Research paper

Cross-cultural validation and measurement invariance of anxiety and depression symptoms: A study of the Brief Symptom Inventory (BSI) in 42 countries

Gonzalo R. Quintana [a,*], Fernando P. Ponce [b], Javier I. Escudero-Pastén [a], Juan F. Santibáñez-Palma [a], Léna Nagy [c,d], Mónika Koós [d], Shane W. Kraus [e], Zsolt Demetrovics [d,f], Marc N. Potenza [g,h], Rafael Ballester-Arnal [i], Dominik Batthyány [j], Sophie Bergeron [k], Joël Billieux [l,m], Peer Briken [n], Julius Burkauskas [o], Georgina Cárdenas-López [p], Joana Carvalho [q], Jesús Castro-Calvo [r], Lijun Chen [s], Giacomo Ciocca [t], Ornella Corazza [u,v], Rita I. Csako [w], David P. Fernandez [x], Elaine F. Fernandez [y], Hironobu Fujiwara [z,aa], Johannes Fuss [ab], Roman Gabrhelík [ac,ad], Ateret Gewirtz-Meydan [ae], Biljana Gjoneska [af], Mateusz Gola [ag,ah], Joshua B. Grubbs [ai,aj], Hashim T. Hashim [ak,al], Md. Saiful Islam [am,an], Mustafa Ismail [ak], Martha C. Jiménez-Martínez [ao,ap], Tanja Jurin [aq], Ondrej Kalina [ar], Verena Klein [as], András Költő [at], Chih-Ting Lee [au], Sang-Kyu Lee [av,aw], Karol Lewczuk [ax], Chung-Ying Lin [ay,az], Christine Lochner [ba], Silvia López-Alvarado [bb], Kateřina Lukavská [ac,bc], Percy Mayta-Tristán [bd], Dan J. Miller [be], Oľga Orosová [ar], Gábor Orosz [bf], Sungkyunkwan University's research team NA[bg,1], Gabriel C. Quintero Garzola [bh,bi], Jano Ramos-Diaz [bj], Kévin Rigaud [bf], Ann Rousseau [bk], Marco De Tubino Scanavino [bl,bm], Marion K. Schulmeyer [bn], Pratap Sharan [bo], Mami Shibata [z], Sheikh Shoib [bp,bq], Vera Sigre-Leirós [l], Luke Sniewski [br], Ognen Spasovski [bs,bt], Vesta Steibliene [bu], Dan J. Stein [bv], Berk C. Ünsal [c,d], Marie-Pier Vaillancourt-Morel [bw], Marie Claire Van Hout [bx], Beáta Bőthe [by,bz]

[a] Departamento de Psicología y Filosofía, Facultad de Ciencias Sociales, Universidad de Tarapacá, Arica, Arica y Parinacota, Chile
[b] Facultad de Psicología, Universidad de Talca, Chile
[c] Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
[d] Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
[e] Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV, USA
[f] Centre of Excellence in Responsible Gaming, University of Gibraltar, Gibraltar, Gibraltar
[g] Yale University School of Medicine, New Haven, CT, USA
[h] Connecticut Council on Problem Gambling, Wethersfield, CT, USA
[i] Departamento de Psicología Básica, Clínica y Psicobiología, University Jaume I of Castellón, Spain
[j] Institute for Behavioural Addictions, Sigmund Freud University Vienna, Austria
[k] Département de psychologie, Université de Montréal, Montréal, Canada
[l] Institute of Psychology, University of Lausanne, Lausanne, Switzerland
[m] Center for Excessive Gambling, Addiction Medicine, Lausanne University Hospitals (CHUV), Lausanne, Switzerland
[n] Institute for Sex Research, Sexual Medicine, and Forensic Psychiatry, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
[o] Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health Sciences, Palanga, Lithuania
[p] Virtual Teaching and Cyberpsychology Laboratory, School of Psychology, National Autonomous University of Mexico, Mexico
[q] William James Center for Research, Departamento de Educaçao e Psicologia, Universidade de Aveiro, Aveiro, Portugal
[r] Department of Personality, Assessment, and Psychological Treatments, University of Valencia, Spain
[s] Department of Psychology, College of Humanity and Social Science, Fuzhou University, China
[t] Section of Sexual Psychopathology, Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, Rome, Italy
[u] Department of Clinical, Pharmaceutical and Biological Sciences, University of Hertfordshire, United Kingdom
[v] Department of Psychology and Cognitive Science, University of Trento, Italy
[w] Department of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand

* Corresponding author at: Universidad de Tarapacá, 2222, 18 de Septiembre Av., PO 1000007, Arica, Arica y Parinacota, Chile.
E-mail address: gquintanaz@academicos.uta.cl (G.R. Quintana).

[x] *Nottingham Trent University, United Kingdom*
[y] *HELP University, Malaysia*
[z] *Department of Neuropsychiatry, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
[aa] *Decentralized Big Data Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan*
[ab] *Institute of Forensic Psychiatry and Sex Research, Center for Translational Neuro- and Behavioral Sciences, University of Duisburg-Essen, Essen, Germany*
[ac] *Charles University, First Faculty of Medicine, Department of Addictology, Prague, Czech Republic*
[ad] *General University Hospital in Prague, Department of Addictology, Czech Republic*
[ae] *School of Social Work, Faculty of Social Welfare and Health Sciences, University of Haifa, Israel*
[af] *Macedonian Academy of Sciences and Arts, North Macedonia*
[ag] *Institute of Psychology, Polish Academy of Sciences, Poland*
[ah] *Institute for Neural Computations, University of California San Diego, USA*
[ai] *Center on Alcohol, Substance use, Addictions University of New Mexico, Albuquerque, USA*
[aj] *Department of Psychology, University of New Mexico, Albuquerque, USA*
[ak] *University of Baghdad, College of Medicine, Iraq*
[al] *University of Warith Al-anbya, College of Medicine, Karbala, Iraq*
[am] *Department of Public Health and Informatics, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh*
[an] *Centre for Advanced Research Excellence in Public Health, Savar, Dhaka 1342, Bangladesh*
[ao] *Universidad Pedagógca y Tecnológica de Colombia, Colombia*
[ap] *Grupo de Investigación Biomédica y de Patología, Colombia*
[aq] *Department of Psychology, Humanities and Social Sciences, University of Zagreb, Croatia*
[ar] *Department of Educational Psychology and Psychology of Health, Pavol Jozef Safarik University in Kosice, Slovakia*
[as] *School of Psychology, University of Southampton, United Kingdom*
[at] *Health Promotion Research Centre, University of Galway, Ireland*
[au] *Department of Family Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan*
[av] *Department of Psychiatry, Hallym University Chuncheon Sacred Heart Hospital, South Korea*
[aw] *Chuncheon Addiction Management Center, South Korea*
[ax] *Institute of Psychology, Cardinal Stefan Wyszynski University, Warsaw, Poland*
[ay] *Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan*
[az] *Biostatistics Consulting Center, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan*
[ba] *SAMRC Unit on Risk & Resilience in Mental Disorders, Stellenbosch University, South Africa*
[bb] *University of Cuenca, Ecuador*
[bc] *Charles University, Faculty of Education, Department of Psychology, Prague, Czech Republic*
[bd] *Facultad de Medicina, Universidad Científica del Sur, Lima, Peru*
[be] *College of Healthcare Sciences, James Cook University, Australia*
[bf] *Artois University, France*
[bg] *Department of Psychology, Sungkyunkwan University, South Korea*
[bh] *Florida State University, Republic of Panama*
[bi] *Sistema Nacional de Investigación (SNI), SENACYT, Panama*
[bj] *Facultad de Ciencias de la Salud, Universidad Privada del Norte, Lima, Peru*
[bk] *Leuven School For Mass Communication, KU Leuven, Leuven, Belgium*
[bl] *Department of Psychiatry, Schulich School of Medicine & Dentistry, Western University, St. Joseph's Health Care London and London Health Sciences Centre, London, Canada*
[bm] *Lawson Health Research Institute, London, Canada*
[bn] *Universidad Privada de Santa Cruz de la Sierra, Bolivia*
[bo] *Department of Psychiatry, All India Institute of Medical Sciences, New Delhi 110029, India*
[bp] *Department of Psychology, Shardha University, India*
[bq] *Department of Health Services, Srinagar 190001, India*
[br] *Compassionate Inquiry, Canada*
[bs] *Faculty of Philosophy, Ss. Cyril and Methodius University in Skopje, North Macedonia*
[bt] *Faculty of Philosophy, University of Ss. Cyril and Methodius in Trnava, Slovakia*
[bu] *Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health sciences, Palanga, Lithuania*
[bv] *SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry, Neuroscience Institute, University of Cape Town, South Africa*
[bw] *Département de Psychologie, Université du Québec à Trois-Rivières, Trois-Rivières, Canada*
[bx] *Public Health Institute, Faculty of Health, Liverpool John Moores University, United Kingdom*
[by] *Département de Psychologie, Université de Montréal, Montréal, Canada*
[bz] *Centre de recherche interdisciplinaire sur les problèmes conjugaux et les agressions sexuelles (CRIPCAS)*

ARTICLE INFO

ABSTRACT

Background: Depression and anxiety are among the most prevalent mental health issues experienced worldwide. However, whereas cross-cultural studies utilize psychometrically valid and reliable scales, fewer can meaningfully compare these conditions across different groups. To address this gap, the current study aimed to psychometrically assess the Brief Symptomatology Index (BSI) in 42 countries.

Methods: Using data from the International Sex Survey ($N = 82{,}243$; $M_{age} = 32.39$; $SD_{age} = 12.52$; women: $n = 46{,}874$; 57 %), we examined the reliability of depression and anxiety symptom scores of the BSI-18, as well as evaluated evidence of construct, invariance, and criterion-related validity in predicting clinically relevant variables across countries, languages, genders, and sexual orientations.

Results: Results corroborated an invariant, two-factor structure across all groups tested, exhibiting excellent reliability estimates for both subscales. The 'caseness' criterion effectively discriminated among those at low and high risk of depression and anxiety, yielding differential effects on the clinical criteria examined.

Limitations: The predictive validation was not made against a clinical diagnosis, and the full BSI-18 scale was not examined (excluding the somatization sub-dimension), limiting the validation scope of the BSI-18. Finally, the

[1] The Sungkyunkwan University's research team comprises Dr. H. Chang and Mr. K. Park.

study was conducted online, mainly by advertisements through social media, ultimately skewing our sample towards women, younger, and highly educated populations.

Conclusions: The results support that the BSI-12 is a valid and reliable assessment tool for assessing depression and anxiety symptoms across countries, languages, genders, and sexual orientations. Further, its caseness criterion can discriminate well between participants at high and low risk of depression and anxiety.

1. Introduction

Depression and anxiety are affective disorders characterized by abnormal emotional states. Diagnostic criteria in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013) and the International Classification of Diseases 11th Revision (ICD-11; WHO, 2022a) focus on specific symptoms. Depression is characterized by prolonged periods of melancholic feelings of grief or unhappiness, despair, and loss of pleasure or interest in activities for extended periods, whereas anxiety involves feelings of fear, dread, and uneasiness, often accompanied by physical sensations. These disorders are among the most common mental health conditions experienced worldwide, affecting approximately 300 million people globally (Saha et al., 2021; WHO, 2017, 2022b). However, these estimates are unequally distributed across groups of people (Steel et al., 2014).

