ARTÍCULO ORIGINAL

Gac Méd Caracas 2026;134(Supl 2):S502-S513
DOI: 10.47307/GMC.2026.134.S2.22

Psychometric evaluation of the Spanish DASS-21 among Venezuelan medical students: Internal consistency and three-factor structure

Evaluación psicométrica del DASS-21 en español entre estudiantes de medicina venezolanos: consistencia interna y estructura de tres factores

Mervin Chávez-Castillo[1a], Daniela Ariza[2a], María Gotera[3a], Jim Palmar[4a], Paola Ramírez[5a], María Suárez[6a], Wheeler Torrez[7a], Juan Hernández-Lalinde[8b], Valmore Bermúdez[9c], Diego Rivera-Porras[10d]

SUMMARY

Background: *Medical students commonly report elevated psychological distress. In contexts where locally validated instruments are scarce, brief symptom screening tools with stable measurement properties are required. This study examined the psychometric performance of the Spanish 21-item Depression, Anxiety and Stress Scale (DASS-21) in Venezuelan medical students.*

Material and Methods: *A cross-sectional field study was conducted with 229 students from years 1-6, recruited via non-probabilistic sampling. Participants completed the Spanish DASS-21. Internal consistency was estimated using Cronbach's alpha. Construct adequacy and factorability were assessed using sampling adequacy statistics and factor-analytic procedures.*

Results: *Internal consistency was high (Cronbach's α = 0.924). Sampling adequacy supported factor extraction (KMO = 0.908). A three-factor configuration was supported; the retained CFA specification showed RMSEA = 0.084 (0.067–0.102), CFI = 0.985, TLI = 0.980, and SRMR = 0.068. Measurement invariance by sex met incremental-fit criteria (max $|\Delta CFI|$ = 0.002; max $|\Delta RMSEA|$ = 0.011).*

Conclusions: *In this cohort, the Spanish DASS-21 demonstrated adequate internal psychometric performance, supporting its use as a brief dimensional*

DOI: https://doi.org/10.47307/GMC.2026.134.S2.22

ORCID: 0000-0001-8511-0230[1]
ORCID: 0000-0002-0123-3629[2]
ORCID: 0009-0004-3959-9820[3]
ORCID: 0000-0002-8582-2155[4]
ORCID: 0000-0002-5986-0246[5]
ORCID: 0000-0003-0478-8950[6]
ORCID: 0000-0002-3692-9917[7]
ORCID: 0000-0001-6768-1873[8]
ORCID: 0000-0003-1880-8887[9*]
ORCID: 0000-0003-2169-3208[10*]

Recibido: 20 de diciembre 2025
Aceptado: 1 de febrero 2026

[a]Endocrine and Metabolic Diseases Research Center. School of Medicine, University of Zulia. Maracaibo 4001, Venezuela.
[b]Instituto de Estadística Aplicada y Computación, IEAC. Universidad de Los Andes; Mérida 5115, Venezuela.
[c]Universidad Simón Bolívar, Centro de Investigaciones en Ciencias de la Vida; Barranquilla 080001, Atlántico, Colombia; E-mail: valmore.bermudez@unisimon.edu.co (V.B.)
[d]Universidad de la Costa, Departamento de Productividad e Innovación; Barranquilla, Atlántico, Colombia, 080001; E-mail: drivera23@cuc.edu.co (D.R.P.)

*Corresponding author: Universidad Simón Bolívar, Centro de Investigaciones en Ciencias de la Vida; Barranquilla 080001, Atlántico, Colombia; E-mail: valmore.bermudez@unisimon.edu.co and Diego Rivera-Porras, Universidad de la Costa, Departamento de Productividad e Innovación; Barranquilla, Atlántico, Colombia, 080001 . E-mail: drivera23@cuc.edu.co

symptom measure for screening and monitoring emotional distress in medical education settings, without implying diagnostic validity.

Keywords: *DASS-21, depression, anxiety, stress, medical students, construct validity, factor analysis, internal consistency.*

