The Development of the DASS-6 for In-Person and Telehealth Settings: A Community-Driven Empirical Approach

Lissette M. Saavedra, Antonio A. Morgan-López, Kiersten L. Johnson, Shannon M. Blakey, Alexandra T. Tonigan, Luke Smith, and Molly B. Hayes

RTI INTERNATIONAL

Suggested Citation

https://doi.org/10.3768/rtipress.2025.mr.0055.2504   **www.rti.org/rti-press**

Contents

About the Authors

Lissette M. Saavedra, PhD, is a Public Health Fellow at RTI International. https://orcid.org/0000-0001-8880 -0624

Antonio A. Morgan-López, PhD, is Statistics and Psychometrics Senior Fellow at RTI International. https://orcid.org/0000-0003-4706-9964

Kiersten L. Johnson, PhD, is a research public health analyst at RTI International. https://orcid.org/0000 -0002-0266-2521

Shannon M. Blakey, PhD, is a research public health analyst at RTI International. https://orcid.org/0000 -0002-9291-3688

Alexandra T. Tonigan, PhD, is a research public health analyst at RTI International. https://orcid.org/0000 -0003-0501-9598

Luke Smith, MD, is the executive director and medical director of El Futuro, Durham, NC.

Molly B. Hayes, MEd, is the director of clinical enhancement services at El Futuro, Durham, NC.

RTI Press Associate Editor

Marcus Berzofsky

Abstract

Brief, reliable instruments are critical for mental health care providers to efficiently capture psychometrically sound and clinically useful information during in-person and telehealth service provision. This study used data from 949 community clinic patients with primary anxiety and depressive disorders. We developed a shortened (6-item) form of the English and Spanish versions of the 21-item Depression, Anxiety, and Stress Scales (DASS) for in-person, telehealth, and hybrid mental health services using moderated nonlinear factor analysis, designed to maintain maximum reliability across the clinical range of mental health treatment-seekers. Results indicate minimal loss in local reliability in reducing the DASS to 6 items (DASS-6), with reliability exceeding 0.85 throughout the relevant clinical range of distress. Moreover, likely clinical distress is a minimum total score of 8. Test-retest with an independent sample that included in person and telehealth administration was in the good-to-excellent range for each subscale. The DASS-6 is an efficient tool for assessing depression, anxiety, and stress in both in-person and telehealth settings.

Introduction

In the past two decades, considerable progress has been made in research around mental health service provision for underserved Latinx people in the United States—both immigrants and those born in the United States (Castro, Barrera, & Holleran Steiker, 2010). Despite improvements in mental health access, Latinx people continue to seek mental health services less often than other racial and ethnic groups (Chang & Biegel, 2018; Fortuna et al. 2016; Derr, 2016). Latinx people, particularly first-generation immigrants, face significant barriers to treatment utilization, including lack of health insurance, language challenges, fear of discrimination, concerns about citizenship status, and stigma (Fortuna et al., 2016). As such there is an emphasis in Latinx communities on increasing feasibility of identifying mental health problems in primary care and specialized mental health settings where people feel more comfortable disclosing information and more open to receiving treatment (Escovar, Craske, Roy-Byrne, Stein, Sullivan, Sherbourne, Bystritsky, & Chavira, 2018).

It is well documented that underserved groups of people experienced more risk during the COVID-19 pandemic (Bambra, Riordan, Ford, & Matthews, 2020). Social distancing and shelter-in-place ordinances enacted across the globe in response to the COVID-19 pandemic created a major hurdle in access to care. Many behavioral health providers have been quick to embrace telehealth solutions to continue caring for their clients. Telehealth service provision via telephone and video were the main mental health service modality available during the pandemic (Saavedra, Johnson, Counts, et al., 2024; Uscher-Pines, Sousa, Jones et al. 2021). Consequently, there has been a synchronized pivot to conducting treatment monitoring and other assessments via virtual delivery methods (video and audio-only) as well via in-person delivery. Post COVID-19 pandemic, this need is still present for people living in rural areas (Patel, Saavedra, Rodriguez-Borja, et al., 2025).

One of the most widely used instruments to assess self-reported depression, anxiety, and stress in both English and Spanish is the Depression, Anxiety, and Stress Scales (DASS; Lovibond & Lovibond,

1995; Lovibond, 1998; Gomez, 2016). The DASS is commonly used in primary care and community-based clinical settings, as well as research settings. The DASS is unique in that it allows for the assessment of depression, anxiety, and stress in one instrument. Moreover, the DASS is not tied to a particular edition of the *Diagnostic and Statistical Manual* (DSM; American Psychiatric Association, 2013). The original DASS, which consisted of 42 items, was shortened to the DASS-21, which has undergone considerable psychometric evaluation (see Clara, Cox, & Enns, 2001; Henry & Crawford, 2005; Daza, Novy, Stanley, & Averill, 2002; Norton, 2007).

The need for brief instruments that can quickly and accurately assess symptoms of depression, anxiety, and stress to inform clinical decision-making is critical in most settings but especially in primary care and community-based organizations that provide mental health treatment, including via telehealth. In such settings, providers tend to be generalists with limited time for evaluation and assessment, yet a large amount of information needs to be gathered to establish a clinical diagnosis. Lengthy assessment instruments used to compliment the clinical diagnosis contribute to patient fatigue, which can reduce the validity of the assessment data. Given that Federally Qualified Health Centers, community-based organizations, and primary care entities often work with Latinx individuals, these clinical settings would benefit from a validated short form of an instrument with strong psychometric properties that is available in both English and Spanish to meet their patients' language preferences (Saavedra et al., 2024; Sanchez et al., 2024). This is especially true in the context of telehealth service delivery.

The aim of this study was to examine the use of moderated nonlinear factor analysis (MNLFA; Bauer & Hussong, 2009; Bauer, 2017) to develop a shorter version of the DASS-21 using data from a community-based treatment clinic primarily serving Latinx patients presenting with depression, anxiety, and stress. The MNLFA model is a flexible psychometric model that, unlike item response theory (IRT) models (e.g., Edelen & Reeve, 2007; Samejima, 1969, 2016), accounts for differential item functioning (DIF) in both factor loadings and item

thresholds across multiple predictors simultaneously. Providers have emphasized a need for a shorter instrument for continued telehealth (telephone and video) service delivery during the pandemic but also after the pandemic where segments of the population will continue to have limited access to mental health services (Saavedra et al., 2024; Uscher-Pines, Sousa, Jones et al. 2021).

The DASS has been translated to more than 20 languages, but most of the psychometric work with translated versions of the DASS has been conducted using classical test theory (CTT) approaches to score the DASS even when factor analysis methods have been used to assess item properties (e.g., Daza et al., 2002; Ruiz et al., 2017). We used MNLFA in this study to (a) assess differences in the item properties across language, gender, age and primary diagnosis simultaneously; (b) estimate scale scores that account for differential weighting of symptoms (Bauer, 2017; Curran et al., 2008); and (c) use MNLFA item information functions to develop a short screener that maximally distinguish clinical from nonclinical levels of negative affectivity (NA) among Hispanic mental health treatment-seekers. Further, this analysis attempts to balance (a) the incorporation of at least one depression, anxiety, and stress item that was found to be invariant across DIF factors and/or (b) one depression, anxiety, and stress item that had maximum item information at clinical levels of NA (even if the item was noninvariant across language and/or other DIF factors; see Millsap & Kwok, 2004 discussion of partial invariance). Such a short-form DASS instrument could be easily integrated into community clinical settings to improve treatment planning and monitoring for English- and Spanish-speaking patients.