These common mental disorders (CMD, defined as a range of anxiety and depressive disorders; Lund et al., 2010) may also impact people's lives differently depending on the intersectionality of various attributes and risk factors. Systematic reviews show that low- and middle-income countries experience a higher burden of CMDs, which may lead to multiple socio-economic stressors and challenges, thus perpetuating a vicious cycle (Kessler, 2007; Ridley et al., 2020). Among anxiety and depression vulnerable populations, women (Steel et al., 2014) and sexual/gender minorities (Wittgens et al., 2022) commonly experience depression and anxiety during their lifetime. For instance, once controlling for sociodemographic factors, gay and bisexual males exhibit a significantly higher risk of experiencing depression (OR = 2.91 and 2.41, respectively) in comparison to their heterosexual counterparts (Gonzales and Henning-Smith, 2017). Whereas these relations often replicate across studies, what people experience and express through their symptoms may be influenced by cultural factors (Noguera et al., 2009; Smirnova et al., 2018).

Cross-cultural and multi-national studies of depression and anxiety using self-report measures show differences across nations, cultures, and ethnicities (see Wiesner et al., 2010). However, it is generally assumed that people experience, express, and understand symptoms the same way across cultures, and that diagnostic/screening tools measure the same latent constructs.

The first assumption may not fully consider how cultural factors influence how people view, think, experience, and express their symptoms, as well as how the context and social norms may modulate or interact with these (Bredström, 2019; Hofmann and Hinton, 2014; Lewis-Fernández et al., 2010). Indeed, studies suggest that different words used as descriptors of symptoms differ in their meanings across languages, thus modulating the diagnosis of depression and anxiety disorders (Noguera et al., 2009; Smirnova et al., 2018). These differences may be diluted, under-recognized, or misidentified through cross-cultural/multi-national studies or even within diverse populations within the same study sample, with potentially serious implications, especially for under-represented populations (Kirmayer, 2001; Kirmayer and Groleau, 2001).

The second assumption, related to measurement invariance (Putnick and Bornstein, 2016), pertains to the capacity of any screening tool to measure—and therefore compare—the same latent construct, reflecting the same meaning (psychometrically speaking) across specific groups or repeated measures. Despite its significance, measurement invariance remains largely overlooked, even in validated instruments routinely administered to measure depression and anxiety symptoms (Akena

et al., 2012; Gilbody et al., 2001; Yeung et al., 2020). Several studies and systematic reviews highlight the importance of testing measurement invariance in cross-cultural and multi-national research when using psychometric scales (Gregorich, 2006; Lacko et al., 2022; Stevanovic et al., 2017; Uysal-Bozkir et al., 2013). This may be of particular interest when ascertaining what may be culture-independent or culture-specific dimensions when assessing depression and anxiety (Canino and Alegria, 2008).

Of many scales available to measure depression or anxiety, fewer have been designed to measure both simultaneously. The Brief Symptom Index (BSI-18) (Derogatis, 1982; Derogatis, 2000) is capable of measuring depression and anxiety symptoms (and somatization), as well as their severity. Whereas several studies have examined this scale's measurement invariance in patient populations (e.g., Hoe and Brekke, 2008; Wang et al., 2010), large cross-cultural assessments of its measurement invariance have not been conducted to date.

This study examined the psychometric properties of the depression and anxiety scores of the BSI-18. Specifically, the study examined the following aims: a) factor structure of the BSI in the entire sample; b) measurement invariance based on language, country, gender, and sexual orientation to ensure meaningful and unbiased subgroup comparisons; c) reliability and validity of the BSI through associations with theoretically relevant correlates; and d) feasibility of using previously established thresholds to identify participants at high risk of depression or anxiety.

2. Method

2.1. Procedure

The validation of the BSI was conducted within the framework of the "International Sex Survey,"[2] project, a cross-sectional international collaboration involving 42 countries.[3] Details regarding this study and the entire cross-sectional collaboration, such as procedures and eligibility criteria, are described in the project's study protocol (Bőthe et al., 2021) that was also pre-registered (https://osf.io/csyjq). The ISS, and therefore this study, was reviewed and approved by each institution's ethics committee board (see OSF registry of ethics approval https://osf.io/n3k2c), and conducted in accordance with the Helsinki Declaration, while every participant provided full consent before answering the survey. Furthermore, as part of the ISS project's transparency policy, readers may find a detailed description of the methods used and data cleaning process through its OSF registries (https://osf.io/dk78r).[4]

---

[2] Study website: https://internationalsexsurvey.org/

[3] The study protocol paper included Egypt, Iran, Pakistan, and Romania as collaborating countries. However, it was not possible to get ethical approval for the study in a timely manner in these countries. Chile was not included in the study protocol paper as a collaborating country as it joined the study after publishing the study protocol. Therefore, instead of the planned 45 countries (Bőthe et al., 2021), only 42 individual countries are considered in the present study (see details at https://osf.io/n3k2c/)

[4] Publications: https://osf.io/jb6ey/?view_only=0014d87bb2b54 6f7a2693543389b934d, Conference presentations: https://osf.io/c695n/?vie w_only=7cae32e642b54d049e600ceb8971053e

2.2. Participants

Data cleaning of the ISS project database yielded 82,243 participants (M_{age} = 32.39; *SD* = 12.52). Participants' sociodemographic characteristics are detailed in Table 1.

2.3. Measures

Translated versions of scales used here and on the ISS are available on the project's OSF page.[5]

2.3.1. Sociodemographic and health related variables

The survey included several sociodemographic and general questions about physical, sexual, and mental health, including specific questions about the psychological impact of the COVID-19 pandemic.

2.3.2. Brief Symptom Inventory (BSI-12)

The BSI is a self-report measure designed to screen for psychological distress and psychiatric disorders and their severities. Using factor analysis, the BSI-18 was developed to measure somatization, depression, and anxiety, each with six items (Derogatis, 1982; Derogatis and Fitzpatrick, 2004). In the current study, however, we only assessed the two latter (i.e., depression and anxiety) in accordance with the ISS' goals (Bőthe et al., 2021). Participants rated a series of symptoms using a 5-point Likert scale (0 = Not at all; 4 = Extremely) according to how much they disturbed respondents in the prior week. Furthermore, the scale provides the 'caseness' criterion (Derogatis, 2000) to distinguished between participants at high or low risk of depression and anxiety.

2.3.3. P4 suicidality risk screener

The P4 screener (Dube et al., 2010) is a questionnaire assessing suicide thoughts and risk through four screening questions regarding: past suicide attempts, a plan for suicide, the probability of completing suicide, and preventive factors. The scale was chosen for analytic purposes as a criterion variable for the 'caseness' criterion of the BSI scale, and for its widespread used to measure suicidality. The score was calculated by assigning and summing one point to the responses concerning suicidal ideation (see Fig. 1 in Dube et al., 2010).

2.3.4. Compulsive Personality Assessment Scale (CPAS)

The CPAS (Burkauskas and Fineberg, 2020; Fineberg et al., 2007) is an eight-item scale designed to assess compulsive personality traits including need for control and rigidity, among others. Respondents rated items using a 5-point Likert scale (0 = Not at all characteristic of me; 4 = Entirely characteristic of me). The scale was also chosen for analytic purposes as a criterion variable for the 'caseness' criterion of the BSI scale, and for its widespread used to measure compulsivity.

2.4. Analysis plan

All validation studies included in the ISS survey followed a pre-registered analysis protocol.[6] We first examined descriptive statistics, and explored for patterns in the missing data using all 82,243 participants. Out of the total, 6178 participants provided no answer in all BSI items, whereas 270 had at least one missing value. Nonetheless, missing values on the BSI-12 items were minimal (about 7.5 %) and not missing completely at random (MCAR), as per assessed through Little's MCAR Test (Little, 1988; $\chi 2$ = 512.170, df = 523, *p* = 0.623). Therefore, given the MCAR pattern results, the large sample size, and the ease of analyses, we employed a pairwise deletion of participants who provided no item response in the BSI, leaving a sample size of 76,065. Nevertheless, when conducting the criterion validity analyses, to improve the validity of this

procedure and its results and its applicability for other studies and potential clinical applications, we opted not to impute any missing values while excluding cases by listwise deletion, leaving a sample size of 75,795. This constituted 270 (< 0.01 %) cases less than the total sample used in other analyses based on the gender variable.

We used Confirmatory Factor Analysis (CFA) to estimate the two-correlated-factors model with items loading on one of the factors of anxiety or depression, consistent with the notion of multifaceted distress (Durá et al., 2006; Recklitis et al., 2017; Torres et al., 2013; Wiesner et al., 2010). Following the ISS project guidelines (Bőthe et al., 2021), and due to the ordinal nature of the BSI-12 items, as well as the lack of multivariate normality in the data detected through Mardia's test (Mardia, 1970) (b_{Skew} = 13.969; b_{Kurt} = 228.358; *ps* < 0.001), the Weighted Least Squares Means- and Variance-Adjusted (WLSMV) was used as the estimation method for CFA and invariance testing (Finney and DiStefano, 2013). The goodness of fit was evaluated using the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). For the CFI and TLI indices, estimated values above 0.90 and 0.95 indicate acceptable and good fit degrees, respectively (Hu and Bentler, 1999; Marsh et al., 2004). For the SRMR and RMSEA, values equal to or lower than 0.05 and 0.08 are considered good and acceptable, respectively (Browne and Cudeck, 1992; Hu and Bentler, 1999). All statistical analyses were conducted using Mplus v.8.2. (Muthén and Muthén, 2017). We also estimated reliability in terms of internal consistency in each subscale/factor, using both Cronbach's alpha (1951) and McDonald's omega (1970), respectively.[7]

Second, due to the large-scale and cross-cultural nature of the ISS study, we conducted measurement invariance to investigate the potential presence of structural differences in the BSI-12 based on participants' language, country, gender, and sexual orientation (van de Schoot et al., 2012; Vandenberg and Lance, 2000). Monte Carlo simulations[8] revealed that each group should have a minimum of 460 participants as the optimal sample size by level in each grouping variable examined. Based on this criterion, some groups were excluded from the analysis language (4 groups: n's ≤ 332) and country of residence (9 groups: n's ≤ 385; see Table 1). Alternately, for invariance analysis by gender and sexual orientation, some groups were collapsed into broader categories to maximize the number of participants studied (Borgogna et al., 2019; Feinstein et al., 2023; A description of the collapsed response options for the gender and sexual orientation variables can be found in Table 1). Further information regarding the rationale behind creating these groupings can be found in the ISS pre-registration document.