RESUMEN

Antecedentes: *Los estudiantes de medicina suelen reportar niveles elevados de malestar psicológico. En contextos con evidencia psicométrica local limitada, se requieren instrumentos breves de tamizaje sintomático con propiedades de medición estables. Este estudio examinó el desempeño psicométrico de la versión en español de la Escala de Depresión, Ansiedad y Estrés de 21 ítems (DASS-21) en estudiantes de medicina venezolanos.*
Materiales y Métodos: *Estudio de campo transversal con 229 estudiantes de 1.º a 6.º año, reclutados mediante muestreo no probabilístico. Los participantes completaron la DASS-21 en español. La consistencia interna se estimó mediante el alfa de Cronbach. La adecuación muestral y la factorizabilidad se evaluaron mediante estadísticas de adecuación y procedimientos de análisis factorial.*
Resultados: *La consistencia interna fue alta (α = 0,924). La adecuación muestral respaldó la extracción factorial (KMO = 0,908). La estructura latente fue compatible con una configuración de tres factores; el modelo CFA retenido mostró RMSEA = 0,084 (0,067–0,102), CFI = 0,985, TLI = 0,980 y SRMR = 0,068. La invarianza por sexo cumplió los criterios de cambio incremental (máx $|\Delta$CFI| = 0,002; máx $|\Delta$RMSEA| = 0,011).*
Conclusiones: *En esta cohorte, la DASS-21 en español mostró un desempeño psicométrico interno adecuado, lo que respalda su uso como medida dimensional breve para tamizaje y monitoreo de malestar emocional en contextos de educación médica, sin implicar validez diagnóstica.*

Palabras clave: *DASS-21; depresión; ansiedad; estrés; estudiantes de medicina; validez de constructo; análisis factorial; consistencia interna.*

INTRODUCTION

Depressive and anxiety disorders remain among the leading contributors to non-fatal health loss worldwide, and the COVID-19 period added a measurable excess burden in many settings (1,2). Within this wider epidemiological landscape, medical undergraduates occupy a distinctive risk niche. Academic intensity, high-stakes assessment, early clinical exposure, and professional identity pressures concentrate during a developmental window in which emotional symptoms are already common in the general population. A sizeable synthesis reported pooled prevalence estimates of depression or depressive symptoms in medical students at 27.2 %, with suicidal ideation at 11.1 % (3). A separate meta-analysis focusing on diagnostic and screening-based depression estimates also documented substantial pooled prevalence in this population (4). Anxiety shows a similarly concerning pattern: a global meta-analysis estimated anxiety prevalence at 33.8 % in medical students (5). Burnout adds an overlapping, partly distinct dimension of distress, with recent pooled estimates indicating that it affects more than one-third of students in available syntheses (6). These figures matter because psychological distress in training is linked to impaired functioning, reduced learning engagement, and risk-relevant trajectories that may persist into residency and practice (7).

At the level of educational systems, student well-being is not a purely individual matter. Interventions targeting the learning environment—such as changes to assessment structures, mentoring models, and support programmes—have been associated with improved emotional well-being in systematic evidence. However, effects vary by intervention type and study quality (8). Recent overviews emphasise that mindfulness-based initiatives, structured mental health programmes, and pass/fail grading are recurrent candidates, while also noting methodological limitations across the evidence base (9). Complementary evidence suggests that perceptions of learning mindsets and learning environments correlate with both well-being and ill-being outcomes in medical training contexts(10). Even with upstream educational reforms, credible measurement remains indispensable: screening, programme evaluation, and research on determinants of distress all depend on instruments with defensible measurement properties.

The Depression Anxiety Stress Scales were developed to separate core symptom clusters of

depression, anxiety, and stress within a common negative-affect space (11). The 21-item short form (DASS-21) is widely used because it is brief, public-facing, and suitable for community, clinical, and student samples (12). Items are rated on a 4-point frequency/severity scale, producing three subscale scores that can be interpreted as dimensional indicators of symptom severity rather than categorical diagnoses (11). For medical education research, this profile is practical: it permits efficient monitoring across academic cycles and facilitates comparisons across cohorts and interventions.

Despite its widespread use, the DASS-21's internal structure remains debated. Many studies support a three-factor model that is correlated, consistent with the intended subscales (13,14). Other work argues for hierarchical or bifactor representations, in which a broad distress factor accounts for shared variance across items, alongside narrower domain factors (15,16). Measurement invariance is an additional concern. If items function differently across gender or other subgroups, observed score differences may reflect measurement artefacts rather than genuine symptom variation (17,18). These issues are not merely technical: in medical student populations, where overlapping symptom profiles are common, a scale that collapses domains into a dominant general factor may behave differently from one with strongly differentiated subscales, affecting both prevalence estimates and intervention sensitivity.

Language and cultural context further condition measurement performance. A Spanish-language translation and validation were reported for a Hispanic sample in early work (19), and subsequent evidence has raised questions about construct validity patterns in Hispanic populations, including the degree to which the three domains are empirically distinguishable (20). Even when translation quality is strong, psychometric behaviour can shift across populations because symptom expression, response styles, and contextual stressors vary. Contemporary reporting standards underline the need to document structural validity, reliability, and related properties transparently (21). Medical students in Latin American settings constitute an important case: training conditions, socioeconomic pressures, and institutional support structures may differ from those in settings where the scale has been most frequently studied, and these differences can alter item functioning and factor structure.