Methods

Participants

The sample consisted of 949 individuals (64 percent female; mean age = 32.65 [12.74]) who presented to a community-based mental health treatment center that provides bilingual and culturally informed behavioral health treatment for underserved Spanish-speaking individuals and families. This organization serves more than 2,400 patients per year, the majority of whom are first- and second-generation immigrants. Most of the patient population falls below 200 percent of the federal poverty level. All assessors and therapists are fully bilingual in Spanish and English, and all instruments are available in both languages. Patients are given the option to complete assessment and treatment in their preferred language (Spanish or English). The total sample of 949 was divided into a calibration sample and a validation sample using a random 50/50 split based on the SAS random (binary) number generator function RANBIN. Table 1 presents demographic and diagnostic information, subset by language and sample type (calibration or validation). In a similar albeit independent subsample (n = 161), we also examined test-retest reliability of responses from the 6 items from the fully administered DASS-21 and the items from the shortened version (DASS-6).

Measures

Depression, Anxiety, and Stress Scale-21

The DASS-21 is used to assess past-week depression, anxiety, and stress along three 7-item self-report scales (Lovibond & Lovibond, 1995). Items are rated using a 4-point Likert-type scale (0 = "did not apply to me at all" to 3 = "applied to me very much or most of the time"). The English version of the DASS-21 has good psychometric properties (e.g., internal consistency, convergent and discriminant validity) based on CTT methods (McDonald, 1999; Scholten et al., 2017), as does the Spanish translation of the DASS-21 (Bados et al., 2005). The properties using IRT methods (e.g., local reliability; Chiesi et al., 2017), however, are not yet known.

Psychiatric Diagnosis

Primary diagnoses of anxiety, depression, or posttraumatic stress disorder (PTSD) were based on a diagnostic interview with a licensed service provider (psychiatrist, psychologist, counselor). Although this interview is not a standard semi-structured as in clinical research, clinicians have access to modules from semi-structured interviews. Table 1 presents DSM-5 diagnoses for the sample.

Table 1. Demographics and primary diagnosis by language and sample type

Demographic or diagnostic category	Calibration sample English ($n = 150$)	Calibration sample Spanish ($n = 347$)	Validation sample English ($n = 143$)	Validation sample Spanish ($n = 309$)	Test-retest* ($n = 161$)
Female	56.60%	68.60%	55.90%	68.20%	63.30%
Age	23.21	36.90	21.55	37.61	31.82
	-8.06	-11.75	-6.46	-12.18	(5.45)
Anxiety	8.00%	5.48%	6.29%	5.83%	N/A
Depression	38.67%	34.29%	41.26%	36.57%	N/A
Trauma	24.00%	38.33%	23.08%	35.28%	N/A
Substance Use	3.33%	2.02%	4.90%	3.24%	N/A
Other Psychiatric	26.00%	19.89%	24.47%	19.09%	N/A

Note: Standard deviations for age in parentheses.

After the clinical assessment, participants completed the DASS-21 in addition to other demographic information as part of their intake.

Inputs from Community Clinical Experts

As part of this study, we had regular inputs from community mental health providers with considerable experience implementing evidence-based and evidence-informed approaches while working in mental health settings for broad mental health problems and for heterogeneous groups of Latinx people. This includes expertise delivering services with monolingual and bilingual individuals with various language preferences. Members included leadership and clinical staff from El Futuro (one psychiatrist, two social workers, and three mental health counselors—all licensed in their respective fields).

Importantly, these community mental health providers approached us with a specific challenge: they needed a tool to maintain the utility and reliability of the DASS-21 while also meeting the unique demands of telehealth settings, particularly in the rural areas across the state of North Carolina. They emphasized the growing need for evidence-based approaches to monitoring mental health and integrating measurement-based care into their evidence-informed, evaluation-friendly practice. However, the DASS-21, in its current form was perceived as too lengthy and burdensome for consistent use in these settings. This feedback highlighted the importance of balancing psychometric rigor and practical feasibility for providers invested in quality evidence-informed care in real-world, resource-constrained contexts.

Procedure

Assessments were conducted at one of the El Futuro sites, a community-based mental health treatment center that serves patients from rural and urban backgrounds in North Carolina. Data for the development of the DASS-6 were collected between 2017 and 2019. Patients were assessed by a trained and licensed clinician (MD, PhD) or therapist (PhD-level licensed psychologist, master's-level licensed counselor or clinical social worker) with experience working with Latinx individuals seeking mental health services and trained in the assessment of DSM-5 disorders, including differential diagnosis between anxiety, depression, PTSD, adjustment and stress disorders, and other related disorders (APA, 2013). To ensure reliability using a post Covid-19 pandemic independent sample, 161 individuals were administered the DASS-21 and the DASS-6 for subscale test-retest reliability (interval 5–10 days). These were collected between 2020 and 2023.

Data Analysis Strategy

Tests for Model Fit: Single-Factor and "Total Score" Analog Model

Before fitting of MNLFA models, essential unidimensionality was examined in the calibration sample to see if a single factor generally underlies the DASS-21 items using means and variance-adjusted weighted least squares estimation in M*plus* version 8 (Muthén & Muthén, 2017). Within tests of fit of a general single factor, we also fit a more restrictive model where factor loadings/discrimination parameters were constrained to equality to assess

whether the psychometric model that underlies DASS-21 total scores fit the data (the IRT analog of a one-parameter logistic model; Andrich, 1978). Given the proliferation of the use of DASS-21 total scores, fitting this model is consistent with recent calls in the quantitative methods literature for explicit testing and justification of the use of total scores by ensuring that the psychometric model that is assumed by the total score fits the data (He et al., 2014; McNeish & Wolf, 2020).

MNLFA: Calibration Sample

First, in the calibration sample, an initial base, single-factor MNLFA model was fit where the 21 DASS symptoms were modeled assuming no DIF across language, gender, age, and primary diagnosis. Mean differences in latent NA were also examined in this base model across these variables as the basis for later examination of differential test functioning (DTF; Jones, 2006; Jones & Gallo, 2002). Next, a series of 21 models were fit for each item/symptom where thresholds and loadings were tested to see if they varied across language, gender, age, and primary diagnosis (i.e., DIF). Thresholds indicate the level of the underlying construct (i.e., latent negative affectivity) at which the probabilities of crossing one category to another on an item (e.g., 0 to 1) are equal. Threshold and/or factor loading DIF values that exceeded Cohen's d > |.25|, per DIF effect size recommendations from Educational Testing Service (Zwick, 2012), from this series of models were retained for inclusion in an interim model where all DIF parameters were estimated simultaneously. DIF parameters from the interim model that remained above an effect size of d > |.25| were retained for the final scoring model. For each model, including the final MNLFA scoring model, DIF predictors were centered so that the interpretation of the "main" item parameters was at the mean levels of the predictors. The final MNLFA scoring model was then used to generate latent NA scores in the validation sample that considered DIF across all symptoms. This final MNLFA model also assessed mean differences in latent NA (after accounting for DIF) and compared these mean differences across language, gender, age, and primary diagnosis to those from the base model; because all models had the NA scale set to N(0,1),

these "difference-in-difference" estimates indicate the level of impact of cumulative DIF on NA scores for estimation of DTF in Cohen's d (i.e., effect size) metric (Teresi & Jones, 2016).

