. Second, to the authors' knowledge, the equivalence of the IRT DIF and SEM multiple groups approaches has not been formally studied under conditions in which normal theory maximum likelihood is inappropriate—that is, when items have fewer than five response alternatives and should thus not be treated as continuous (Rhemtulla et al., 2012). Specifically, if one were to conduct an SEM multiple-group analyses with the current data, which have only four response categories, it would be more justifiable to use a limited information estimator (e.g., WLSMV) that uses estimates of the polychoric correlations rather than the raw data (Rhemtulla et al., 2012). The IRT DIF approach, in contrast, is a full information method that uses the raw data. Thus, the two methodological approaches may be more comparable under the conditions described in Thissen (2017), who justified the use of maximum likelihood within the SEM framework due to the large number of response categories, than the conditions described in the present study. Future methodological work should more rigorously examine the comparability of the IRT and SEM approaches when the SEM approach involves a limited information estimator.

Unfortunately, the current study did not compare the ATSPPH-SF with other measures of help seeking, thus limiting the ability to evaluate the construct validity of this instrument. Furthermore, the majority of the current sample comprised individuals of Mexican descent, thus, not allowing for comparisons across Latino subgroups. Future research should seek to assess the psychometric properties of the ATSPPH-SF within heterogeneous Latino samples and with other ethnic groups. Additionally, the present study did not assess past help-seeking behaviors, which is likely to influence the attitudes toward seeking psychological services. That is, a self-selection bias may have occurred with those who volunteered to participate. Finally, although participants completed the questionnaires in their preferred language, we did not include formal measures of language fluency making it difficult to determine overall language ability or differentiate monolingual from bilingual speakers. Presumably, participants chose the language in which they were the most fluent and comfortable with when completing paper–pencil surveys. Despite these constraints, the current findings provide insight into the utility of the ATSPPH-SF with Latino adults and suggests that further work is needed to understand the factors that contribute to an individual's attitudes toward seeking psychological services. As such, future research can investigate which ATSPPH-SF subscale may be most predictive of treatment utilization while taking into account critical demographic factors including gender, language preference, and nativity status.

Author's Note

Declaration of Conflicting Interests

Funding

References

Alegría, M., Mulvaney-Day, N., Woo, M., Torres, M., Gao, S., Oddo, V. (2007). Correlates of past-year mental health service use among Latinos: Results from the national Latino and Asian American study. American Journal of Public Health, 97(1), 76-83.

Ang, R. P., Lau, S., Tan, A.-G., Lim, K. M. (2007). Refining the Attitudes Toward Seeking Professional Psychological Help Scale: Factorial invariance across two Asian samples. Measurement and Evaluation in Counseling and Development, 40(3), 130-141.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B: Methodological, 57(1), 289-300.

Berdahl, T. A., Stone, R. A. T. (2009). Examining Latino differences in mental healthcare use: The roles of acculturation and attitudes towards healthcare. Community Mental Health Journal, 45(5), 393-403.

Bridges, A. J., Andrews, A. R., Deen, T. L. (2012). Mental health needs and service utilization by Hispanic immigrants residing in mid-southern United States. Journal of Transcultural Nursing, 23(4), 359-368.

Browne, M. W., Cudeck, R. (1993). Alternative ways of assessing model fit. Sage Focus Editions, 154, 136-136.

Burnett-Zeigler, I., Lee, Y., Bohnert, K. M. (2018). Ethnic identity, acculturation, and 12-month psychiatric service utilization among Black and Hispanic adults in the U.S. Journal of Behavioral Health Services & Research, 45(1), 13-30.

Cabassa, L. J. (2007). Latino immigrant men's perceptions of depression and attitudes toward help seeking. Hispanic Journal of Behavioral Sciences, 29(4), 492-509.