Objective

The objective of this psychometric evaluation was to examine the structural validity and internal consistency of the Spanish-language DASS-21 in a sample of undergraduate medical students drawn from a Latin American university setting (institutional identifiers blinded).

Research question

To what extent does the Spanish-language DASS-21 demonstrate (a) an interpretable latent structure consistent with its theoretical subscales and (b) satisfactory internal consistency in undergraduate medical students within a Latin American training context?

Analytic expectations

Based on prior evidence, it was expected that: (i) a three-factor representation would provide an acceptable account of item covariation in this population, while acknowledging competing hierarchical accounts in the literature; and (ii) reliability indices for the total scale and subscales would fall within ranges considered adequate for research and screening applications (12,13,15).

MATERIALS AND METHODS

Study Design and Setting

A cross-sectional, instrumentally psychometric study was conducted in 2018 within an undergraduate medical programme at a higher-education institution. Enrolment records for the year indicated a total population of 3 480 medical students distributed across six academic cohorts (first to sixth year).

Participants and Sampling

A non-probabilistic quota sampling strategy was used to recruit students across cohorts. The final sample comprised 229 undergraduate medical students: first year (n = 53), second year (n = 49), third year (n = 45), fourth year (n = 31), fifth year (n = 28), and sixth year (n = 23). Eligibility criteria were: (i) being an adult student, (ii) being enrolled in the medical programme, and (iii) agreeing to participate; students of both sexes were eligible.

Measure: Depression Anxiety Stress Scales—21 (DASS-21)

Symptoms of depression, anxiety, and stress were assessed using the 21-item version of the Depression Anxiety Stress Scales (DASS-21 (11). The DASS-21 comprises three 7-item subscales (Depression, Anxiety, Stress) and is rated on a 4-point response scale ranging from 0 ("did not apply to me at all") to 3 ("applied to me very much or most of the time"), referring to experiences during the previous week (11). Subscale scores are obtained by summing item responses within each domain; higher scores reflect greater symptom severity (12,13). Prior evidence supports its internal consistency and construct validity in clinical and non-clinical samples (12,13), including Spanish-language validation work (19).

Translation and Cultural Adaptation Procedure

A Spanish-language version was prepared through a multistep procedure consistent with widely used test-translation and adaptation frameworks (22,23). Two independent forward translations were produced by physicians (in blinded roles), and English-language specialists produced two independent translations. A reconciled Spanish draft was agreed by consensus and then back-translated into English. An expert review panel—two professors of psychiatry, one associate professor of English language, and one assistant professor of clinical psychology—evaluated semantic, conceptual, and cultural equivalence against the source instrument. The original instructions, item format, and response options were retained. A pilot test with 20 medical students was conducted to confirm the comprehensibility and feasibility of administration.

Data Collection Procedures

The questionnaire was administered online using a web-based form. The survey link was distributed to students via institutional email. Participants provided informed consent electronically before completing the questionnaire. A printed consent form was additionally made available through the faculty for participants who preferred a paper record. Responses were collected anonymously and used exclusively for research purposes.

Statistical Analysis

Descriptive statistics were computed (mean and standard deviation). Normality of score distributions was assessed using the Kolmogorov–Smirnov test (24). Internal consistency was examined using Cronbach's alpha (25), interpreted with attention to its dependence on scale length and dimensionality (26). Corrected item–total correlations were inspected as complementary evidence during scale evaluation (27).

Construct validity was examined through exploratory and confirmatory approaches. Exploratory analyses were performed using a principal components extraction strategy, preceded by checks of factorability (Kaiser–Meyer–Olkin and Bartlett's test of sphericity). Factor retention was based on eigenvalues greater than 1.0 and solution interpretability (28). Rotations included an oblique oblimin solution (Kaiser normalisation) and an orthogonal varimax solution (29) to aid interpretation, including cohort-wise inspection via split-file procedures. Salient loadings were evaluated using a minimum loading criterion of 0.30 (30,31).

Confirmatory factor analysis (CFA) was subsequently used to evaluate the latent structure within a structural equation modelling framework, using standard goodness-of-fit indices (32) and estimation approaches commonly recommended for ordinal indicators (33). Statistical significance

was set at $p < 0.05$. Analyses were conducted using IBM SPSS Statistics (version 20.0).

Ethical Considerations

This study was approved by the Endocrine and Metabolic Diseases Research Center's Bioethics Committee. Participants provided informed consent electronically before completing the online questionnaire.