For measure invariance, we followed a specification sequence of analyses that successively incorporated more restrictive models by applying parameter constraints at the model structure (i.e., configural model) and factor loadings (i.e., metric model) levels. For these models, factors' means and variances were fixed to 0 and 1 in each group for identification, respectively. The equality of items' thresholds (four per item given five response options) and variances across groups were then examined by fitting first a scalar model with all residual variances freely estimated and then compared with a residual model in which all residual variances were fixed to be equal in all groups. Finally, we examined the structural invariance of the model at the level of the variances and covariance between anxiety and depression factors (i.e., latent variance-covariance invariance) and their respective means (i.e., latent mean

---

[7] The tau-equivalence assumption (i.e., equal factor loadings for all items in factor models) is required for alpha to be comparable to the reliability coefficient (Cronbach, 1951). If this assumption is violated (also referred to as congeneric models), the reliability value will be underestimated depending on the severity of the violation (Green & Yang, 2009). Here, we opted to focus on interpreting the omegas because it corrects the underestimation bias of alpha in congeneric models (Dunn et al., 2014; Revelle & Zinbarg, 2009).

[8] See details https://osf.io/dk78r

---

[5] https://osf.io/jcz96/?view_only=9af0068dde81488db54638a01c8ae118

[6] https://osf.io/dk78r

Table 1
Summary table of sociodemographic characteristics of the study population.

Country of residence		n	%			n	%
	Algeria[a]	24	<0.1		Austria	746	0.9
Africa	South Africa	1849	2.2		Belgium	644	0.8
	Bolivia[a]	385	0.5		Croatia	2390	2.9
	Brazil	3579	4.4		Czech Republic	1640	2.0
	Canada	2541	3.1		France	1706	2.1
	Chile	1173	1.4		Germany	3271	4.0
	Colombia	1913	2.3		Gibraltar[a]	64	<0.1
	Ecuador[a]	276	0.3		Hungary	11,200	13.6
	Mexico	2137	2.6		Ireland	1702	2.1
	Panama[a]	333	0.4		Italy	2401	2.9
	Peru	2672	3.2		Lithuania	2015	2.5
America	United States	2398	2.9		North Macedonia	1251	1.5
	Bangladesh[a]	373	0.5		Poland	9892	12.0
	China	2428	3.0		Portugal	2262	2.8
	India[a]	194	0.2		Slovakia	1134	1.4
	Iraq[a]	99	0.1		Spain	2327	2.8
	Israel	1334	1.6		Switzerland	1144	1.4
	Japan	562	0.7		Turkey	820	1.0
	Malaysia	1170	1.4	Europe	United Kingdom	1412	1.7
	South Korea	1464	1.8		Australia	639	0.8
Asia	Taiwan	2668	3.2	Oceania	New Zealand	2834	3.4

	n	%		n	%
Language					
Arabic[a]	142	0.2	Korean	1437	1.7
Bangla[a]	332	0.4	Lithuanian	2094	2.5
Croatian	2522	3.1	Macedonian	1301	1.6
Czech	1583	1.9	Mandarin – simplified	2474	3.0
Dutch	518	0.6	Mandarin – traditional	2685	3.3
English	13,994	17.0	Polish	10,343	12.6
French	3941	4.8	Portuguese – Brazil	3650	4.4
German	3494	4.2	Portuguese – Portugal	2277	2.8
Hebrew	1315	1.6	Slovak	2118	2.6
Hindi[a]	17	<0.1	Spanish – Latin America	8926	10.9
Hungarian	10,937	13.3	Spanish – Spain	2312	2.8
Italian	2437	3.0	Turkish	853	1.0
Japanese	466	0.6			
Religion					
Buddhist	1526	1.9	Sikh	33	<0.1
Christian	24,080	29.3	Spiritism	426	0.5
Confucianist	18	<0.1	Taoist	627	0.8
Hindu	246	0.3		11,330	13.8
Jain	14	<0.1	Spiritual but no committed to one religion		
Jewish	1288	1.6	I am not religious	39,331	47.8
Muslim	1097	1.3	Other	2166	2.6

	n	%		n	%
Current place of residence			**Highest level of education**		
Metropolis (>1 million people)	26,441	32.1	Primary (e.g., Elementary school)	1002	1.2
City (100,000–999,999 people)	29,920	36.4	Secondary (e.g., High school)	20,325	24.7
Town (1000–99,999 people)	21,103	25.7	Tertiary (e.g., College/University)	60,896	74.1
Village (< 1000 people)	4764	5.8			
Currently studying?			**Currently working?**		
No	49,802	60.6	No	20,853	23.4
Yes, in primary education	64	<0.1	Yes, full time	42,981	52.3
Yes, in secondary education	1571	1.9	Yes, part-time	11,356	13.8
Yes, in tertiary education	30,762	37.4	Yes, I do odd jobs	7029	8.5
Socio-economical status					
Among the worst	227	0.3	Better than average	31,567	38.4
Much worse than average	773	0.9	Much better than average	14,736	17.9
Worse than average	4232	5.1	Among the best	3957	4.8
Average	26,742	32.5			
Gender (original)[b]			**Gender (used)[c]**		
Masculine/man	32,549	39.6	Man	32,549	39.6
Feminine/woman	46,874	57.0	Woman	46,874	57.0
			Gender-diverse individuals	2783	3.4
Indigenous or other cultural gender minority identity (e.g., two-spirit)	166	0.2			
Non-binary, gender fluid, or something else (e.g., genderqueer)	2315	2.8			

(continued on next page)

Table 1 (*continued*)

	n	%		n	%
Other	302	0.4			
Trans status			**Relationship status**		
No, I am not a trans person	79,280	96.4	Single	27,541	33.5
Yes, I am a trans man	357	0.4	In a relationship	27,440	33.4
Yes, I am a trans woman	295	0.4	Married or common-law partners	24,338	29.6
Yes, I am a non-binary trans person	881	1.1	Widow or widower	428	0.5
I am questioning my gender	1137	1.4	Divorced	2472	3.0
I don't know what it means	269	0.3			
Sexual orientation (original)[b]			**Sexual orientation (used)**[c]		
Heterosexual/Straight	56,125	68.2	Heterosexual	56,125	68.2
Gay or lesbian or Homosexual	4607	5.6	Gay or lesbian or Homosexual	4607	5.6
Heteroflexible	6200	7.5	Bisexual	7688	9.3
Homoflexible	534	0.6	Queer or pansexual	2926	3.6
Bisexual	7688	9.3	Homo- or heteroflexible identity	6734	8.2
Queer	957	1.2	Asexual	1064	1.3
Pansexual	1969	2.4	Questioning	1951	2.4
Asexual	1064	1.3	Other	807	1.0
I don't know yet or I am currently questioning my sexual orientation	1951	2.4			
None of the above	807	1.0			
I don't want to answer	308	0.4			

Note. Percentages might not add up to 100 % due to missing data in each variable.

[a] This group was discarded from invariance analysis due to insufficient sample size.
[b] Original response options available in the survey.
[c] Collapsed response options used in the analysis.

invariance; Vandenberg and Lance, 2000). We used the WLSMV estimator for all models, including a probit link and the theta parameterization.

We did not inspect χ^2 differences because, given our large sample size, they are prone to overestimating slight misfits (Meade et al., 2008). Differences in the relative fit indices (i.e., CFI, TLI, RMSEA) were examined following the criteria defined by Chen (2007) for large sample sizes, where ΔCFI values ≤ 0.010 and ΔRMSEA values ≤ 0.015 indicate invariance between groups, but including a more liberal cutoff for the RMSEA (≤ 0.10). We also considered ΔRMSEA (≤ 0.050) and ΔCFI (≤ 0.020) indexes when using large sample size and multiple categories in measurement invariance tests (i.e., when there are at least ten groups in a two-factor model and a non-normal data distribution; Svetina and Rutkowski, 2017; Svetina et al., 2019). Finally, we included the change in TLI (i.e., ΔTLI) fit index as a control measure for parsimony, following Chen (2007)'s criterion (ΔTLI ≤ 0.010). Multiple comparisons across all levels of countries, languages, sexual orientations, and genders were performed using Kruskall Wallis' tests (see Supplementary Tables 4, 5, 6, and 7, respectively).

Finally, to assess the criterion-related validity of the BSI-12, we compared participants categorized as having a high or low risk for anxiety and depression on various criterion variables (i.e., physical, sexual, and mental health [see Table 3. for details]). To do this, following the standard case-rule scoring (cutoff point, ≥ 63 T-score for both scales; Derogatis, 2000) based on self-reported gender (i.e., men, women, and gender-diverse individuals) was used to identify cases with a high risk of anxiety or depression. Using this cutoff parameter, we compared participants in both groups (i.e., low- and high-risk groups) in each disorder along the aforementioned correlates using Mann-Whitney tests for continuous and ordinal variables and chi-square tests for categorical variables. For the Mann-Whitney tests, the r indicator, based on the statistic z (Cohen, 1988), was reported as effect size, where 0.10 was considered as small, 0.30 as a medium, and 0.50 or greater as large differences. Likewise, for chi-square tests, Cramer's V and Phi (1946) were reported as effect size indicators, where 0.20 was considered as weak, 0.30 as a moderate, and higher than 0.40 as strong associations (Cohen, 1988; Rea and Parker, 1992).

3. Results

3.1. Factor analysis and reliabilities

Standardized factor loadings and reliability indices of the two-factor model CFA can be seen in Supplemental Table 1. This model exhibited an excellent fit to the data (WLSMV χ^2 (53) = 23,070.197; $p < 0.001$; CFI = 0.984; TLI = 0.981; RMSEA = 0.076 [90 % CI = 0.075–0.076]). Standardized factor loadings ranged from 0.702 to 0.891 ($ps < 0.001$), whereas inter-factor correlation was positive and strong ($\Phi = 0.771$; $p < 0.001$).

The internal consistency indices were excellent, and Cronbach's alpha (α) and McDonald's omega (ω) were consistently equal or larger than 0.90.

3.2. BSI-12 measurement invariance analyses

Table 2 shows the results of the multigroup analysis of cross-cultural invariance of the BSI-12. Fit indices for the two-factor model across languages, countries, gender, and sexual orientation are depicted in Supplemental Tables 2 and 3, respectively. All factor loadings were statistically significant and similar across the groups tested ($ps < 0.001$).

The country-based invariance was examined using two pseudo-randomly selected groups (based on alphabetical order), including the same number of countries, to avoid the non-convergence of parameter estimates due to having too many groups (Bou Malham and Saucier, 2014). The configural model showed good fit indices in both groups (CFIs ≥ 0.982; TLIs ≥ 0.978; RMSEA ≤ 0.095). Likewise, the values of ΔCFI, ΔTLI, and ΔRMSEA were in the acceptable range, indicating that the two-factor model of BSI-12 was invariant at the structural level, including factor means, variances, and covariances.

Furthermore, the factor structure of the BSI-12 was also evaluated for each language, obtaining evidence of a good fit of the two-factor solution for most groups (see Supplemental Table 2). Specifically, the configural model exhibited satisfactory fit indices (CFI = 0.983; TLI = 0.979; RMSEA = 0.085 [90 % CI = 0.084–0.086]). In addition, the multigroup analysis yielded evidence that the measurement model was

Table 2

Multigroup analysis of the Brief Symptom Inventory (BSI-12) by country of residence, language and self-reported gender and sexual orientation.