Further, selection of the items for the newly proposed DASS-6 was based on three criteria. First, item selection was prioritized by the rank-ordering of item information function values *at the mean of the NA scale for the most severe diagnostic grouping* (i.e., depression diagnosis; see Results section) as recommended by Wyse and Babcock (2016) when maximizing classification accuracy with short forms. Second, priority was given to items that were invariant across all four predictors of DIF. Finally, conditional of the 1st two priorities, two items each were selected from the depression, anxiety, and stress subscales to maximize representation of all three disorders, given the criterion variables were the three diagnoses.

MNLFA: Validation Sample

In the validation sample, both the full DASS-21 and the DASS-6 were scored treating the MNLFA item parameter estimates in the calibration sample as fixed item parameters in the validation sample. Local reliability was then assessed for both the full DASS-21 and the DASS-6. Under CTT, the most common measure of reliability (at least as defined here as internal consistency) used in practice is, of course, Cronbach's α. Under CTT, α presumes that the reliability of a score is constant throughout the range of the construct, which is often unrealistic in practice. In factor analysis/item response theory (FA/IRT), the concept of reliability is "local," or specific to different levels of the construct (Embretson & Reise, 2000); for health outcomes research, a measure is ideally at its maximum reliability at the level of the construct at which a diagnostic decision is made (Chiesi et al., 2017; Morgan-López et al., 2020). To calculate and graph local reliability (LR), test information function (TIF) values are output, where TIF values are the expected value of the inverse of the error variances for each estimated value of the latent construct score; these values can be requested as output from *Mplus*. Then, the TIF values are converted to LR values using $1 - (1/(TIF))$ for each specific value of the latent construct score.

DASS-6 MNLFA Score Differences by Baseline Diagnosis

As part of the scoring model in the validation sample for generation of DASS-6 scores under MNLFA, paths from diagnostic grouping dummy variables to the latent NA with a random intercept structure (to accommodate clustering among raters) was performed with baseline primary diagnosis groupings (depression, anxiety, PTSD, no diagnosis) as the grouping variable. This analysis to test the significance of the differences between groups DASS-6 severity scores generated under MNLFA (e.g., a *de facto* validity check).

Clinical Validation Metrics

To assess the clinical utility of the MNLFA score estimates for the DASS-6, positive and negative predictive values (PPV, NPV) were calculated at specific cut points on the MNLFA score, with the corresponding range of the DASS-6 total score (which, for the 6 items could range from 0 to 18) for assessing the approximate correspondence of the total symptom score to the MNLFA score.

Test-Retest Reliability

After identification of the DASS-6 items, we collected 1-week test-retest reliability data on an independent sample of participants. The DASS-21 was administered the first week and the DASS-6 items on their return visit 1-week later (range: 5–10 days).

Results

Preliminary Tests of Model Fit

The conventional one-factor model for DASS-21 items fit the data well, with comparative fit index = 0.97; root mean square error of approximation = 0.054, 95 percent confidence interval [.051, .057]. A more restrictive model with equality constraints on factor loadings fit comparatively worse against the conventional one-factor model ($\Delta\chi^2(20) = 679.75$; $P < 0.001$). Thus, a psychometric model that assumes that the DASS-21 items have equal weight (i.e., total scores) is likely to produce significant bias in scale scoring.

DASS-21 Calibration Sample MNLFA

Overall descriptive statistics across DASS-21 language groups are shown in Table 1 which includes information about age, gender, and clinical diagnosis. In the calibration sample, the Spanish sample was significantly older (b = 13.69 [3.49]; $P < 0.001$), had higher proportions of women ($\chi^2[1] = 6.54$; $P < 0.001$), and had differences in the diagnostic makeup ($\chi^2[4] = 10.51$; $P = 0.03$). The English sample was most predominant on depression and the Spanish sample was most prominent on stress. The final model MNLFA primary item parameters are shown in Table 2 shows the MNLFA primary item parameters, and Table 3 shows the final model MNLFA factor loading and threshold DIF parameters (that exceeded d > |.25|). As Table 3 shows, 16 of the 21 items were invariant across language (5 showed threshold [i.e., uniform] DIF across language, and none showed factor loading DIF), while 9 items were fully invariant across all DIF factors. The final DASS-21 MNLFA scoring model, with DIF incorporated across all other items, fit significantly better than the base model $\chi^2(16) = 267.50$; $P < 0.0001$.

In the final MNLFA model, Spanish speakers scored higher on NA than English speakers (b = 0.451 [0.115]; $t = 3.929$; $P < 0.001$). Note that, because the latent variable variance was set to 1 for NA, mean differences are already in Cohen's d/standardized mean difference units. Women ($d = 0.42$, $P < 0.001$) and patients with a primary diagnosis of depression ($d = 0.389$, $P = 0.001$) had both statistically significant and meaningful effect size differences compared with men and the other diagnosis grouping, respectively. Patients with a primary diagnosis of anxiety ($d = 0.37$, $P = 0.058$) showed nonsignificant but meaningful elevations on NA, whereas patients with primary PTSD did not show differences ($d = 0.12$, $P = 0.31$). Assessment of DTF (Teresi & Jones, 2016) was conducted to examine the practical impact of collective DIF on scale score estimates. This was done by comparing the effect sizes for differences in latent NA across language/age/gender/diagnosis from the final model against the corresponding effect sizes from the base model with no DIF. For example, the language effect size on latent NA for the base model ($d = 0.461$) versus the language effect size

Table 2. MNLFA main item parameters (calibration sample *N* = 497)

Item	Factor Loading (λ)	Threshold (τ; 0 to 1)	Threshold (1 to 2)	Threshold (2 to 3)
Hard to wind down (*Costo mucho relajarme*)	1.80	-2.89	-0.75	1.41
Dryness of the mouth (*Tenia la boca seca*)	1.09	-0.80	0.47	1.92
Couldn't experience positive feeling (*No podía sentir sentimiento positivo*)	2.04	-1.76	0.24	2.27
Difficulty breathing (*Difícil respirar*)	1.69	-0.35	1.24	3.33
Lack of initiative (*Difícil tomar iniciativa*)	1.79	-2.08	-0.28	1.83
Overreacting (*Reaccione exageradamente*)	1.48	-1.99	-0.37	1.73
Trembling hands (*Mis manos temblaban*)	2.01	-0.86	0.55	2.41
Nervous energy (*Tenía muchos nervios*)	2.48	-2.82	-0.55	1.66
Worried about panic/fool (*Preocupado de pánico o ser ridículo*)	1.95	-0.92	0.41	2.24
Nothing to look forward to (*No tenía nada por que vivir*)	2.46	0.20	1.91	3.71
Agitated (*Me agitaba*)	1.87	-1.78	0.38	2.50
Difficult to relax (*Difícil relajarme*)	2.41	-2.79	-0.80	1.82
Feeling blue (*Me siento triste y deprimido*)	2.75	-3.52	-1.03	1.66
Intolerant (*No tolere nada*)	1.79	-1.69	0.06	2.05
Close to panic (*Al punto de panico*)	2.64	-0.64	1.23	3.53
Unable to become enthusiastic (*No me podia entusiasmar*)	2.49	-1.26	0.89	3.26
Worthless (*Valia muy poco*)	2.43	-0.82	0.69	2.82
Touchy (*Irritable*)	1.88	-1.67	-0.04	2.15
Aware of increased heart rate (*Sentí latidos del corazón*)	2.08	-0.90	0.61	2.78
Scared (*Miedo sin razón*)	2.38	-1.23	0.64	2.55
Life was meaningless (*La vida no tiene sentido*)	2.16	0.09	1.55	3.27

Note: MNLFA = moderated nonlinear factor analysis.

from the final model (*d* = 0.451) show a "difference-in-difference" effect size (ES) that was negligible (Δ*d* = |.01|). All other similar Δ*d* values were less than |.042|, suggesting little-to-no practical impact on NA scale scores of modeled item-level DIF on DTF.