RESULTS

Sample characteristics and scale scores

The analytical plan combined descriptive statistics with internal consistency, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and additional validity checks. A total of 229 medical students participated (54.6 % women), with a mean age of 21 (2) years and representation across all six academic years (Table 1). The mean total score for the DASS-21 was 16.03 (10.92); subscale means were 4.74 (4.70) for depression, 4.55 (3.79) for anxiety, and 6.74 (4.05) for stress.

Table 1. Sociodemographic profile of the sample (N = 229)

Variable	Category/statistic	n (%) / Mean (SD)
Sex	Female	125 (54.6)
	Male	104 (45.4)
Age (years)	Mean (SD)	21 (2)
	First year	53 (23.1)
	Second year	49 (21.4)
Academic year	Third year	45 (19.7)
	Fourth year	31 (13.5)
	Fifth year	28 (12.2)
	Sixth year	23 (10.0)
	Total score	16.03 (10.92)
DASS-21	Depression	4.74 (4.70)
	Anxiety	4.55 (3.79)
	Stress	6.74 (4.05)

Source: Own elaboration

Internal consistency

Internal consistency for the full DASS-21 scale reached $\alpha = 0.924$, with subscale coefficients of $\alpha = 0.883$ (depression), $\alpha = 0.789$ (anxiety), and $\alpha = 0.838$ (stress). The additivity assumptions for the depression and anxiety subscales were met, while the stress subscale showed a deviation from additivity. From a reporting standpoint, Cronbach's α is widely used for internal consistency, while complementary coefficients (e.g., ω) are often preferred when tau-equivalence is doubtful.

Exploratory factor analysis

Factorability was supported by a KMO of .908 and Bartlett's test of sphericity. Communalities ranged from 0.331 to 0.814. A three-factor solution was retained, accounting for 56.115 % of the total variance. Several items displayed meaningful cross-loadings, particularly Item 5 (loading on stress while theoretically aligned with depression) and Items 8, 11, and 14 (loading on anxiety while theoretically aligned with stress).

Table 2. Summary of cross-loading items identified in EFA

Item	Theoretically aligned subscale	Observed pattern in EFA
Item 5	Depression	Loaded under stress
Item 8	Stress	Loaded with anxiety
Item 11	Stress	Loaded with anxiety
Item 14	Stress	Loaded with anxiety

Source: Own elaboration

Confirmatory factor analysis and model comparison

CFA tested four hierarchical structures using a WLSMV estimator and compared competing specifications (Models 1–4). Model fit was judged through χ^2/df, RMSEA, SRMR, CFI, TLI, GFI, and AGFI (Table 2). Fit-index interpretation commonly relies on established guidance (e.g., prioritising incremental fit indices and residual-based indices alongside RMSEA).

Models 1–3 showed comparatively weaker fit profiles, with RMSEA values ranging from 0.089 to 0.119 and CFI values around 0.931–0.943 (Table 2). Model 4 improved the overall pattern of fit while enforcing a more conservative retention rule, keeping items with standardised loadings ≥ 0.50.

In Model 4, fit indices were: $\chi^2 = 133.59$, df = 51, χ^2/df = 2.62, RMSEA = 0.084 (LCI = 0.067; UCI = 0.102), SRMR = 0.068, CFI = 0.985, TLI = 0.980, GFI = 0.994, and AGFI = 0.989 (Table 3).

Table 3. Confirmatory factor analysis (Models 1–4)

Model	χ^2	df	p	χ^2/df	RMSEA	LCI	UCI	p ≤ 0.05	SRMR	CFI	TLI	GFI	AGFI
Model 1	531.97	186	0.004	2.86	0.089	0.080	0.098	0.0001	0.860	0.943	0.935	0.940	0.926
Model 2	665.49	157	0.002	4.24	0.119	0.110	0.128	0.0001	0.078	0.932	0.923	0.933	0.916
Model 3	640.37	187	0.001	3.41	0.104	0.095	0.112	0.0001	0.066	0.931	0.920	0.935	0.920
Model 4	133.59	51	<0.001	2.62	0.084	0.067	0.102	0.001	0.068	0.985	0.980	0.994	0.989

Source: Own elaboration.

Convergent validity

Convergent validity was assessed through AVE alongside internal consistency coefficients (α and ω) across the competing models. In Model 4, AVE values were 0.582 (stress), 0.552 (depression), and 0.504 (anxiety), accompanied by α values of 0.89, 0.88, and 0.88, and ω values of 0.90, 0.88, and 0.89, respectively.

Discriminant validity

Discriminant validity was examined using both the Fornell–Larcker criterion and the HTMT ratio. In Model 4, the Fornell–Larcker matrix (Table 4) and the HTMT values (Table 5) supported separation among constructs.