Groups	Model	Fit indices					Model comparison				
		WLSMV χ^2	df	CFI	TLI	RMSEA [CI$_{90\%}$]	Pair	Δdf	Δ_{CFI}	Δ_{TLI}	Δ_{RMSEA}
Country (Group 1)	M1. Configural	14,448.594*	901	0.982	0.978	0.083 [0.082–0.084]					
	M2. Metric	15,596.609*	1061	0.981	0.980	0.079 [0.078–0.080]	M2-M1	160	−0.001	0.002	−0.004
	M3. Scalar	33,063.223*	1605	0.961	0.971	0.095 [0.094–0.096]	M3-M2	544	−0.020	−0.009	0.016
	M4. Residual	41,115.505*	1785	0.948	0.967	0.101 [0.100–0.101]	M4-M3	180	−0.011	−0.004	0.005
	M5. Latent variance-covariance	22,867.147*	1833	0.972	0.983	0.073 [0.072–0.073]	M5-M4	48	0.024	0.016	−0.028
	M6. Latent means	**29,982.336***	**1863**	**0.963**	**0.978**	**0.083 [0.082–0.084]**	**M6-M5**	**30**	**−0.009**	**−0.005**	**0.010**
Country (Group 2)	M1. Configural	15,299.593*	901	0.984	0.980	0.086 [0.085–0.088]					
	M2. Metric	14,056.682*	1061	0.985	0.984	0.076 [0.075–0.077]	M2-M1	160	0.001	0.004	−0.010
	M3. Scalar	30,442.069*	1605	0.967	0.977	0.092 [0.091–0.093]	M3-M2	544	−0.018	−0.007	0.016
	M4. Residual	31,525.637*	1785	0.966	0.979	0.088 [0.087–0.089]	M4-M3	180	−0.001	0.002	−0.004
	M5. Latent variance-covariance	16,720.571*	1833	0.983	0.990	0.062 [0.061–0.062]	M5-M4	48	0.017	0.010	−0.025
	M6. Latent means	**20,185.463***	**1863**	**0.979**	**0.987**	**0.068 [0.067–0.069]**	**M6-M5**	**30**	**−0.004**	**−0.003**	**0.006**
Language	M1. Configural	30,160.712*	1166	0.983	0.979	0.085 [0.084–0.086]					
	M2. Metric	30,441.380*	1376	0.983	0.982	0.078 [0.078–0.079]	M2-M1	210	<0.001	0.003	−0.007
	M3. Scalar	66,957.846*	2090	0.963	0.974	0.095 [0.094–0.096]	M3-M2	714	−0.020	−0.009	0.019
	M4. Residual	77,839.525*	2330	0.956	0.973	0.097 [0.097–0.098]	M4-M3	240	−0.007	−0.001	0.002
	M5. Latent variance-covariance	41,651.910*	2393	0.976	0.986	0.069 [0.069–0.070]	M5-M4	63	0.020	0.013	−0.028
	M6. Latent means	**53,950.260***	**2433**	**0.970**	**0.982**	**0.079 [0.078–0.079]**	**M6-M5**	**40**	**−0.006**	**−0.004**	**0.010**
Gender	M1. Configural	22,454.722*	159	0.984	0.980	0.074 [0.074–0.075]					
	M2. Metric	15,486.690*	179	0.989	0.988	0.058 [0.057–0.059]	M2-M1	20	0.005	0.008	−0.017
	M3. Scalar	22,452.771*	247	0.984	0.987	0.060 [0.059–0.060]	M3-M2	68	−0.005	−0.001	0.002
	M4. Residual	20,330.587*	259	0.986	0.989	0.055 [0.055–0.056]	M4-M3	12	0.002	0.002	−0.005
	M5. Latent variance-covariance	9506.150*	265	0.993	0.995	0.037 [0.036–0.038]	M5-M4	6	0.007	0.006	−0.018
	M6. Latent means	**13,731.841***	**267**	**0.990**	**0.993**	**0.045 [0.044–0.045]**	**M6-M5**	**2**	**−0.003**	**−0.002**	**0.008**
Sexual orientation	M1. Configural	22,699.096*	424	0.984	0.980	0.074 [0.074–0.075]					
	M2. Metric	15,880.136*	494	0.989	0.988	0.057 [0.057–0.058]	M2-M1	70	0.005	0.008	−0.017
	M3. Scalar	20,763.284*	732	0.985	0.989	0.054 [0.053–0.054]	M3-M2	238	−0.004	0.001	−0.003
	M4. Residual	17,715.159*	804	0.988	0.992	0.047 [0.047–0.048]	M4-M3	72	0.003	0.003	−0.007
	M5. Latent variance-covariance	8332.040*	825	0.994	0.996	0.031 [0.030–0.032]	M5-M4	21	0.006	0.004	−0.016
	M6. Latent means	**10,539.393***	**837**	**0.993**	**0.995**	**0.035 [0.034–0.036]**	**M6-M5**	**12**	**−0.001**	**−0.001**	**0.004**

Note. χ^2 = Chi-square; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; Δ_{CFI}, Δ_{TLI}, and Δ_{RMSEA} = change in CFI/TLI, RMSEA values compared to the preceding model. Bold letters indicate the final level of invariance that was achieved. * $p < 0.001$.

psychometrically equivalent by language when models were increasingly constrained, revealing that no noteworthy variations in overall fit were detected according to ΔCFI, ΔTLI, and ΔRMSEA. These results support the cross-cultural comparison (by language or country) of anxiety and depression levels among these conditions.

Subsequently, we examined potential measurement biases associated with participants' self-reported gender and sexual orientation conditions separately. Fit indices for the two-factor model are shown in Supplemental Table 3 for each group examined. All factor loadings were statistically significant and similar between the groups ($ps < 0.001$).

For gender, the configural model showed good fit indices (CFI = 0.984; TLI = 0.980; RMSEA = 0.074 [90 % CI = 0.074–0.075]). Importantly, the lack of substantive differences to the fit detriment after imposing constraints (ΔCFIs ≤0.007; ΔTLIs ≤0.008; ΔRMSEA ≥0.018; see Table 2) supports full structural invariance, suggesting that all parameters were equivalent regardless of participants' gender.

Regarding sexual orientation (Table 2), the configural invariance was supported by good fit indices (CFI = 0.984; TLI = 0.980; RMSEA = 0.074 [90 % CI = 0.074–0.075]). Likewise, the two-factor model of BSI-12 demonstrated structural invariance across groups. All models showed acceptable fit, and no notable deviation was evident between constrained and freely estimated models.

3.3. BSI-12 cut-off scores and criterion-related validity

BSI-12 anxiety and depression T scores were transformed, in each gender condition, using raw scores of each subscale by summing all items. As a result, we identified participants with a high or low risk of anxiety or depression, and tested criterion-related evidence. Tables 3, 4, and 5 compare men, women, or gender-diverse participants' physical, sexual, and mental health in the low- and high-risk cut-off scores of anxiety and depression groups, respectively.

Overall, 13.75 % of participants scored above caseness thresholds for high risk for depression and anxiety. Women had a higher proportion of high-risk cases of both depression (14 %) and anxiety (14.1 %) than men and gender-diverse individuals (< 13.5 %). For anxiety and depression subscale scores, there were significant differences regardless of whether we compared high- and low-risk participants for anxiety or depression ($ps < 0.001$), reflecting their comorbidity. However, the effect size of the differences in anxiety was consistently larger between participants at high- and low-risk for anxiety ($rs \geq 0.585$) than for depression ($rs \leq 0.400$). Conversely, the differences in depression were more notable between the high- and low-risk groups for depression ($rs \geq 0.578$) than for anxiety ($rs \leq 0.406$), reflecting specificity in the expression and frequency of symptoms. This effect was generalized in men, women, and gender-diverse participants.

Table 3

Comparison of men's clinical characteristics in the low- and high-risk of anxiety and depression groups based on the Brief Symptom Inventory (BSI-12).

Variables	Low-risk anxiety (n = 25,969; 86.7 %)			High-risk anxiety (n = 3989; 13.3 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	p	r effect size
Brief Symptom Inventory total score (BSI-12)	8.55	7.11	7.00	29.57	7.54	29.00	2,854,190.00	−96.338	<0.001	0.557
Anxiety Factor	3.70	3.18	3.00	15.52	3.20	15.00	0	−102.334	<0.001	0.591
Depression Factor	4.85	4.85	3.00	14.06	5.68	14.00	11,997,382.50	−78.574	<0.001	0.454
Valuation of life circumstances compared to others[a]	4.92	1.01	5.00	4.52	1.09	5.00	41,555,591.00	−21.064	<0.001	0.122
Emotionally affected by the COVID-19 pandemic[b]	4.02	1.76	4.00	4.58	1.80	5.00	39,083,733.50	−17.704	<0.001	0.104
Perceived stress related to the COVID-19 pandemic[c]	3.42	1.67	3.00	4.41	1.82	5.00	32,844,111.50	−31.029	<0.001	0.183
Compulsive Personality Assessment Scale total score[d]	12.79	4.92	13.00	16.24	5.06	16.00	32,094,036.00	−38.339	<0.001	0.222
Potential suicide risk (P4 Screener)	0.35	0.78	0.00	0.97	1.19	0.00	33,707,750.00	−37.747	<0.001	0.223

Variables	Low-risk anxiety (n = 25,969; 86.7 %)		High-risk anxiety (n = 3989; 13.3 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	4475	17.83	1940	50.87	2092.804	<0.001	0.269	0.269
No	20,623	82.17	1874	49.13				
Suffering from any physical illness?								
Yes	5501	21.94	802	21.04	1.546	0.214	0.007	0.007
No	19,575	78.06	3811	78.96				
Suffering from any sexual problems?								
Yes	2861	11.44	687	18.06	134.082	<0.001	0.068	0.068
No	22,151	88.56	3117	81.94				
Have you had thoughts of actually hurting yourself?								
Yes	5789	23.31	1924	50.90	1267.719	<0.001	0.211	0.211
No	19,041	76.69	1856	49.10				

Variables	Low-risk depression (n = 25,917; 86.5 %)			High-risk depression (n = 4041; 13.5 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	P	r effect size
Brief Symptom Inventory total score (BSI-12)	8.48	7.01	7.00	29.71	7.34	29.00	2,470,152.50	−97.681	<0.001	0.564
Anxiety Factor	4.21	4.10	3.00	12.07	5.78	12.00	14,222,835.50	−74.948	<0.001	0.433
Depression Factor	4.27	3.78	3.00	17.64	3.00	17.00	0	−102.821	<0.001	0.594
Valuation of life circumstances compared to others[a]	4.94	1.00	5.00	4.41	1.10	4.00	38,704,618.00	−27.966	<0.001	0.162
Emotionally affected by the COVID-19 pandemic[b]	4.03	1.76	4.00	4.52	1.85	5.00	40,683,776.50	−15.774	<0.001	0.093
Perceived stress related to the COVID-19 pandemic[c]	3.44	1.67	3.00	4.31	1.87	5.00	35,110,311.00	−27.570	<0.001	0.162
Compulsive Personality Assessment Scale total score[d]	12.85	4.93	13.00	15.82	5.19	16.00	35,319,987.50	−32.986	<0.001	0.191
Potential suicide risk (P4 Screener)	0.31	0.73	0.00	1.22	1.24	1.00	27,592,883.50	−56.238	<0.001	0.332

Variables	Low-risk depression (n = 25,917; 86.5 %)		High-risk depression (n = 4041; 13.5 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	4337	17.33	2078	53.54	2552.636	<0.001	0.297	0.297
No	20,694	82.67	1803	46.46				
Suffering from any physical illness?								
Yes	5412	21.64	891	22.97	3.476	0.062	0.011	0.011
No	19,596	78.36	3879	77.03				
Suffering from any sexual problems?								
Yes	2795	11.21	753	19.45	210.896	<0.001	0.086	0.086
No	22,149	88.79	3119	80.55				
Have you had thoughts of actually hurting yourself?								
Yes	5344	21.57	2369	61.82	2730.990	<0.001	0.309	0.309
No	19,434	78.43	1463	38.18				

Note. M = Mean; SD = Standard deviation; Mdn = Median.