Item information curves (IICs) were examined to select the optimal items for the proposed DASS-6 in this order of priority: (a) rank ordering of the IIC values at a "clinical cut point" (e.g., the midpoint of the distribution of latent NA for patients with depression, +.39 SDs above the mean on latent NA); (b) invariance across all DIF factors (if possible); (c) invariance across language (even if DIF by other predictors is present); (d) representation across DASS-21 subdimensions (despite the unidimensional fit noted previously); and (e) preference for the items based of provider input. Table 4 shows the selected items (with IICs in Figure 1) and the joint criteria by which the DASS-6 items were selected: *Difficulty*

breathing (item 4; anxiety), *Close to panic* (item 15; anxiety), *Nothing to look forward to* (item 10; depression), *Unable to become enthusiastic* (item 16; depression), *Nervous energy* (item 8; stress), *Difficulty relaxing* (item 12; stress). Note that two items that may have reflected superior item properties (*Feeling blue*, *Aware of increased heart rate*) were not included. *Feeling blue* (item 13; depression), despite having the second largest IIC overall, was not selected because of DIF by language; this likely reflects the conceptual differences in the English colloquialism "feeling blue" not having a direct analog in Spanish. Providers for the study remarked that *Difficulty breathing* might more clearly reflect a physiological symptom that is clearly associated with distress than might an increase in heart rate (item 19; anxiety), which could be due to emotions such as anger or excitement.

Table 3. MNLFA differential item functioning parameters (calibration sample N = 497)

| Item	Language τ DIF	Gender λ DIF	Gender τ DIF	Age τ DIF	Depression λ DIF	Depression τ DIF	Anxiety τ DIF	PTSD λ DIF	PTSD τ DIF
Hard to wind down	0.79								
Dryness of the mouth	0.49				0.59				0.69
Couldn't experience positive feeling									
Difficulty breathing									
Lack of initiative				-0.03					
Overreacting				-0.02					
Trembling hands									
Nervous energy									
Worried about panic/fool							1.07		
Nothing to look forward to	-1.66		-0.78		0.76				
Agitated	-1.29								
Difficult to relax									
Feeling blue	1.04								
Intolerant				-0.02					
Close to panic									
Unable to become enthusiastic									
Worthless						0.88			
Touchy									
Aware of increased heart rate									
Scared								0.82	
Life was meaningless		-1.08							

Notes: MNLFA = moderated nonlinear factor analysis; τ = Threshold; DIF = differential item functioning; λ = factor loading; PTSD = posttraumatic stress disorder.

Table 4. DASS-6 candidate items

Item name	Item information curve (IIC) value	Item number	Overall anchor item?	Disorder dimension
Close to panic	**2.05**	**Item 15**	**Yes**	**Anxiety**
Feeling blue*	2.01	Item 13	No	Depression
Nothing to look forward to	**1.79**	**Item 10**	**No**	**Depression**
Unable to become enthusiastic	**1.78**	**Item 16**	**Yes**	**Depression**
Feeling worthless	1.75	Item 17	No	Depression
Nervous energy	**1.73**	**Item 8**	**Yes**	**Stress**
Difficult to relax	**1.55**	**Item 12**	**Yes**	**Stress**
Scared	1.31	Item 20	No	Anxiety
Aware of increased heart rate*	1.28	Item 19	Yes	Anxiety
Trembling hands	1.22	Item 7	Yes	Anxiety
Difficulty breathing	**0.85**	**Item 4**	**Yes**	**Anxiety**

Notes: DASS = Depression, Anxiety, and Stress Scales. Final model calibration sample moderated nonlinear factor analysis IIC values output at 0.39 standard deviations above the mean.

* Indicates additional candidate items that were under consideration.

Figure 1. Item information curves
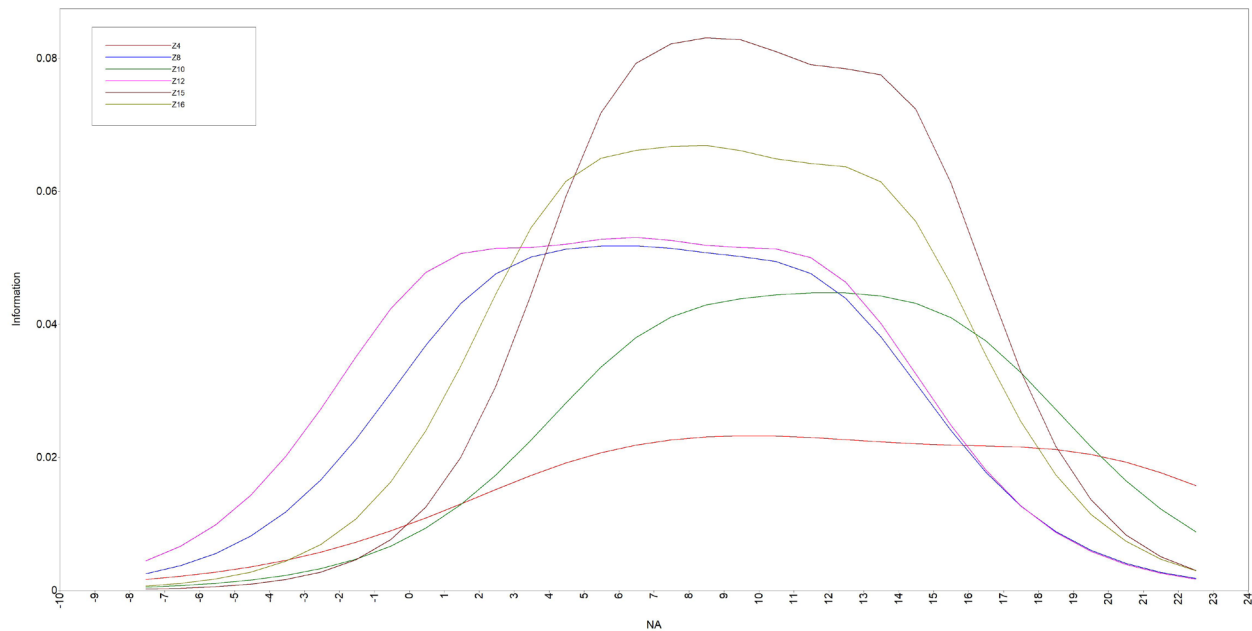
## Validation Sample: Scoring and Local Reliability