Measurement invariance by sex (Model 4)

Measurement invariance was tested across sex groups through sequential constraints (configural, threshold, metric, scalar, residual). Table 6 and 7 report fit indices and the incremental changes (ΔCFI, ΔRMSEA, ΔSRMR). The observed changes were small across steps, consistent with common invariance decision rules that focus on stability of CFI and RMSEA rather than χ^2 difference testing alone.

Table 4. Convergent validity indices (Models 1–4)

Model	Factor	AVE	α	ω
Model 1	Depression	0.578	0.88	0.88
	Anxiety	0.599	0.84	0.85
	Stress	0.459	0.85	0.85
Model 2	Depression	0.553	0.89	0.89
	Anxiety	0.605	0.83	0.83
	Stress	0.900	0.87	0.89
Model 3	Depression	0.578	0.81	0.81
	Anxiety	0.713	0.86	0.87
	Stress	0.716	0.85	0.87
Model 4	Stress	0.582	0.89	0.90
	Depression	0.552	0.88	0.88
	Anxiety	0.504	0.88	0.89

Source: Own elaboration
Note: (Where ω was used, it provides a reliability estimate less dependent on restrictive assumptions than α in many practical settings.)

Table 5. Fornell–Larcker criterion (Model 4).

	F1	F2	F3
F1	0.620		
F2	0.404	0.830	
F3	0.407	0.225	0.900

Source: Own elaboration

Table 6. HTMT ratio (Model 4)

Pair	HTMT
F1–F2	0.60
F1–F3	0.64
F2–F3	0.49

Source: Own elaboration
Note: HTMT is frequently recommended as a sensitive diagnostic for discriminant validity.

Table 7. Measurement invariance by sex (Model 4)

Model	χ^2	df	RMSEA	SRMR	CFI	ΔCFI	ΔRMSEA	ΔSRMR
Configural	247.1	102	0.087	0.074	0.987	—	—	—
Threshold	270.4	117	0.076	0.075	0.989	0.002	-0.011	0.001
Metric	259.0	114	0.075	0.075	0.990	0.001	-0.001	0.000
Scalar	269.8	117	0.076	0.075	0.989	-0.001	0.001	0.000
Residual	282.1	126	0.074	0.075	0.989	0.000	-0.002	0.000

Source: Own elaboration.

DISCUSSION

This psychometric evaluation set out to clarify two linked issues: whether the Spanish-language DASS-21 yields (i) an interpretable latent structure that corresponds to its depression, anxiety, and stress subscales and (ii) reliability levels adequate for research and symptom screening/monitoring use in undergraduate medical education.

Internal consistency and score coherence

At the scale level, internal consistency was strong (α = 0.924), placing the total score in the range commonly reported for the short-form DASS in non-clinical and student samples (13,34). In subscale terms, coefficients were also acceptable to high (α = 0.897 for depression; α = 0.789 for anxiety; α = 0.838 for stress). The anxiety coefficient is the lowest of the three, yet it remains suitable for symptom screening and group comparisons when the measure is treated as dimensional rather than diagnostic (11,12). A more nuanced point emerges from the item-level diagnostics: the stress subscale was flagged as deviating from strict additivity, which can occur when items are not equally strong indicators of the same latent trait or when the construct has a heterogeneous symptom content.

Given long-standing concerns about relying exclusively on α under violations of tau-equivalence, the inclusion of McDonald's ω is an asset, because ω is less restrictive in its assumptions and is often preferable for congeneric scales (35). In this evaluation, ω coefficients for the retained CFA model were high (ω = 0.880 for depression; ω = 0.901 for anxiety; ω = 0.897 for stress), reinforcing the interpretation that observed score consistency reflects substantive coherence rather than an artefact of a single coefficient choice.

Latent structure: three domains, with predictable points of overlap

The exploratory solution supported factor extraction and yielded a tripartite configuration aligned with the intended DASS domains. Sampling adequacy was strong (KMO = 0.908; Bartlett's test $p < 0.001$), supporting the suitability of the correlation matrix for factor analysis. Three factors accounted for 56.115 % of the variance, consistent with an interpretable multidimensional structure for a brief affective symptom scale in community-like samples.

At the same time, several items displayed cross-loadings (items 5, 8, 11, and 14), and item 5 clustered with stress rather than depression. This pattern is not anomalous in the DASS literature. The DASS-21 was designed to differentiate symptom clusters within a shared negative-affect space, and even in large normative datasets, domain factors often share variance through a broader distress dimension (11,13). Theoretical overlap is particularly plausible in medical student cohorts, where fatigue, tension, worry, and low mood can co-occur in response to sustained workload, sleep disruption, and high-stakes assessment conditions (3,7).