Cai, L., Thissen, D., du Toit, S. H. (2011). IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Scientific Software International. http://swmath.org/software/13414

Chang, H. (2007). Depressive symptom manifestation and help-seeking among Chinese College students in Taiwan. International Journal of Psychology, 42(3), 200-206. https://doi.org/10.1080/00207590600878665

Davis, J. M., Liang, C. T. (2015). A test of the mediating role of gender role conflict: Latino masculinities and help-seeking attitudes. Psychology of Men & Masculinity, 16(1), 23-32. https://doi.org/10.1037/a0035320

Elhai, J. D., Schweinle, W., Anderson, S. M. (2008). Reliability and validity of the Attitudes Toward Seeking Professional Psychological Help Scale-Short Form. Psychiatry Research, 159(3), 320-329.

Fang, K., Pieterse, A. L., Friedlander, M., Cao, J. (2011). Assessing the psychometric properties of the Attitudes Toward Seeking Professional Psychological Help Scale-Short Form in Mainland China. International Journal for the Advancement of Counseling, 33(4), 309-321.

Fischer, E. H., Farina, A. (1995). Attitudes toward seeking professional psychological help: A shortened form and considerations for research. Journal of College Student Development, 36(4), 368-373.

Fischer, E. H., Turner, J. L. (1970). Orientation to seeking professional psychological help: Development and research utility of an attitude scale. Journal of Consulting and Clinical Psychology, 35(1), 79-90.

Gloria, A. M., Castellanos, J., Segura-Herrera, T. A., Mayorga, M. (2010). Assessing cultural orientation, cultural fit, and help-seeking attitudes of Latina undergraduates. Journal of College Counseling, 13(2), 126-140.

Gonzalez, J. M., Alegría, M., Prihoda, T. J. (2005). How do attitudes toward mental health treatment vary by age, gender, and ethnicity/race in young adults? Journal of Community Psychology, 33(5), 611-629.

Gonzalez, J. M., Alegría, M., Prihoda, T. J., Copeland, L. A., Zeber, J. E. (2011). How the relationship of attitudes toward mental health treatment and service use differs by age, gender, ethnicity/race and education. Social Psychiatry and Psychiatric epidemiology, 46(1), 45-57.

Hammer, J. H., Parent, M. C., Spiker, D. A. (2018). Mental Help Seeking Attitudes Scale (MHSAS): Development, reliability, validity, and comparison with the ATSPPH-SF and IASMHS-PO. Journal of Counseling Psychology, 65, 74-85.

Hartlaub, M. G., Biedenharn, P., Brouillard, P., Seidel, S. (2014). Attitudes toward treatment and potential barriers to access of mental health services in a sample of elderly Hispanic and Anglo adults. Texas Public Health Journal, 66, 18-20.

Holland, P. W., Wainer, H. (Eds.). (1993). Differential item functioning. Lawrence Erlbaum.

Hu, L., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.

Jang, Y., Chiriboga, D. A., Herrera, J. R., Martinez Tyson, D., Schonfeld, L. (2011). Attitudes toward mental health services in Hispanic older adults: The role of misconceptions and personal beliefs. Community Mental Health Journal, 47, 164-170.

Keyes, K. M., Martins, S. S., Hatzenbuehler, M. L., Blanco, C., Bates, L. M., Hasin, D. S. (2012). Mental health service utilization for psychiatric disorders among Latinos living in the United States: The role of ethnic subgroup, ethnic identity, and language/social preferences. Social Psychiatry and Psychiatric Epidemiology, 47, 383-394.

Kim, G., Aguado Loi, C. X., Chiriboga, D. A., Jang, Y., Parmelee, P., Allen, R. S. (2011). Limited English proficiency as a barrier to mental health service use: A study of Latino and Asian immigrants with psychiatric disorders. Journal of Psychiatric Research, 45(1), 104-110.

Knight, G. P., Roosa, M. W., Umaña-Taylor, A. J. (2009). Studying ethnic minority and economically disadvantaged populations: Methodological challenges and best practices. American Psychological Association.

Liu, Y., Magnus, B. E., O'Connor, H., Thissen, D. (2018). Multidimensional item response theory. In Irwing, P., Booth, T., Hughes, D. (Eds.), The Wiley-Blackwell handbook of psychometric testing (pp. 445-493). John Wiley.

Mackenzie, C. S., Know, V. J., Gekoski, W. L., Macaulay, H. L. (2004). An adaptation and extension of the Attitudes Toward Seeking Professional Psychological Help Scale. Journal of Applied Social Psychology, 34, 2410-2435.