In the validation sample, both the full DASS-21 and DASS-6 were scored for the English and Spanish versions using the item parameter estimates from either the full 21-item set or the selected set of 6 items based on parameter estimates in Tables 2 and 3. Local reliability (overall and conditional on predictors) was then assessed for scale scores estimated from the full DASS-21 and the DASS-6 based on conversion of the test information function values (Chiesi et al., 2017). Figure 2 shows the local reliabilities for the DASS-21 (solid line) and the proposed DASS-6 for English, Spanish, and overall (dashed lines). Local reliability values for the DASS-21 MNLFA scores remained above 0.92 throughout the range of latent NA, and these reliability values did not vary across language, gender, age, or diagnosis. The DASS-6 scale scores overall and across language (no DIF was observed in the six selected items other than threshold DIF by language for "Nothing to Look Forward To") remained above 0.85 reliability throughout most of the practical range of latent negative affectivity, with a peak reliability of 0.90 at the +.39 standard deviation (SD) clinical cut point.

Random Effects ANOVA: DASS-6 MNLFA Score Differences

An ANOVA model with a random intercept structure was performed with baseline diagnosis groupings (depression only [D], anxiety only [A], PTSD only [P], no depression, anxiety or PTSD [no DAP]) as the grouping variable; the no DAP diagnosis group was used as the reference group. The model accounted for a significant amount of variability in DASS-6 MNLFA scores in the validation sample (F [3, 488] = 11.21; $P < 0.0001$; $r^2 = 0.064$). The depression ($d = 0.85$; $P <0.001$), anxiety ($d = 0.99$; $P < 0.001$), and PTSD ($d = 0.55$; $P = 0.018$) diagnostic groups showed significant and meaningful differences in DASS-6 NA scale scores compared with the no DAP group. Further, none of these effect size differences varied by more than |.02| in a comparable random effects ANOVA for MNLFA scale scores from the full DASS-21 in the validation sample.

Clinical Validation Metrics

Table 5 presents clinical validation metrics across a selected range of NA. At 0.5 SDs) above the mean DASS-6 MNLFA scale score (close to the 0.39 SDs used for IIC selection), PPVs (e.g., the proportion of patients with a depression diagnosis that would have been detected as such with the DASS-6) ranged from

Figure 2. Overall and conditional local reliability plots
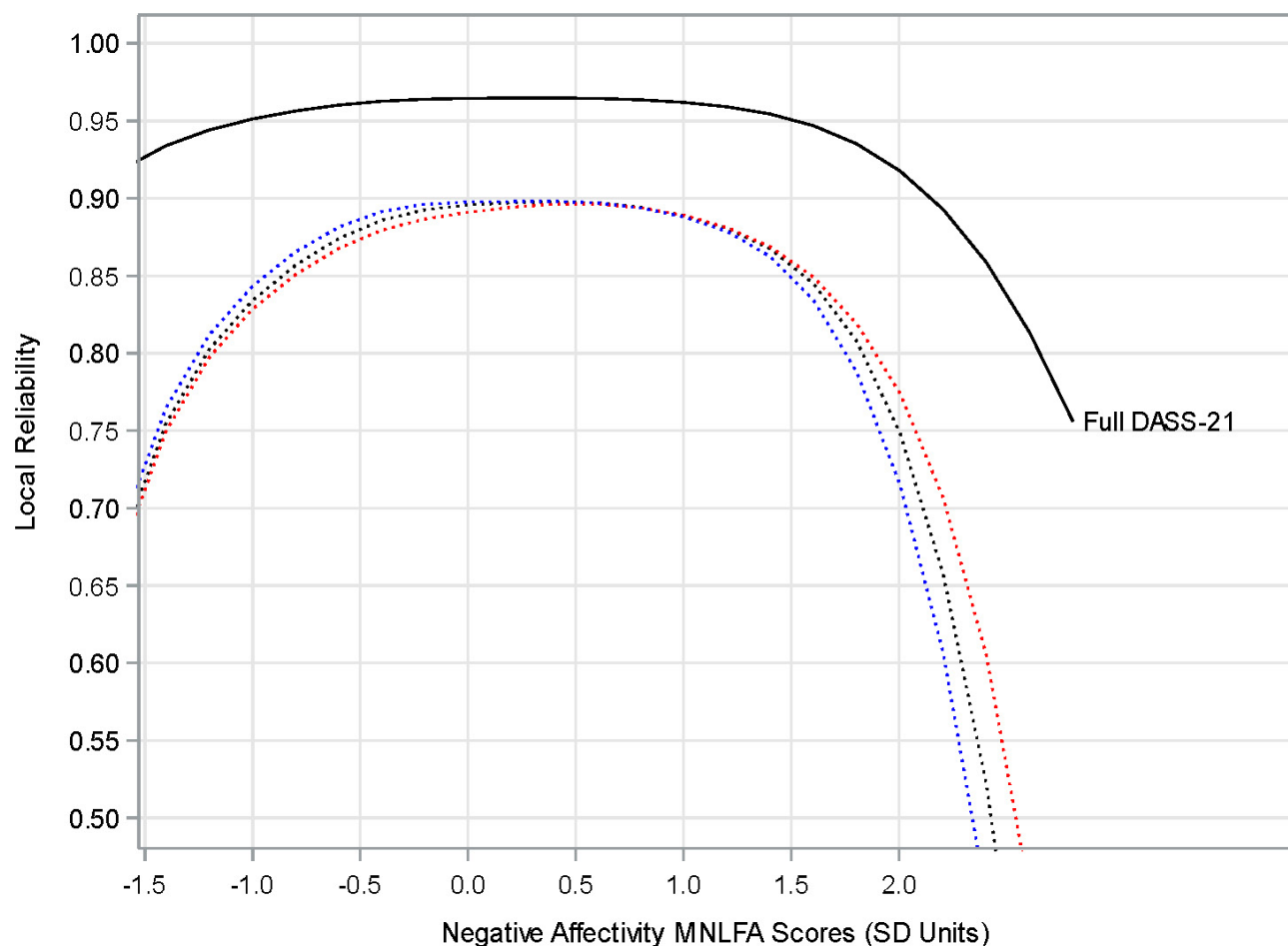
0.88 to 0.95 across the three disorders, while NPVs (e.g., the proportion of patients without a depression diagnosis that would have depression ruled out with the DASS-6) ranged from 0.53 to 0.64. The DASS-6 MNLFA cut score roughly translates to a DASS-6 total score of 10, with some patients reaching the clinical range with scores of 8 (depending on the combination of symptoms); because of the differences in item weighting via factor loading variation across items, the same total score can lead to a different MNLFA score estimate and vice versa (Morgan-López et al., 2024; Saavedra et al., 2021 Saavedra et al., 2022). The correlation between the DASS-6 total scores and MNLFA scores (r = 0.98) with a scatterplot shown in Figure 3. Selection of a DASS-6 cut score with a high PPV and low-to-moderate NPV reflects a clinical context where false negatives are more acceptable than false positives, such as patients who may be in subthreshold clinical distress that could still benefit from treatment (Trevethan, 2017).