The confirmatory modelling sequence adds detail to that interpretation. The initial CFA models (Models 1–3) produced comparatively weaker fit (e.g., CFI $\approx$ 0.931–0.943; RMSEA $\approx$ .089–0.119), whereas the retained specification (Model 4) achieved markedly better incremental fit (CFI = 0.985; TLI = 0.980; SRMR = 0.068; RMSEA = 0.084).

The RMSEA estimate remains above the most conservative "close-fit" thresholds sometimes cited for simple CFA models, yet that index is sensitive to model constraints, degrees of freedom, and estimation under ordinal indicators; over-reliance on a single cut-off can be misleading, particularly when CFI/TLI and SRMR converge on an acceptable solution (32,33). The modelling strategy adopted here — contrasting alternative structures and retaining a specification with stronger item loadings — aligns with contemporary expectations that structural validity evidence should be presented transparently and triangulated across indices rather than anchored to a single statistic (21).

The broader literature helps contextualise the model comparison. Many studies support a correlated three-factor structure, whereas others argue that hierarchical or bifactor models better represent shared distress variance across narrower domains (13-15). Within that debate, the present pattern — three interpretable factors with some cross-loading items and relatively high interfactor associations — sits comfortably with the view that the DASS-21 can be used meaningfully at both the subscale and total-score level, provided the analyst recognises that subscales may not function as sharply separated diagnostic proxies (11,13,34).

Convergent and discriminant signals within the measurement model

Beyond factor fit, convergent validity checks for the retained model also met standard expectations. Average variance extracted exceeded 0.50 for each domain (depression AVE = 0.582; anxiety AVE = 0.504; stress AVE = 0.552), and composite reliability indices were high (CR ≈ 0.878–0.898). These values support the claim that the retained items capture their intended latent constructs with a satisfactory signal-to-noise ratio, a particularly relevant point for short scales that trade breadth for brevity (21,36).

Discriminant evidence was also favourable. Both the Fornell–Larcker check and HTMT ratios supported separability of the three domains (HTMT range = 0.49–0.64). Although depression, anxiety, and stress are expected to correlate, HTMT values comfortably below commonly used decision bands are compatible with a model in which domains remain distinguishable enough for subscale interpretation in research and screening contexts (37). The practical implication is that the DASS-21 can be used to profile symptom emphasis across domains, while still acknowledging that a substantial portion of variance is shared across affective dimensions in this population.

Measurement invariance by sex: comparability of score meaning

A key applied question in medical education research is whether subgroup comparisons reflect true differences in symptom severity or artefacts of measurement non-equivalence. Invariance testing across sex groups supported the stability of the retained model under sequential constraints (configural, threshold, metric, scalar, residual), with minimal changes in fit indices (e.g., ΔCFI values within ± 0.002).

This degree of stability falls within widely used invariance decision rules that prioritise incremental fit over χ^2 difference testing alone, particularly in moderate sample sizes (38,39). With scalar invariance supported, comparisons of latent means (and, pragmatically, observed scores) between women and men in similar cohorts are less likely to be confounded by differential item functioning, strengthening the measure's utility for equity-focused monitoring and programme evaluation in medical training.

Educational and clinical relevance in a high-burden context

The descriptive profile of the cohort offers a clinical-educational context for interpreting measurement performance. The sample comprised 229 medical students across years 1-6 (54.6 % women; mean age 21 ± 2), and mean symptom scores were non-trivial (total DASS-21 mean 16.03 ± 10.92; depression 4.74 ± 4.70; anxiety 4.55 ± 3.79; stress 6.74 ± 4.05). While mean scores are not directly comparable to prevalence estimates from diagnostic meta-analyses, they sit within an international evidence landscape in which distress during medical training is common: pooled estimates for depressive symptoms and anxiety often exceed one quarter to one third of students, and burnout is also frequent in syntheses (3-7).

In that setting, a brief tool that performs consistently and comparably across key subgroups can play several roles: (i) symptom screening to support early identification pathways (not diagnosis), (ii) monitoring across academic cycles, and (iii) evaluation of learning-environment interventions intended to improve well-being. Evidence syntheses suggest that curricular and environmental changes (e.g., assessment structures, mentoring, structured support programmes) can influence student well-being, even if effect sizes vary and primary studies remain heterogeneous (8,9).

In practical terms, the psychometric performance documented here positions the Spanish-language DASS-21 as a defensible option for these uses in comparable Latin American medical education contexts, with the additional advantage of established interpretability as a dimensional symptom measure rather than a diagnostic test (11,12).