Marín, G., Marín, B. (1991). Research with Hispanic populations. Sage.

Meade, A. W., Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organizational Research Methods, 7(4), 361-388.

Miville, M. L., Constantine, M. G. (2006). Sociocultural predictors of psychological help-seeking attitudes and behavior among Mexican American college students. Cultural Diversity & Ethnic Minority Psychology, 12(3), 420-423.

Mojtabai, R., Olfson, M., Mechanic, D (2002). Perceived need and help-seeking in adults with mood, anxiety, or substance use disorders. Archives of General Psychiatry, 59(1), 77-84.

Nam, S. K., Chu, H. J., Lee, M. K., Lee, J. H., Kim, N., Lee, S. M. (2010). A meta-analysis of gender differences in attitudes toward seeking professional psychological help. Journal of American College Health, 59(2), 110-116.

Orlando, M., Marshall, G. N. (2002). Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. Psychological Assessment, 14(1), 50-59.

Picco, L., Abdin, E., Chong, S. A., Pang, S., Shafie, S., Chua, B. Y., Vaingankar, J. A., Ong, L. P., Tay, J., Subramaniam, M. (2016). Attitudes toward seeking professional psychological help: Factor structure and socio-demographic predictors. Frontiers in Psychology, 7, 547-557.

Raju, N. S., van der Linden, W. J., Fleer, P. F. (1995). IRT-based internal measures of differential item functioning. Applied Psychological Measurement, 19(4), 353-368.

Ramos-Sanchez, L., Atkinson, D. R. (2009). The relationship between Mexican American acculturation, cultural values, gender, and help-seeking intentions. Journal of Counseling & Development, 87(1), 62-71.

Reise, S. P., Waller, N. G. (2009). Item response theory and clinical measurement. Annual Review of Clinical Psychology, 5, 27-48.

Rhemtulla, M., Brosseau-Liard, P. É., Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological Methods, 17, 354-373.

Schwartz, S. J., Benet-Martínez, V., Knight, G. P., Unger, J. B., Zamboanga, B. L., Des Rosiers, S. E., Stephens, D., Huang, S., Szapocznik, J. (2014). Effects of language of assessment on the measurement of acculturation: Measurement equivalence and cultural frame switching. Psychological Assessment, 26(1), 100-114.

Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Cook, K. F., Crane, P. K., Gibbons, L. E., Morales, L. S., Orlando-Edelen, M., Cella, D. (2007). Evaluating measurement equivalence using item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): Applications (with illustrations) to measures of physical functioning ability and general distress. Quality of Life Research, 16(Suppl. 1), 43-68.

Thissen, D. (2017). Similar DIFs: Differential item functioning and factorial invariance for scales with seven ("plus or minus two") response alternatives. In van der Ark, L. A., Wilberg, D. M., Culpepper, S.A., Douglas, J. A., Wang, W-C. (Eds.), Quantitative psychology: The 81st Annual Meeting of the Psychometric Society, Asheville, North Carolina, 2016 (pp. 81-91). Springer.

U.S. Census Bureau . (2018). Hispanic heritage month 2018. https://www.census.gov/newsroom/facts-for-features/2018/hispanic-heritage-month.html

U.S. Department of Health and Human Services . (2001). Mental health: Culture, race, and ethnicity—A supplement to mental health: A report of the surgeon general. https://drum.lib.umd.edu/handle/1903/22834

Vega, W. A., Kolody, B., Aguilar-Gaxiola, S. (2001). Help seeking for mental health problems among Mexican Americans. Journal of Immigrant Health, 3(3), 133-140.

Wei, U., McGrath, P. J., Hayden, J., Kutcher, S. (2017). Measurement properties of mental health literacy tools measuring help-seeking: A systematic review. Journal of Mental Health, 26, 543-555.

Zhao, Y. (2014). The performance of model fit measures by robust weighted least squares estimators in confirmatory factor analysis [Unpublished doctoral dissertation]. The Pennsylvania State University.

Zumbo, B. D. (2003). Does item-level DIF manifests itself in scale-level analyses? Implications for translating language tests. Language Testing, 20(2), 136-147.