Independent Sample Test-Retest Reliability

We also included an independent sample of patients (*n* = 161). For these individuals, the Cronbach's alpha coefficient was 0.93 for the DASS-6; it was 0.88 for depression, 0.88 for anxiety, and 0.89 for stress, indicating the acceptable internal consistency of the scale. In addition, the test-retest reliability of the DASS-21 and DASS-6 was in the good range (0.89) and was acceptable, and the ICC of all domains was adequate, ranging from 0.75 to 0.86.

Table 5. Positive and negative predictive values at each MNLFA score cut point by diagnosis (validation sample *N* = 452)

DASS-6 MNLFA cut point (in SD units)	PPV depression	NPV depression	PPV anxiety	NPV anxiety	PPV PTSD	NPV PTSD	DASS-6 raw score mean	DASS-6 raw score minimum	DASS-6 raw score maximum
-0.5 [-.75 to -0.25]	0.73	0.72	0.74	0.93	0.69	0.62	4.3	2	7
0 [-.25 to 0.25]	0.81	0.63	0.88	0.81	0.79	0.60	7.1	5	10
0.5 [.25 to 0.75]	**.92**	**.55**	**.95**	**.64**	**.88**	**.53**	**10.3**	**8**	**14**
1 [.75 to 1.25]	0.96	0.46	0.98	0.52	0.98	0.46	13.0	10	17

Notes: MNFLA = moderated nonlinear factor analysis; DASS = Depression, Anxiety, and Stress Scales; SD = standard deviation; PPV = positive predictive value; NPV = negative predictive value; PTSD = posttraumatic stress disorder. Clinical threshold in boldface.

Discussion

The systematic and precise measurement of symptoms before, during, and after treatment is an important tenet in evidence-based practice (Joiner et al., 2005). This is routinely challenging for providers in real-life settings where there is limited time to gather needed information. The development of reliable brief instruments that allow providers and clients to maximize their in-session time while still capturing psychometrically sound and clinically useful information is critical. Additionally, for services provided to the US Latinx community, having an instrument that assesses depression, anxiety, and stress is particularly important (Chang

Figure 3. DASS-6 MNLFA scores and total scores scatterplot
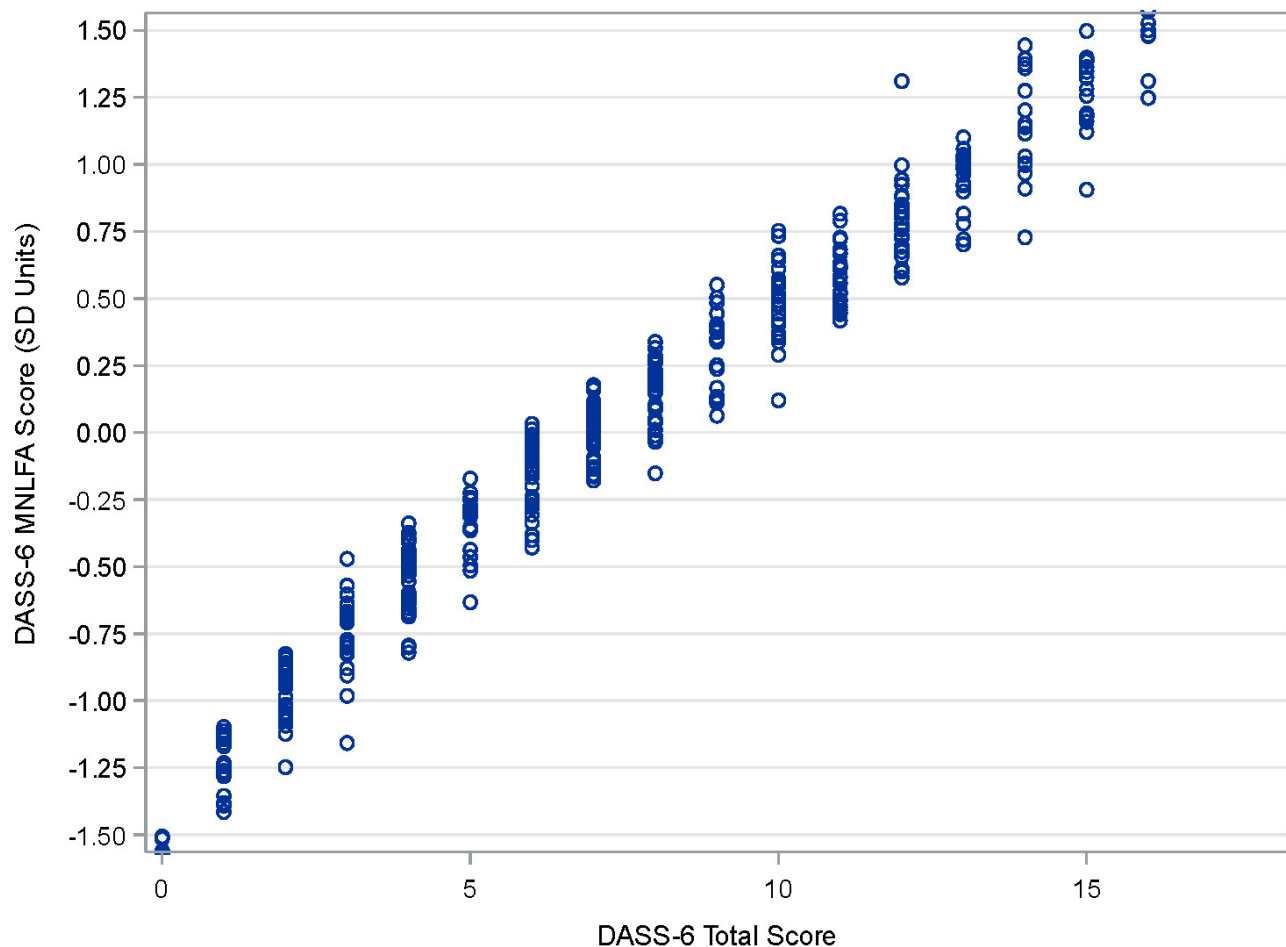
Notes: DASS = Depression, Anxiety, and Stress Scales; MNFLA = moderated nonlinear factor analysis; SD = standard deviation.

& Biegel, 2018; Derr, 2016; Stone, Fernandez & DeSantiago, 2022); this is especially relevant for clinicians serving both English- and Spanish-speaking clients. Having one assessment that can be reliably administered in both languages reduces additional burden placed on clinicians to learn and master multiple new instruments.

In close consultation with community mental health providers, this study used advances in MNLFA to identify common mental health symptoms in a Latinx clinical population of adult patients who present to a community-based outpatient treatment center. The six items that make up the DASS-6 include: *Difficulty breathing* (DASS-21 item 4); *Nervous energy* (DASS-21 item 8); *Nothing to look forward to* (DASS-21 item 10); *Difficult to relax* (DASS-21 item 12); *Close to panic* (DASS-21 item 15); and *Unable to become enthusiastic* (DASS-21 item 16). Results showed that these six items can reliably discriminate anxiety and depression severity levels in both the Spanish and English versions of the abbreviated instrument. This is critical for monitoring in telehealth settings where shorter instruments are easier to implement. This tool aids providers in accurately identifying and monitoring the severity of mental health conditions in a culturally and linguistically competent manner. The DASS-6 can be an especially valuable tool in early identification and prevention planning for individuals who may require more time to develop rapport and for whom there is a higher likelihood for their symptoms to escalate to more severe levels.