LIMITATIONS

Several constraints temper interpretation and generalizability. First, the design is cross-sectional and relies on a single administration of a

self-report questionnaire. This structure supports internal consistency and structural validity checks, yet it does not address temporal stability. Without test–retest evidence, the reliability of scores over time and their sensitivity to academic-cycle fluctuations remain unresolved, which is important in medical education settings where stress exposure can vary sharply across semesters and clinical transitions.

Second, sampling was non-probabilistic and drawn from a single medical school context, which limits external validity. Even when internal structure is robust, factor solutions and item functioning can vary across institutions, regions, and socioeconomic contexts, particularly in Latin American settings, where training conditions and support infrastructure may differ across universities (19,20). Replication across multiple programmes and across public/private institutional environments would clarify whether the cross-loading items observed here remain the same pressure points elsewhere.

Third, the analytic approach in the exploratory phase used principal components extraction, which is common in applied settings but is not equivalent to common-factor modelling and can influence the distribution of variance across factors. Alongside the decision to retain only items with loadings ≥ 0.50 in the retained confirmatory specification, this may prioritise a cleaner structure at the cost of content coverage for some facets within each domain. The retained model's fit and reliability are strong enough to justify its use, yet different extraction/rotation choices or modern hybrid approaches (e.g., ESEM or bifactor-ESEM) could yield alternative representations of the same item set and may help disentangle domain-specific variance from a broad distress factor (15,16)

Fourth, the validity evidence here is predominantly internal (structure, reliability, convergent/discriminant checks within the measurement model, and sex invariance). Participants completed the DASS-21 as the focal measure, which limits the scope for external convergent, criterion, or diagnostic validation against established symptom inventories or clinical interviews. Accordingly, the present findings should be interpreted as evidence of internal validity and measurement comparability rather than external or diagnostic validity. In applied screening contexts, external anchoring is valuable because it supports interpretation of score ranges and clarifies how subscale elevations map onto clinically meaningful impairment (12,34). Finally, invariance was assessed across sex; however, other meaningful groupings in medical education — academic year, preclinical versus clinical stage, work/commute burden, and financial stressors — may also shape item interpretation and warrant examination when data permit (17,40).