[a] 1 = Among the worst; 2 = Much worse than average; 3 = Worse than average; 4 = Average; 5 = Better than average; 6 = Much better than average; 7 = Among the best.

[b] 1 = Not at all; 7 = Very much.

[c] 1 = No stress; 7 = Extreme stress.

[d] 0 = Not at all characteristic of me; 1 = Rather not characteristic of me; 2 = Somewhat characteristic of me; 3 = Characteristic of me; 4 = Entirely characteristic of me.

Table 4

Comparison of women's clinical characteristics in the low- and high-risk of anxiety and depression groups based on the Brief Symptom Inventory (BSI-12).

Variables	Low-risk anxiety (n = 37,156; 85.9 %)			High-risk anxiety (n = 6100; 14.1 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	p	r effect size
Brief Symptom Inventory total score (BSI-12)	11.52	8.09	10.00	32.34	7.12	32.00	7,862,401.50	−116.740	<0.001	0.561
Anxiety Factor	5.69	4.03	5.00	17.95	2.63	17.00		−125.615	<0.001	0.604
Depression Factor	5.83	5.09	4.00	14.39	5.79	15.00	31,961,916.50	−90.218	<0.001	0.434
Valuation of life circumstances compared to others[a]	4.82	0.95	5.00	4.58	1.02	5.00	98,583,516.00	−17.221	<0.001	0.083
Emotionally affected by the COVID-19 pandemic[b]	4.24	1.63	4.00	4.73	1.70	5.00	85,724,823.50	−21.374	<0.001	0.105
Perceived stress related to the COVID-19 pandemic[c]	3.93	1.61	4.00	4.79	1.68	5.00	73,083,263.00	−36.591	<0.001	0.180
Compulsive Personality Assessment Scale total score[d]	13.03	4.82	13.00	15.97	5.18	16.00	75,716,280.50	−40.990	<0.001	0.197
Potential suicide risk (P4 Screener)	0.54	0.94	0.00	1.27	1.30	1.00	69,002,733.50	−46.828	<0.001	0.231

Variables	Low-risk anxiety (n = 37,156; 85.9 %)		High-risk anxiety (n = 6100; 14.1 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	10,580	29.56	3725	63.60	2585.649	<0.001	0.249	0.249
No	25,208	70.44	2132	36.40				
Suffering from any physical illness?								
Yes	7152	20.01	1368	23.38	35.168	<0.001	0.029	0.029
No	28,598	79.99	4483	76.62				
Suffering from any sexual problems?								
Yes	2749	7.72	724	12.42	144.448	<0.001	0.059	0.059
No	32,878	92.28	5105	87.58				
Have you had thoughts of actually hurting yourself?								
Yes	11,536	32.77	3508	60.99	1694.070	<0.001	0.203	0.203
No	23,668	67.23	2244	39.01				

Variables	Low-risk depression (n = 37,201; 86.0 %)			High-risk depression (n = 6055; 14.0 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	p	r effect size
Brief Symptom Inventory total score (BSI-12)	11.49	8.00	10.00	32.73	6.83	33.00	6,348,574.00	−118.006	<0.001	0.567
Anxiety Factor	6.27	4.94	5.00	14.47	5.39	15.00	31,264,099.00	−90.465	<0.001	0.435
Depression Factor	5.21	4.08	4.00	18.26	2.71	18.00		−125.269	<0.001	0.602
Valuation of life circumstances compared to others[a]	4.84	0.94	5.00	4.45	1.01	4.00	88,321,033.50	−28.489	<0.001	0.137
Emotionally affected by the COVID-19 pandemic[b]	4.24	1.62	4.00	4.68	1.72	5.00	87,058,862.50	−19.274	<0.001	0.095
Perceived stress related to the COVID-19 pandemic[c]	3.95	1.61	4.00	4.66	1.72	5.00	77,700,887.50	−30.572	<0.001	0.150
Compulsive Personality Assessment Scale total score[d]	13.09	4.86	13.00	15.61	5.15	15.00	80,503,691.50	−34.927	<0.001	0.168
Potential suicide risk (P4 Screener)	0.49	0.89	0.00	1.55	1.32	2.00	55,088,234.00	−66.519	<0.001	0.328

Variables	Low-risk depression (n = 37,201; 86.0 %)		High-risk depression (n = 6055; 14.0 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	10,633	29.68	3672	63.09	2478.568	<0.001	0.244	0.244
No	25,192	70.32	2148	36.91				
Suffering from any physical illness?								
Yes	7140	19.96	1380	23.71	43.276	<0.001	0.032	0.032
No	28,640	80.04	4441	76.29				
Suffering from any sexual problems?								
Yes	2679	7.51	794	13.69	248.082	<0.001	0.077	0.077
No	32,978	92.49	5005	86.31				
Have you had thoughts of actually hurting yourself?								
Yes	10,986	31.19	4058	70.82	3331.089	<0.001	0.285	0.285
No	24,240	66.81	1672	29.18				

Note. M = Mean; SD = Standard deviation; Mdn = Median.

[a] 1 = Among the worst; 2 = Much worse than average; 3 = Worse than average; 4 = Average; 5 = Better than average; 6 = Much better than average; 7 = Among the best.

[b] 1 = Not at all; 7 = Very much.

[c] 1 = No stress; 7 = Extreme stress.

[d] 0 = Not at all characteristic of me; 1 = Rather not characteristic of me; 2 = Somewhat characteristic of me; 3 = Characteristic of me; 4 = Entirely characteristic of me.

Table 5

Comparison of gender-diverse participants' clinical characteristics in the low- and high-risk of anxiety and depression groups based on the Brief Symptom Inventory (BSI-12).

Variables	Low-risk anxiety (n = 2243; 86.9 %)			High-risk anxiety (n = 338; 13.1 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	p	r effect size
Brief Symptom Inventory total score (BSI-12)	17.62	9.92	17.00	38.64	6.27	40.00	31,382.00	−27.231	<0.001	0.536
Anxiety Factor	8.12	4.96	8.00	20.51	2.10	20.00		−29.713	<0.001	0.585
Depression Factor	9.50	6.24	9.00	18.13	5.15	19.00	115,878.00	−20.626	<0.001	0.406
Valuation of life circumstances compared to others[a]	4.53	1.08	5.00	4.08	1.22	4.00	295,338.00	−6.744	<0.001	0.133
Emotionally affected by the COVID-19 pandemic[b]	4.49	1.63	5.00	4.92	1.69	5.00	297,680.00	−4.447	<0.001	0.089
Perceived stress related to the COVID-19 pandemic[c]	4.17	1.66	4.00	4.81	1.88	5.00	272,512.00	−6.641	<0.001	0.133
Compulsive Personality Assessment Scale total score[d]	14.27	5.04	14.00	16.96	5.33	17.00	265,841.50	−8.667	<0.001	0.171
Potential suicide risk (P4 Screener)	1.29	1.28	1.00	2.24	1.29	2.00	209,988.50	−11.798	<0.001	0.238

Variables	Low-risk anxiety (n = 2243; 86.9 %)		High-risk anxiety (n = 338; 13.1 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	1390	64.50	289	87.84	70.987	<0.001	0.169	0.169
No	765	35.50	40	12.16				
Suffering from any physical illness?								
Yes	569	26.29	111	33.84	8.178	0.004	0.057	0.057
No	1595	73.71	217	66.16				
Suffering from any sexual problems?								
Yes	302	13.99	65	19.88	7.808	0.005	0.056	0.056
No	1856	86.01	262	80.12				
Have you had thoughts of actually hurting yourself?								
Yes	1280	60.00	271	84.16	70.150	<0.001	0.169	0.169
No	853	40.00	51	15.84				

Variables	Low-risk depression (n = 2252; 87.3 %)			High-risk depression (n = 329; 12.7 %)			Mann-Whitney U tests			
	M	SD	Mdn	M	SD	Mdn	U	Z	p	r effect size
Brief Symptom Inventory total score (BSI-12)	17.66	9.93	17.00	38.95	5.98	39.00	26,791.00	−27.227	<0.001	0.536
Anxiety Factor	8.69	5.69	8.00	16.92	5.33	17.00	114,001.00	−20.334	<0.001	0.400
Depression Factor	8.96	5.51	9.00	22.03	1.48	22.00		−29.368	<0.001	0.578
Valuation of life circumstances compared to others[a]	4.55	1.08	5.00	3.96	0.13	4.00	263,243.00	−8.766	<0.001	0.173
Emotionally affected by the COVID-19 pandemic[b]	4.52	1.63	5.00	4.74	1.73	5.00	311,809.00	−2.282	0.023	0.046
Perceived stress related to the COVID-19 pandemic[c]	4.20	1.67	4.00	4.62	1.86	5.00	288,277.00	−4.382	<0.001	0.088
Compulsive Personality Assessment Scale total score[d]	14.33	5.06	14.00	16.62	5.38	17.00	275,076.50	−7.411	<0.001	0.146
Potential suicide risk (P4 Screener)	1.25	1.25	1.00	2.56	1.26	3.00	156,663.50	−15.723	<0.001	0.317

Variables	Low-risk depression (n = 2252; 87.3 %)		High-risk depression (n = 329; 12.7 %)		χ^2 tests			
	n	%	n	%	χ^2	p	Phi	Cramer's V
Suffering from any mental illness or emotional problems?								
Yes	1414	65.19	265	84.13	45.023	<0.001	0.135	0.135
No	755	34.81	50	15.87				
Suffering from any physical illness?								
Yes	575	26.41	105	33.33	6.643	0.010	0.052	0.052
No	1602	73.59	210	66.67				
Suffering from any sexual problems?								
Yes	306	14.09	61	19.49	6.339	0.012	0.051	0.051
No	1866	85.91	252	80.51				
Have you had thoughts of actually hurting yourself?								
Yes	1281	59.64	270	87.95	92.545	<0.001	0.194	0.194
No	867	40.36	37	12.05				

Note. M = Mean; SD = Standard deviation; Mdn = Median.

[a] 1 = Among the worst; 2 = Much worse than average; 3 = Worse than average; 4 = Average; 5 = Better than average; 6 = Much better than average; 7 = Among the best.

[b] 1 = Not at all; 7 = Very much.

[c] 1 = No stress; 7 = Extreme stress.