We also underscore additional benefits to the DASS-6. The DASS-6 is an efficient tool for measuring the severity of overall NA (i.e., symptoms of depression, anxiety, and stress) while accounting for measurement bias across Spanish and English language usage. An important contribution is the development of the shortened version in both English and Spanish. Although one must take caution when using this screening instrument as a diagnostic tool, for many settings such as telehealth and primary care, this reliable brief form, which requires significantly less time to administer and score, can be used to quickly identify patients who require more assessment and eventually treatment.

Moreover, this instrument can be useful for monitoring treatment progress, especially in settings where time is particularly limited. Initial test-retest reliability and internal consistency of the DASS-6 provide preliminary support for its suitability for ongoing monitoring of symptom severity over time. In telehealth settings, where longitudinal patient engagement can be challenging, the DASS-6 offers a reliable means of tracking treatment progress and outcomes, supporting more personalized and dynamic treatment planning as well as measurement-based care. Further, previous analysis of the DASS-21 showed that, while several items showed differential item functioning across time (which could compromise precision of estimates of individual patient changes), this DIF was not observed among the items selected for the DASS-6 (Morgan-López et al., 2022). Despite these initial findings, additional research will be needed to determine the utility of the DASS-6 in capturing change over time, particularly after an intervention targeting depression, anxiety, and stress symptoms. Future research should also examine whether information gathered by the DASS-6 shows similar group- and individual-level change in response to treatment relative to the DASS-21.

An important strength of this study lies in the collaboration with community mental health providers, who emphasized the need for a brief yet reliable tool that could be integrated with telehealth and in-person settings. Providers working in rural areas identified a gap in tools that could efficiently monitor symptoms while adhering to evidence-informed approaches for monitoring and measurement-based care. Although the DASS-21 offers a shortened alternative to the original DASS, it remained too lengthy to be practical, especially in telehealth contexts, where time is a premium and patient engagement is key. By using established methodological approaches like MNLFA for deriving the DASS-6, a meaningful solution for both clinicians and clients, the instrument ensures utility, reliability, and feasibility in real-world settings. The partnership with clinicians underscores the importance of designing tools that are both psychometrically sound and meet the practical demands of providers working with underserved populations.

Limitations

Results from this study may be specific to Latinx populations presenting to community-based care organizations. Additional research should examine the performance of the DASS-6 as a brief screener with non-Latinx individuals in similar primary care and/or outpatient community-based organizations. Although this does limit generalizability to other populations, there is a need for more precise measurement tools for this heterogeneous and underserved group of people. Nevertheless, additional studies with different populations should be conducted. Also, the DASS-6 (and the DASS-21) measure general anxiety and not specific anxiety disorders. Future research should examine the utility of the DASS-6 in differentiating between specific anxiety disorders and explore whether item severity has a transdiagnostic component for measuring anxiety disorder severity, as these research questions were beyond the scope of the current study. Finally, in this study, assessments were conducted by experienced and licensed clinicians and therapists all with substantial experience with anxiety and depressive disorders, but fidelity was not monitored.

Limitations notwithstanding, the present study provides an approach for reliably developing other abbreviated symptom monitoring instruments that would be beneficial to clinicians. One of the silver linings of the COVID-19 pandemic has been the expansive capacity built around telehealth service provision. Post-pandemic, there will undoubtably be a continued need for telehealth services to reach individuals with limited resources to access care. Telehealth service provision has opened avenues for this work, but more research is needed on assessment and treatment monitoring. This study using MNLFA to reliably identify the most relevant symptoms can help researchers and practitioners to continue to work toward access to quality mental health services in clinical settings.

Overall, the validation of a shortened, bilingual assessment tool for depression, anxiety, and stress represents a significant step forward in making mental health care more accessible and effective for the Latinx community. It aligns with the broader goal of enhancing mental health service delivery through improved assessment and treatment monitoring, particularly in the context of the burgeoning field of telehealth. This work not only benefits clinicians by providing efficient, reliable tools in dual languages, but also supports the wider aim of improving mental health outcomes for underserved populations.

Data Availability Statement

The data supporting the current study are available from the authors upon reasonable request.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Bados, A., Solanas, A., & Andrés, R. (2005). Psychometric properties of the Spanish version of Depression, Anxiety, and Stress Scales (DASS). *Psicothema*, *17*, 679–683.

Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2020). The COVID-19 pandemic and health inequalities. *Journal of Epidemiology and Community Health*, *74*(11), 964–968. https://doi.org/10.1136/jech-2020-214401

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological Methods*, *14*(2), 101. https://doi.org/10.1037/a0015583

Castro, F. G., Barrera, M., Jr., & Holleran Steiker, L. K. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology*, *6*(1), 213–239. https://doi.org/10.1146/annurev-clinpsy-033109-132032

Chang, C. W., & Biegel, D. E. (2018). Factors affecting mental health service utilization among Latino Americans with mental health issues. *Journal of Mental Health (Abingdon, England)*, *27*(6), 552–559. https://doi .org/10.1080/09638237.2017.1385742

Chiesi, F., Primi, C., Pigliautile, M., Ercolani, S., Della Staffa, M. C., Longo, A., Boccardi, V., & Mecocci, P. (2017). The local reliability of the 15-item version of the Geriatric Depression Scale: An item response theory (IRT) study. *Journal of Psychosomatic Research*, *96*, 84–88. https://doi.org/10.1016/j.jpsychores.2017.03.013

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory factor analysis of the Depression-Anxiety-Stress Scale in depressed and anxious patients. *Journal of Psychopathology and Behavioral Assessment*, *23*(1), 61–67. https://doi.org/10.1023/A:1011095624717

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*(2), 365–380. https://doi.org/10.1037/ 0012-1649.44.2.365

Daza, P., Novy, D. M., Stanley, M. A., & Averill, P. (2002). The Depression and Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment*, *24*(3), 195–205. https://doi.org/10.1023/A: 1016014818163

Derr, A. S. (2016). Mental health service use among immigrants in the United States: A systematic review. *Psychiatric Services (Washington, D.C.)*, *67*(3), 265–274. https://doi.org/10.1176/appi.ps.201500004

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *16*(S1), 5–18. https://doi.org/10.1007/s11136-007-9198-0

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists* (Multivariate Applications Books Series). Lawrence Erlbaum Associates Publishers.

Escovar, E. L., Craske, M., Roy-Byrne, P., Stein, M. B., Sullivan, G., Sherbourne, C. D., Bystritsky, A., & Chavira, D. A. (2018). Cultural influences on mental health symptoms in a primary care sample of Latinx patients. *Journal of Anxiety Disorders*, *55*, 39–47. https:// doi.org/10.1016/j.janxdis.2018.03.005

Fortuna, L. R., Álvarez, K., Ramos Ortiz, Z., Wang, Y., Mozo Alegría, X., Cook, B. L., & Alegría, M. (2016). Mental health, migration stressors and suicidal ideation among Latino immigrants in Spain and the United States. *European Psychiatry*, *36*, 15–22. https://doi.org/ 10.1016/j.eurpsy.2016.03.001

Gomez, F. (2016). *A guide to the Depression, Anxiety and Stress Scale (DASS 21)*. Central and Eastern Sydney Primary Health Networks.