Author Contributions: Conceptualisation, (M.C.-C.), (D.A.), (M.G.), (J.P.), (P.R.), (M.S.), (W.T.), (J.H.-L.), (V.B.), and (D.R.-P.); methodology, (M.C.-C.), (J.H.-L.), (V.B.), and (D.R.-P.); software, (J.H.-L.), (V.B.), and (D.R.-P.); validation, (M.C.-C.), (J.H.-L.), (V.B.), and (D.R.-P.); formal analysis, (M.C.-C.), (J.H.-L.), (V.B.), and (D.R.-P.); investigation, (M.C.-C.), (D.A.), (M.G.), (J.P.), (P.R.), (M.S.), (W.T.), (V.B.), and (D.R.-P.); resources, (M.C.-C.), (V.B.), and (D.R.-P.); data curation, (M.C.-C.), (D.A.), (M.G.), (J.P.), (P.R.), (M.S.), (W.T.), and (J.H.-L.); writing — original draft preparation, (M.C.-C.), (D.A.), (M.G.), (J.P.), (P.R.), (M.S.), (W.T.), (J.H.-L.), (V.B.), and (D.R.-P.); writing — review and editing, (M.C.-C.), (D.A.), (M.G.), (J.P.), (P.R.), (M.S.), (W.T.), (J.H.-L.), (V.B.), and (D.R.-P.); visualisation, (M.C.-C.), (J.H.-L.), (V.B.), and (D.R.-P.); supervision, (V.B.) and (D.R.-P.); project administration, (M.C.-C.), (V.B.), and (D.R.-P.); funding acquisition, (V.B.) and (D.R.-P.). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Ethical Considerations: This study was approved by the Endocrine and Metabolic Diseases Research Center's Bioethics Committee. Participants provided informed consent electronically prior to completing the online questionnaire.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1204-1222.
2. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet Lond Engl. 2021;398(10312):1700-1712.
3. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, et al. Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students. JAMA. 2016;316(21):2214-2236.
4. Puthran R, Zhang MWB, Tam WW, Ho RC. Prevalence of depression amongst medical students: A meta-analysis. Med Educ. 2016;50(4):456-468.
5. Tian-Ci Quek T, Wai-San Tam W, X. Tran B, Zhang M, Zhang Z, Su-Hui Ho C, et al. The Global Prevalence of Anxiety Among Medical Students: A Meta-Analysis. Int J Environ Res Public Health. 2019;16(15):2735.
6. Almutairi H, Alsubaiei A, Abduljawad S, Alshatti A, Fekih-Romdhane F, Husni M, et al. Prevalence of burnout in medical students: A systematic review and meta-analysis. Int J Soc Psychiatry. 2022;68(6):1157-1170.
7. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic Review of Depression, Anxiety, and Other Indicators of Psychological Distress Among U.S. and Canadian Medical Students. Acad Med. 2006;81(4):354.
8. Wasson LT, Cusmano A, Meli L, Louh I, Falzon L, Hampsey M, et al. Association Between Learning Environment Interventions and Medical Student Well-being: A Systematic Review. JAMA. 2016;316(21):2237-2252.
9. Bennett-Weston A, Keshtkar L, Jones M, Sanders C, Lewis C, Nockels K, et al. Interventions to promote medical student well-being: An overview of systematic reviews. BMJ Open. 2024;14(5):e082910.
10. Tibbetts Y, Himmelberger ZM, Barron KE, Speicher MR, Hulleman CS. Learning Mindsets and Well-Being and Ill-Being Among Osteopathic Medical Students. JAMA Netw Open. 2024;7(6):e2418090.
11. Lovibond PF, Lovibond SH. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther. 1995;33(3):335-343.
12. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assess. 1998;10(2):176-181.
13. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. Br J Clin Psychol. 2005;44(Pt 2):227-239.
14. Vignola RCB, Tucci AM. Adaptation and validation of the depression, anxiety and stress scale (DASS) to Brazilian Portuguese. J Affect Disord. 2014;155:104-109.
15. Yılmaz Koğar E, Koğar H. Using a bifactor exploratory structural equation modeling framework to examine the factor structure of the Depression Anxiety and Stress Scales-21. Curr Psychol. 2023;42(29):25807-25821.
16. Gomez R, Stavropoulos V, Griffiths MD. Confirmatory factor analysis and exploratory structural equation modelling of the factor structure of the Depression Anxiety and Stress Scales-21. PloS One. 2020;15(6):e0233998.
17. Lu S, Hu S, Guan Y, Xiao J, Cai D, Gao Z, et al. Measurement Invariance of the Depression Anxiety Stress Scales-21 Across Gender in a Sample of Chinese University Students. Front Psychol. 2018;9.
18. Escalas de Depresión, Ansiedad y Estrés (DASS-21): Estructura factorial, fiabilidad, invariancia y validez de la versión catalana. Available from: https://www.researchgate.net/publication/384849093_Depression_Anxiety_and_Stress_Scales_DASS21_Factor_structure_reliability_invariance_and_validity_of_the_Catalan_Version
19. Daza P, Novy DM, Stanley MA, Averill P. The Depression Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. J Psychopathol Behav Assess. 2002;24(3):195-205.
20. González-Rivera JA, Pagán-Torres OM, Pérez-Torres EM. Depression, Anxiety and Stress Scales (DASS-21): Construct Validity Problem in Hispanics. Eur J Investig Health Psychol Educ. 2020;10(1):375-389.
21. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197-2218.
22. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. Spine. 2000;25(24):3186.
23. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. J Eval Clin Pract. 2011;17(2):268-274.
24. Massey FJ. The Kolmogorov-Smirnov Test for Goodness of Fit. J Am Stat Assoc. 1951;46(253):68-78.

25. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297-334.
26. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol. 1993;78(1):98-104.
27. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. Psychol Assess. 1995;7(3):309-319.
28. Hayton JC, Allen DG, Scarpello V. Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. Organ Res Methods. 2004;7(2):191-205.
29. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23(3):187-200.
30. Yong AG, Pearce S. A Beginner's Guide to Factor Analysis Focusing on Exploratory Factor Analysis. Tutorials in Quantitative Methods for Psychology. 2013;9:79-94.
31. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluación del uso del análisis factorial exploratorio en la investigación psicológica. Psychol Methods. 1999;4:272-299.
32. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999;6(1):1-55.
33. Flora DB, Curran PJ. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. Psychol Methods. 2004;9(4):466-491.
34. Osman A, Wong JL, Bagge CL, Freedenthal S, Gutierrez PM, Lozano G. The Depression Anxiety Stress Scales-21 (DASS-21): Further examination of dimensions, scale reliability, and correlates. J Clin Psychol. 2012;68(12):1322-1338.
35. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol Lond Engl 1953. 2014;105(3):399-412.
36. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J Mark Res. 1981;18(1):39.
37. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci. 2015;43(1):115-135.
38. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Struct Equ Model Multidiscip J. 2002;9(2):233-255.
39. Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Struct Equ Model Multidiscip J. 2007;14(3):464-504.
40. Malas O, Tolsá MD. Depression, Anxiety and Stress Scales (DASS-21): Factor structure, reliability, invariance and validity of the Catalan Version. Ann Psicol UB J Psychol. 2022;52(3).