[d] 0 = Not at all characteristic of me; 1 = Rather not characteristic of me; 2 = Somewhat characteristic of me; 3 = Characteristic of me; 4 = Entirely characteristic of me.
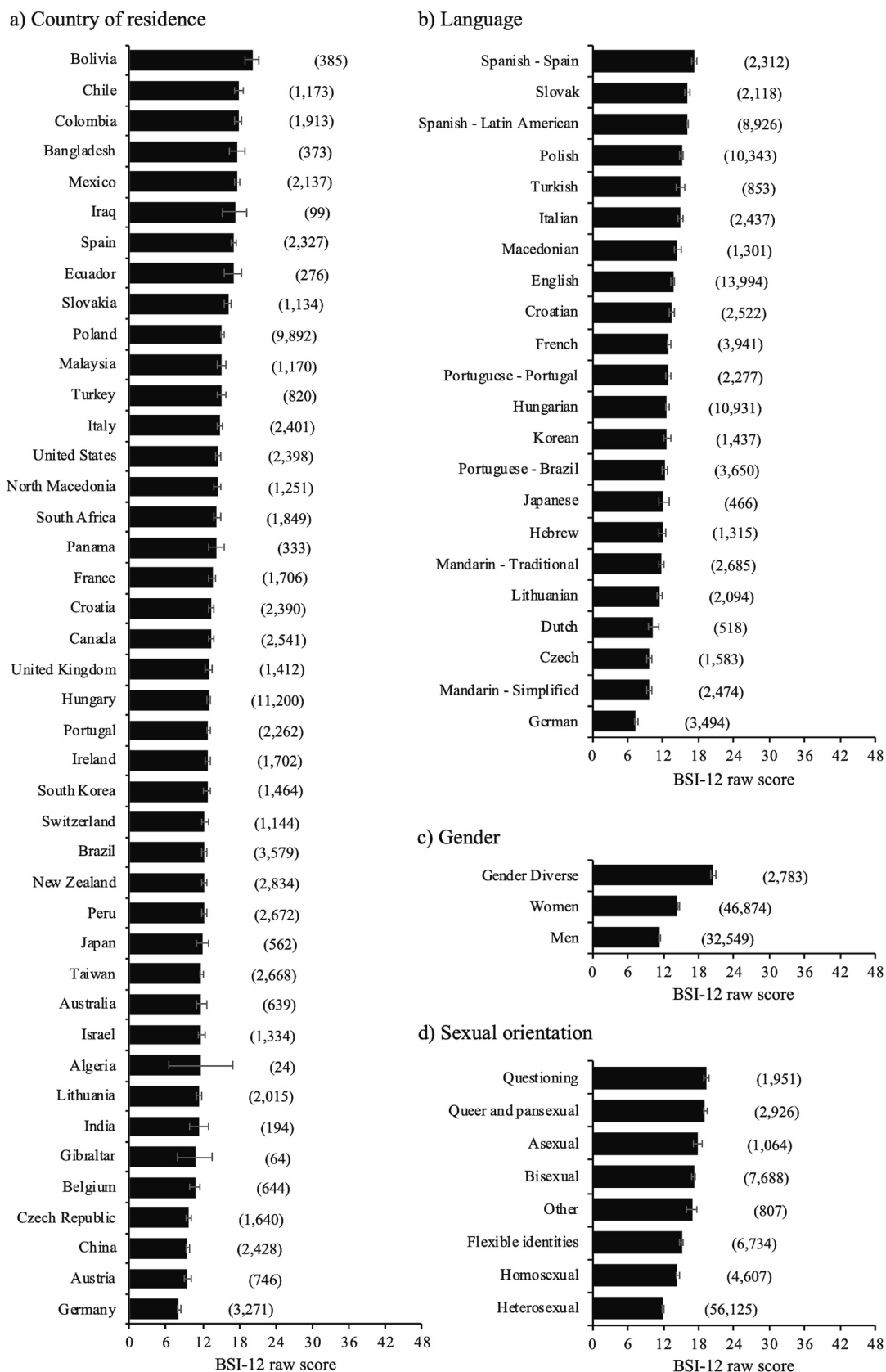
Fig. 1. Brief Symptom Inventory (BSI-12) raw scores by variables included in multigroup analysis. Error bars represent the 95 % confidence interval, and sample sizes are depicted in parentheses.

Regarding physical illnesses, there were no significant differences in relationships between men with high and low risk of anxiety ($\chi^2 = 1.544$; $p = 0.214$) or depression ($\chi^2 = 3.476$; $p = 0.062$). Consistent with the above, although there were statistically significant effects between women ($\chi^2 \geq 35.168$; $ps < 0.001$) and individuals of gender diversities ($\chi^2 \geq 6.643$; $ps \leq 0.010$), in both cases, effect sizes for anxiety (Vs ≤ 0.057) and depression (Vs ≤ 0.052) were marginal. In the same vein, there were significant, yet slight effects on experiencing sexual problems in men (anxiety: $\chi^2 = 134.082$; $p < 0.001$; V = 0.068, depression: $\chi^2 = 210.896$; $p < 0.001$; V = 0.086), women (anxiety: $\chi^2 = 144.448$; $p = 0.001$; V = 0.059, depression: $\chi^2 = 248.082$; $p < 0.001$; V = 0.077), and gender-diverse participants (anxiety: $\chi^2 = 7.808$; $p = 0.005$; V = 0.056, depression: $\chi^2 = 6.339$; $p = 0.012$; V = 0.051).

Contrarily, participants experiencing mental illness or emotional problems showed larger effect sizes than those facing physical and sexual concerns. Among men, those with a high risk of anxiety or depression exhibited significantly higher levels of mental or emotional concerns (anxiety: $\chi^2 = 2092.804$; $p < 0.001$; V = 0.269, depression: $\chi^2 = 2552.636$; $p < 0.001$; V = 0.297). This effect, although slightly less pronounced, was also observed in women (anxiety: $\chi^2 = 2585.649$; $p < 0.001$; V = 0.249, depression: $\chi^2 = 2478.568$; $p < 0.001$; V = 0.244) and gender-diverse individuals (anxiety: $\chi^2 = 70.987$; $p < 0.001$; V = 0.169, depression: $\chi^2 = 45.023$; $p < 0.001$; $V = 0.135$). Similar patterns of differences emerged when examining psychological concerns, specifically emotional concerns and perceived stress related to the COVID-19 pandemic. Participants with a high risk of anxiety had higher perceived stress, with moderate effect sizes ranging from $r = 0.133$ to $r = 0.183$ (Zs ≤ -6.641; $ps < 0.001$). This was also observed in participants at high risk of depression, although with slightly smaller effect sizes ranging from $r = 0.088$ to $r = 0.162$ (Zs ≤ -4.382; $ps < 0.001$). For emotional concerns, significant differences were found for participants at high risk of anxiety (Zs ≤ -4.382; $ps < 0.001$) or depression (Zs ≤ -2.282; $ps \leq 0.023$), albeit relatively small in effect size among those at risk for depression ($rs \leq 0.095$) or anxiety ($rs \leq 0.105$).

When analyzing compulsive personality traits as a criterion variable, although both group of participants identified with a high risk of anxiety (Zs ≤ -8.667; $ps < 0.001$) and depression (Zs ≤ -7.411; $ps < 0.001$) showed significantly higher scores on the CPAS, the effect sizes were slightly larger for participants in the high-risk group for anxiety ($rs \leq 0.222$) compared to those in the high-risk group for depression ($rs \leq 0.191$). These differences in effect size align with the overlap reported between these three disorders (Goodwin, 2015).

This pattern was reversed when examining the potential risk of suicide. First, participants identified as being at high risk of developing depression reported a significant prevalence of self-harm thoughts, with effect sizes ranging from moderate to larger (V = 0.194–0.309; χ^2s ≥ 92.545; $ps < 0.001$). In contrast, participants at high risk of developing anxiety showed smaller effect sizes, ranging from V = 0.169 to V = 0.211 (χ^2s ≥ 70.150; $ps < 0.001$). Second, this difference in effect size was also evident in the comparisons involving the risk of depression when considering the sum scores from the P4 screener. Comparing participants at high risk of anxiety, significant differences were found with moderate effect sizes ($r = 0.223$–0.238; Zs ≤ -11.798; $ps < 0.001$). On the other hand, participants at a higher risk of depression showed consistently significant differences, accompanied by larger effect sizes ($r = 0.317$–0.332; Zs ≤ -15.723; $ps < 0.001$).

4. Discussion

Cross-cultural research is at the forefront when systematically considering similarities and differences across cultures and societies, significantly informing clinical standards across countries. However, pitfalls may hinder replicability of these cross-cultural studies (Allik et al., 2012; Jeong and Lee, 2019). Therefore, the present study comprehensively examined the psychometric properties and validated a depression and anxiety screener (i.e., BSI-12) in 42 countries. Our

findings suggest that the depression and anxiety sub-dimensions of the original BSI-18 demonstrated good structural validity, reliability, and measurement invariance across multiple groups and satisfactorily distinguished low-risk and high-risk participants.

Our results support a two-factor latent structure of the BSI-12 in a diverse sample of adults. The BSI-18, which includes three domains (i.e., depression, anxiety, and somatization) supported a three-dimensional structure. Thus, and in line with previous findings (e.g., Franke et al., 2017; Recklitis et al., 2017), the BSI demonstrated good reliability, showing a positive, strong association with suicidality and small-to-medium with behavioral indicators.

The results supported a latent means level of invariance for all groups. Regarding country-based measurement invariance, differences in total BSI-12 scores across countries ranged between 8 and 20 points (Fig. 1). Countries within the fifth highest scores in the BSI averaged a 2021-Human Development Index (HDI) of 0.744 which is deemed high, whereas the fifth lowest scores average a HDI of 0.891, which is deemed very high. The HDI is a compound measure comparing countries on living standards, education, health and quality of life (UNDP, 2023). Countries with high HDIs tend to invest more heavily in mental health (see Knapp and Wong, 2020), based, in part, on data supporting the cost-utility ratio of treating mental health concerns (e.g., Arias et al., 2022; Brettschneider et al., 2015). However, reports on a phenomenon called the "vulnerability paradox" show the mix results in the literature when it comes to establishing a negative relationship between HDI or other country-wide estimates (e.g., gross domestic product per capita, or a positive relationship with the vulnerability index) and mental health disorder's prevalence (Dückers and Brewin, 2016; Dückers et al., 2019; Jorm and Mulder, 2021). Still, this study supports growing evidence of the world-wide disparities in affective disorders.

Regarding language-based measurement invariance, the BSI-12 could comparably operate across all 26 languages tested, in accordance with previous studies examining different populations (Lu et al., 2019; Torres et al., 2013). This is relevant for cross-cultural research, given that both depression and anxiety are sensitive to cultural and contextual factors (Bredström, 2019; Hofmann and Hinton, 2014; Lewis-Fernández et al., 2010). Nevertheless, while psychometric tools may offer unbiased, systematic, and straightforward ways to assess symptomatology of depression or anxiety, cultural differences remain. Future studies should consider such differences, even when using cross-culturally valid and invariant scales like the BSI-12, as well as when conducting mixed-methods research.