He, Q., Glas, C. A., & Veldkamp, B. P. (2014). Assessing impact of differential symptom functioning on post traumatic stress disorder (PTSD) diagnosis. *International Journal of Methods in Psychiatric Research*, *23*(2), 131–141. https://doi.org/10.1002/mpr .1417

Henry, J. D., & Crawford, J. R. (2005). The short form version of the Depression Anxiety Stress Scales (DASS 21): Construct validity and normative data in a large non clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239. https://doi.org/10.1348/ 014466505X29657

Joiner, T. E., Jr., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment*, *17*(3), 267–277. https://doi.org/10.1037/1040-3590.17.3.267

Jones, R. N. (2006). Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: Detecting differential item functioning using MIMIC modeling. *Medical Care*, *44*(11), S124–S133. https://doi .org/10.1097/01.mlr.0000245250.50114.0f

Jones, R. N., & Gallo, J. J. (2002). Education and sex differences in the mini-mental state examination: Effects of differential item functioning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(6), P548–P558. https://doi.org/10.1093/ geronb/57.6.P548

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology*, *107*(3), 520–526. https://doi.org/10.1037/ 0021-843X.107.3.520

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*. Psychology Foundation.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates Publishers.

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Millsap, R. E., & Kwok, O.-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*(1), 93–115. https://doi.org/10.1037/1082-989X.9.1.93

Morgan-López, A. A., Saavedra, L. M., Hien, D. A., Killeen, T. K., Back, S. E., Ruglass, L. M., Fitzpatrick, S., López-Castro, T., & Patock-Peckham, J. A. (2020). Estimation of equable scale scores and treatment outcomes from patient- and clinician-reported PTSD measures using item response theory calibration. *Psychological Assessment*, *32*(4), 321–335. https://doi.org/10.1037/pas0000789

Morgan-López, A. A., Saavedra, L. M., Prakash, K., West, S. G., & Hien, D. A. (2024). The hidden psychometric model underlying the DSM: What happens when it fails to fit psychiatric data. *Psychiatric Services (Washington, DC)*, *75*(9), 937–938. https://doi.org/10.1176/appi.ps.20230370

Morgan-López, A. A., Saavedra, L. M., Ramirez, D. D., Smith, L. M., & Yaros, A. C. (2022). Adapting the multilevel model for estimation of the reliable change index (RCI) with multiple timepoints and multiple sources of error. *International Journal of Methods in Psychiatric Research*, *31*(2), e1906. https://doi.org/10.1002/mpr.1906

Muthén, B., & Muthén, L. (2017). Mplus. In W. J. van der Linden (Ed.), *Handbook of item response theory* (pp. 507–518). Chapman and Hall/CRC. https://doi.org/10.1201/9781315117430

Norton, P. J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, Stress, and Coping*, *20*(3), 253–265. https://doi.org/10.1080/10615800701309279

Patel, S. V., Saavedra, L. M., Rodriguez Borja, I., Philbrick, S., Schwimmer, M., Ruwala, R., & Viswanathan, M. (2025). Audio-based care for managing mental Health and substance use disorders in adults: A systematic review. *Medical Care*, *63*(2), 134–151. https://doi.org/10.1097/MLR.0000000000002098

Ruiz, F. J., García Martín, M. B., Suárez Falcón, J. C., & Odriozola González, P. (2017). The hierarchical factor structure of the Spanish version of Depression Anxiety and Stress Scale-21. *International Journal of Psychology & Psychological Therapy*, *17*(1), 97–105.

Saavedra, L. M., Johnson, K. L., Counts, J., Spinks, S. N., Yaros, A. C., Hayes, M., & Smith, L. (2024). Provider perspectives on delivering telemental health services in rural North Carolina: Field report. *Rural Mental Health*, *48*(1), 26–35. https://doi.org/10.1037/rmh0000249

Saavedra, L. M., Morgan-López, A. A., Back, S. E., Patel, S. V., Hien, D. A., Killeen, T. K., Norman, S. B., Fitzpatrick, S., Ebrahimi, C. T., & Ruglass, L. M. (2022). Measurement error-corrected estimation of clinically significant change trajectories for interventions targeting comorbid PTSD and substance use disorders in OEF/OIF veterans. *Behavior Therapy*, *53*(5), 1009–1023. https://doi.org/10.1016/j.beth.2022.04.007

Saavedra, L. M., Morgan-López, A. A., Hien, D. A., Killeen, T. K., Back, S. E., Ruglass, L. M., Fitzpatrick, S., & Lopez-Castro, T. (2021). Putting the patient back in clinical significance: Moderated nonlinear factor analysis for estimating clinically significant change in treatment for posttraumatic stress disorder. *Journal of Traumatic Stress*, *34*(2), 454–466. https://doi.org/10.1002/jts.22624

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(S1), 1–97. https://doi.org/10.1007/BF03372160

Samejima, F. (2016). Graded response models. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume one* (pp. 123–136). Chapman and Hall/CRC. https://doi.org/10.1201/9781315374512

Sanchez, M., Saavedra, L., Lozano, A., Morgan-López, A., & Amaro, H. (2024). Promising intervention and measurement directions for substance use treatment research among Hispanics. *Current Addiction Reports*, *11*(3), 492–500. https://doi.org/10.1007/s40429-024-00567-x

Scholten, S., Velten, J., Bieda, A., Zhang, X. C., & Margraf, J. (2017). Testing measurement invariance of the Depression, Anxiety, and Stress Scales (DASS-21) across four countries. *Psychological Assessment*, *29*(11), 1376–1390. https://doi.org/10.1037/pas0000440

Stone, G. A., Fernandez, M., & DeSantiago, A. (2022). Rural Latino health and the built environment: A systematic review. *Ethnicity & Health*, *27*(1), 1–26. https://doi.org/10.1080/13557858.2019.1606899

Teresi, J. A., & Jones, R. N. (2016). Methodological issues in examining measurement equivalence in patient reported outcomes measures: Methods overview to the two-part series, "Measurement equivalence of the patient reported outcomes measurement information system®(PROMIS®) short forms." *Psychological Test and Assessment Modeling*, *58*(1), 37.

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health*, *5*, 307.

Uscher-Pines, L., Sousa, J., Jones, M., Whaley, C., Perrone, C., McCullough, C., & Ober, A. J. (2021). Telehealth use among safety-net organizations in California during the COVID-19 pandemic. *Journal of the American Medical Association*, *325*(11), 1106–1107. https://doi.org/10.1001/jama.2021.0282

Wyse, A. E., & Babcock, B. (2016). How does calibration timing and seasonality affect item parameter estimates? *Educational and Psychological Measurement*, *76*(3), 508–527. https://doi.org/10.1177/0013164415588947

Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, *2012*(1), i-30. https://doi.org/10.1002/j.2333-8504.2012.tb02290.x

RTI International is an independent scientific research institute dedicated to improving the human condition. Our vision is to address the world's most critical problems with technical and science-based solutions in pursuit of a better future. Clients rely on us to answer questions that demand an objective and multidisciplinary approach—one that integrates expertise across social, statistical, data, and laboratory sciences, engineering, and other technical disciplines to solve the world's most challenging problems.