Regarding sexual orientation- and gender-based measurement invariance, in line with previous findings (Ploderl and Tremblay, 2015), cis-gender and heterosexual people scored significantly lower than other sexual minorities, and did so with large magnitudes (see Fig. 1 and Supplementary Tables 6 and 7). These results corroborate previous findings on depression and anxiety where men scored lower than women (Steel et al., 2014), while women scored lower than gender-diverse individuals (Wittgens et al., 2022), and heterosexual individuals scored lower than other sexual orientations (Ross et al., 2018). Gender-diverse individuals' poor mental health is similar to transgender individuals (Chew et al., 2020). The present finding address a gap regarding the omission or exclusion of sexual orientation and gender identity as variables in relation to affective concerns, despite related mental health disparities in these minority groups (Heck et al., 2017). Therefore, the current study provides not only cross-cultural validation, but also supports that the BSI-12 comparably operates across genders and sexual orientations.

The "caseness" criterion (Derogatis, 2000) accurately distinguished between participants with a high risk of depression and anxiety having experienced mental, physical, emotional, or sexual problems. Whereas the criterion was able to distinguish between low- and high-risk depression and anxiety cases, the effect size of the average differences in the non-parametric test was greater among the depression-risk groups than among the anxiety-risk groups. For both anxiety and depression

risk contrasts, there was an effect on the suicidal ideation criterion. However, in the case of anxiety, this effect was the second highest, following the reporting of suffering from mental or emotional problems. Conversely, the effect of suicidal thoughts was the most pronounced among participants at high risk of depression. In contrast, criterion variables related to compulsive personality and sexual problems or physical illness yielded small effect sizes when discriminating between low- and high-risk depression and anxiety cases. Similarly, the emotional toll and perceived stress due to the COVID-19 pandemic yielded significant differences between low- and high-risk groups for both depression and anxiety, yet of small effect sizes.

4.1. Limitations and future studies

Study limitations warrant mention. First, the predictive validation was not made against a clinical diagnosis, which is considered the gold standard diagnostic tool for either anxiety of depression. Second, the full BSI-18 scale was not examined, with the somatization sub-dimension excluded based on the study aims including the assessing of specific psychopathology associated with specific sexual concerns while simultaneously limiting subject burden (Bőthe et al., 2021; Islam et al., 2022; Weinandy et al., 2023). Future studies should examine the entire BSI-18 to be a fully invariant tool. Third, the study was conducted online, mainly by advertisements through social media, national news websites, and the aid of survey services such as Prolific. Their sample pool has been described—by the company itself—as skewed towards women, younger, and highly educated populations, especially in western, educated, industrialized, rich, and democratic (WEIRD) countries (Prolific, 2023). Thus, while a cross-cultural sample helps to improve the generalizability of these results towards more populations, it still struggles with the limitations that stem from predominantly WEIRD samples (Henrich et al., 2010a, 2010b). Therefore, whereas these services make surveying easier and faster, the samples are likely less representative of a world reality, and while oversampling women may help promote women's health, other groups also warrant consideration. Building upon our work, future studies should also investigate underrepresented gender groups and sexual and racial/ethnic diversities (Brance et al., 2023) to help promote equity (Weersing et al., 2022), as well as other demographic variables of interest such as age generation differences (Twenge, 2011; Twenge, 2015). Finally, the ISS has general limitations that are described in the project's OSF page (https://osf.io/n3k2c/?view_only=838146f6027c4e6bb68371 d9d1 4220b5).

5. Conclusions

This study is the first transcultural validation of the BSI-12, demonstrating that the scale is psychometrically reliable and valid to measure depression and anxiety, as well as invariant across countries, languages, genders, and sexual orientations.

Ethics

The authors assert that all procedures contributing to this work comply with the relevant national and institutional committees' ethical standards on human experimentation and the Helsinki Declaration. The study was approved by all collaborating countries' national/institutional ethics review boards: https://osf.io/n3k2c/?view_only=838146f6027c4e6bb68371d9d14220b5.

CRediT authorship contribution statement

Gonzalo R. Quintana: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Fernando P. Ponce:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Javier I. Escudero-Pastén:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – review & editing. **Juan F. Santibáñez-Palma:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – review & editing. **Léna Nagy:** Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Writing – review & editing. **Mónika Koós:** Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Writing – review & editing. **Shane W. Kraus:** Conceptualization, Data curation, Investigation, Writing – review & editing. **Zsolt Demetrovics:** Conceptualization, Data curation, Funding acquisition, Investigation, Writing – review & editing. **Marc N. Potenza:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – review & editing. **Rafael Ballester-Arnal:** Data curation, Investigation, Writing – review & editing. **Dominik Batthyány:** Data curation, Investigation, Writing – review & editing. **Sophie Bergeron:** Data curation, Investigation, Writing – review & editing. **Joël Billieux:** Data curation, Investigation, Writing – review & editing. **Peer Briken:** Data curation, Investigation, Writing – review & editing. **Julius Burkauskas:** Data curation, Investigation, Writing – review & editing. **Georgina Cárdenas-López:** Data curation, Investigation, Writing – review & editing. **Joana Carvalho:** Data curation, Methodology, Writing – review & editing. **Jesús Castro-Calvo:** Data curation, Investigation,

Writing – review & editing. **Lijun Chen:** Data curation, Investigation, Writing – review & editing. **Giacomo Ciocca:** Data curation, Investigation, Writing – review & editing. **Ornella Corazza:** Data curation, Investigation, Writing – review & editing. **Rita I. Csako:** Data curation, Investigation, Writing – review & editing. **David P. Fernandez:** Data curation, Investigation, Writing – review & editing. **Elaine F. Fernandez:** Data curation, Investigation, Writing – review & editing. **Hironobu Fujiwara:** Data curation, Funding acquisition, Writing – review & editing. **Johannes Fuss:** Data curation, Investigation, Writing – review & editing. **Roman Gabrhelík:** Data curation, Funding acquisition, Investigation, Writing – review & editing. **Ateret Gewirtz-Meydan:** Data curation, Investigation, Writing – review & editing. **Biljana Gjoneska:** Data curation, Investigation, Writing – review & editing. **Mateusz Gola:** Data curation, Investigation, Writing – review & editing. **Joshua B. Grubbs:** Data curation, Investigation, Writing – review & editing. **Hashim T. Hashim:** Data curation, Investigation, Writing – review & editing. **Md. Saiful Islam:** Data curation, Investigation, Writing – review & editing. **Mustafa Ismail:** Data curation, Investigation, Writing – review & editing. **Martha C. Jiménez-Martínez:** Data curation, Investigation, Writing – review & editing. **Tanja Jurin:** Data curation, Investigation, Writing – review & editing. **Ondrej Kalina:** Data curation, Investigation, Writing – review & editing. **Verena Klein:** Data curation, Investigation, Writing – review & editing. **András Költő:** Data curation, Investigation, Writing – review & editing. **Chih-Ting Lee:** Data curation, Investigation, Writing – review & editing. **Sang-Kyu Lee:** Data curation, Investigation, Writing – review & editing. **Karol Lewczuk:** Data curation, Investigation, Writing – review & editing. **Chung-Ying Lin:** Data curation, Investigation, Writing – review & editing. **Christine Lochner:** Data curation, Investigation, Writing – review & editing. **Silvia López-Alvarado:** Data curation, Investigation, Writing – review & editing. **Kateřina Lukavská:** Data curation, Investigation, Writing – review & editing. **Percy Mayta-Tristán:** Data curation, Investigation, Writing – review & editing. **Dan J. Miller:** Data curation, Investigation, Writing – review & editing. **Oľga Orosová:** Data curation, Investigation, Writing – review & editing. **Gábor Orosz:** Data curation, Investigation, Writing – review & editing. **Gabriel C. Quintero Garzola:** Data curation, Investigation, Writing – review & editing. **Jano Ramos-Diaz:** Data curation, Investigation, Writing – review & editing. **Kévin Rigaud:** Data curation, Investigation, Writing – review & editing. **Ann Rousseau:** Data curation, Investigation, Writing – review & editing. **Marco De Tubino Scanavino:** Data curation, Investigation, Writing – review & editing. **Marion K. Schulmeyer:** Data curation, Investigation, Writing – review & editing. **Pratap Sharan:** Data curation, Investigation, Writing – review & editing. **Mami Shibata:** Data curation, Investigation, Writing – review & editing. **Sheikh Shoib:** Data curation, Investigation, Writing – review & editing. **Vera Sigre-Leirós:** Data curation, Investigation, Writing – review & editing. **Luke Sniewski:** Data curation, Investigation, Writing – review & editing. **Ognen Spasovski:** Data curation, Investigation, Writing – review & editing. **Vesta Steibliene:** Data curation, Investigation, Writing – review & editing. **Dan J. Stein:** Data curation, Investigation, Writing – review & editing. **Berk C. Ünsal:** Data curation, Investigation, Writing – review & editing. **Marie-Pier Vaillancourt-Morel:** Data curation, Investigation, Writing – review & editing. **Marie Claire Van Hout:** Data curation, Investigation, Writing – review & editing. **Beáta Bőthe:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing.

Declaration of competing interest

JBurkauskas works as a consultant at Cronos. MNP discloses that he has consulted for and advised Game Day Data, Addiction Policy Forum, AXA, Idorsia, Baria-Tek, and Opiant Therapeutics; been involved in a patent application involving Novartis and Yale; received research support from the Mohegan Sun Casino, Children and Screens and the Connecticut Council on Problem Gambling; consulted for or advised legal, gambling and non-profit entities on issues related to impulse control, internet use and addictive behaviors; conducted clinical evaluations related to impulse-control and addictive behaviors; performed grant reviews; edited journals/journal sections; given academic lectures in grand rounds, CME events and other clinical/scientific venues; and generated books or chapters for publishers of mental health texts. The views presented in this manuscript represent those of the authors and not necessarily those of the funding agencies. The rest of the authors report no conflicts of interest with respect to the content of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.01.127.

References

Akena, D., Joska, J., Obuku, E.A., Amos, T., Musisi, S., Stein, D.J., 2012. Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries: a systematic review. BMC Psychiatry 12 (1), 187. https://doi.org/10.1186/1471-244X-12-187.

Allik, J., Massoudi, K., Realo, A., Rossier, J., 2012. Personality and culture cross-cultural psychology at the next crossroads. Swiss J. Psychol. 71 (1), 5–12. https://doi.org/10.1024/1421-0185/a000069.

APA. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed* [doi: https://doi.org/10.1176/appi.books.9780890425596]. American Psychiatric Publishing, Inc. doi:https://doi.org/10.1176/appi.books.9780890425596.

Arias, D., Saxena, S., Verguet, S., 2022. Quantifying the global burden of mental disorders and their economic value. eClinicalMedicine 54, 101675. https://doi.org/10.1016/j.eclinm.2022.101675.

Borgogna, N.C., McDermott, R.C., Aita, S.L., Kridel, M.M., 2019. Anxiety and depression across gender and sexual minorities: implications for transgender, gender nonconforming, pansexual, demisexual, asexual, queer, and questioning individuals. Psychol. Sex. Orientat. Gend. Divers. 6, 54–63. https://doi.org/10.1037/sgd0000306.

Bőthe, B., Koós, M., Nagy, L., Kraus, S., Potenza, M., Demetrovics, Z., 2021. International sex survey: study protocol of a large, cross-cultural collaborative study in 45 countries. J. Behav. Addict. 10 (3), 632–645. https://doi.org/10.1556/2006.2021.00063.

Bou Malham, P., Saucier, G., 2014. Measurement invariance of social axioms in 23 countries. J. Cross Cult. Psychol. 45 (7), 1046–1060. https://doi.org/10.1177/0022022114534771.

Brance, K., Chatzimpyros, V., Bentall, R.P., 2023. Increased social identification is linked with lower depressive and anxiety symptoms among ethnic minorities and migrants: a systematic review and meta-analysis. Clin. Psychol. Rev. 99, 102216 https://doi.org/10.1016/j.cpr.2022.102216.

Bredström, A., 2019. Culture and context in mental health diagnosing: scrutinizing the DSM-5 revision. Journal of Medical Humanities 40 (3), 347–363. https://doi.org/10.1007/s10912-017-9501-1.

Brettschneider, C., Djadran, H., Härter, M., Löwe, B., Riedel-Heller, S., König, H.H., 2015. Cost-utility analyses of cognitive-behavioural therapy of depression: a systematic review. Psychother. Psychosom. 84 (1), 6–21. https://doi.org/10.1159/000365150.

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. Sociol. Methods Res. 21 (2), 230–258. https://doi.org/10.1177/0049124192021002005.

Burkauskas, J., & Fineberg, N. A. (2020). History and epidemiology of OCPD. In M. D. M. P. H. Jon E. Grant, M. P. D. M. Samuel R. Chamberlain, & A. Pinto (Eds.), Obsessive-Compulsive Personality Disorder. American Psychiatric Association Publishing. https://books.google.cl/books?id=1LrPxQEACAAJ.

Canino, G., Alegria, M., 2008. Psychiatric diagnosis - is it universal or relative to culture? J. Child Psychol. Psychiatry 49 (3), 237–250. https://doi.org/10.1111/j.1469-7610.2007.01854.x.

Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct. Equ. Model. Multidiscip. J. 14 (3), 464–504. https://doi.org/10.1080/10705510701301834.

Chew, D., Tollit, M.A., Poulakis, Z., Zwickl, S., Cheung, A.S., Pang, K.C., 2020. Youths with a non-binary gender identity: a review of their sociodemographic and clinical profile. Lancet Child & Adolescent Health 4 (4), 322–330. https://doi.org/10.1016/s2352-4642(19)30403-1.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. https://doi.org/10.4324/9780203771587.

Cramer, H., 1946. Mathematical Methods of Statistics. Princeton University Press.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16 (3), 297–334. https://doi.org/10.1007/BF02310555.

Derogatis, L.R., 1982. Brief symptom inventory (BSI). In a. PsycTests (Ed.).

Derogatis, L.R., 2000. BSI 18 Brief Symptom Inventory 18: Administration, Scoring and Procedures Manual. NCS Pearson.

Derogatis, L. R., & Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18. In *the Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults, Volume 3,* 3rd ed. (pp. 1–41). Lawrence Erlbaum Associates Publishers.

Dube, P., Kurt, K., Bair, M.J., Theobald, D., Williams, L.S., 2010. The p4 screener: evaluation of a brief measure for assessing potential suicide risk in 2 randomized effectiveness trials of primary care and oncology patients. Prim Care Companion J Clin Psychiatry 12 (6). https://doi.org/10.4088/PCC.10m00978blu.

Dunn, T.J., Baguley, T., Brunsden, V., 2014. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Br. J. Psychol. 105, 399–412. https://doi.org/10.1111/bjop.12046.

Dückers, M.L., Brewin, C.R., 2016. A paradox in individual versus National Mental Health Vulnerability: are higher resource levels associated with higher disorder prevalence? J. Trauma. Stress 29 (6), 572–576. https://doi.org/10.1002/jts.22144.

Dückers, M.L.A., Reifels, L., De Beurs, D.P., Brewin, C.R., 2019. The vulnerability paradox in global mental health and its applicability to suicide. Br. J. Psychiatry 1-6. https://doi.org/10.1192/bjp.2019.41.

Durá, E., Andreu, Y., Galdón, M.J., Ferrando, M., Murgui, S., Poveda, R., Jimenez, Y., 2006. Psychological assessment of patients with temporomandibular disorders: confirmatory analysis of the dimensional structure of the brief symptoms inventory 18. J. Psychosom. Res. 60 (4), 365–370. https://doi.org/10.1016/j.jpsychores.2005.10.013.

Feinstein, B.A., Hurtado Jr., M., Dyar, C., Davila, J., 2023. Disclosure, minority stress, and mental health among bisexual, pansexual, and queer (bi+) adults: the roles of primary sexual identity and multiple sexual identity label use. Psychol. Sex. Orientat. Gend. Divers. 10, 181–189. https://doi.org/10.1037/sgd0000532.

Fineberg, N.A., Sharma, P., Sivakumaran, T., Sahakian, B., Chamberlain, S., 2007. Does obsessive-compulsive personality disorder belong within the obsessive-compulsive Spectrum? CNS Spectr. 12 (6), 467–482. https://doi.org/10.1017/S1092852900015340.

Finney, S.J., DiStefano, C., 2013. Nonnormal and categorical data in structural equation modeling. In: Structural Equation Modeling: A Second Course, 2nd ed. IAP Information Age Publishing, pp. 439–492.

Franke, G.H., Jaeger, S., Glaesmer, H., Barkmann, C., Petrowski, K., Braehler, E., 2017. Psychometric analysis of the brief symptom inventory 18 (BSI-18) in a representative German sample. BMC Med. Res. Methodol. 17 (1), 14. https://doi.org/10.1186/s12874-016-0283-3.

Gilbody, S.M., House, A.O., Sheldon, T.A., 2001. Routinely administered questionnaires for depression and anxiety: systematic review. BMJ 322 (7283), 406–409. https://doi.org/10.1136/bmj.322.7283.406.

Gonzales, G., Henning-Smith, C., 2017. Health disparities by sexual orientation: results and implications from the behavioral risk factor surveillance system. J. Community Health 42 (6), 1163–1172. https://doi.org/10.1007/s10900-017-0366-z.

Goodwin, G.M., 2015. The overlap between anxiety, depression, and obsessive-compulsive disorder. Dialogues Clin. Neurosci. 17 (3), 249–260. https://doi.org/10.31887/DCNS.2015.17.3/ggoodwin.

Gregorich, S.E., 2006. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. Med. Care 44 (11), S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f.

Green, S.B., Yang, Y., 2009. Commentary on Coefficient Alpha: A Cautionary Tale. Psychometrika 74, 121–135. https://doi.org/10.1007/s11336-008-9098-4.

Heck, N.C., Mirabito, L.A., LeMaire, K., Livingston, N.A., Flentje, A., 2017. Omitted data in randomized controlled trials for anxiety and depression: a systematic review of the inclusion of sexual orientation and gender identity. J. Consult. Clin. Psychol. 85 (1), 72–76. https://doi.org/10.1037/ccp0000123.

Henrich, J., Heine, S.J., Norenzayan, A., 2010a. Most people are not WEIRD. Nature 466, 29. https://doi.org/10.1038/466029a.

Henrich, J., Heine, S.J., Norenzayan, A., 2010b. The weirdest people in the world? Behav. Brain Sci. 33 (2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X.

Hoe, M., Brekke, J.S., 2008. Cross-ethnic measurement invariance of the brief symptom inventory for individuals with severe and persistent mental illness. Soc. Work Res. 32 (2), 71–78. https://doi.org/10.1093/swr/32.2.71.

Hofmann, S.G., Hinton, D.E., 2014. Cross-cultural aspects of anxiety disorders. Curr. Psychiatry Rep. 16 (6), 450. https://doi.org/10.1007/s11920-014-0450-3.

Hu, L.-t, Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55. https://doi.org/10.1080/10705519909540118.

Islam, M.S., Tasnim, R., Sujan, M.S.H., Bothe, B., Ferdous, M.Z., Sikder, M.T., Kraus, S. W., Potenza, M.N., Team, C. A.-P. H, 2022. Validation and evaluation of the psychometric properties of the Bangla version of the brief pornography screen in men and women. Int. J. Ment. Heal. Addict. https://doi.org/10.1007/s11469-022-00903-0.

Jeong, S., Lee, Y., 2019. Consequences of not conducting measurement invariance tests in cross-cultural studies: a review of current research practices and recommendations. Adv. Dev. Hum. Resour. 21 (4), 466–483. https://doi.org/10.1177/1523422319870726.

Jorm, A.F., Mulder, R.T., 2021. Cross-national differences in the prevalence of mental illness symptoms: evidence against the vulnerability paradox. Aust. N. Z. J. Psychiatry 55 (6), 620–626. https://doi.org/10.1177/0004867421998780.

Kessler, R.C., 2007. The global burden of anxiety and mood disorders: putting the European study of the epidemiology of mental disorders (ESEMeD) findings into perspective. J. Clin. Psychiatry 68 Suppl 2(Suppl 2), 10–19.

Kirmayer, L.J., 2001. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. J. Clin. Psychiatry 62 Suppl 13, 22–28 (discussion 29-30).

Kirmayer, L.J., Groleau, D., 2001. Affective disorders in cultural context [article]. Psychiatr. Clin. N. Am. 24 (3), 465–478. https://doi.org/10.1016/S0193-953X(05)70241-0.

Knapp, M., Wong, G., 2020. Economics and mental health: the current scenario. World Psychiatry 19 (1), 3–14. https://doi.org/10.1002/wps.20692.

Lacko, D., Cenek, J., Tocik, J., Avsec, A., Dordevic, V., Genc, A., Haka, F., Sakotic-Kurbalija, J., Mohoric, T., Neziri, I., Subotic, S., 2022. The necessity of testing measurement invariance in cross-cultural research: potential Bias in cross-cultural comparisons with individualism- collectivism self-report scales. *Cross-cultural research, 56*(2-3), 228-267, article 10693971211068971. https://doi.org/10.1177/10693971211068971.

Lewis-Fernández, R., Hinton, D.E., Laria, A.J., Patterson, E.H., Hofmann, S.G., Craske, M. G., Stein, D.J., Asnaani, A., Liao, B., 2010. Culture and the anxiety disorders: recommendations for DSM-V. Depress. Anxiety 27 (2), 212–229. https://doi.org/10.1002/da.20647.

Little, R.J.A., 1988. A test of missing completely at random for multivariate data with missing values. J. Am. Stat. Assoc. 83 (404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722.

Lu, Y., Alvarez, A.N., Miller, M.J., 2019. Measurement invariance of the brief symptom Inventory-18 (BSI-18) across Asian American ethnic, nativity, and gender groups. Asian Am. J. Psychol. 10 (1), 1–10. https://doi.org/10.1037/aap0000115.

Lund, C., Breen, A., Flisher, A.J., Kakuma, R., Corrigall, J., Joska, J.A., Swartz, L., Patel, V., 2010. Poverty and common mental disorders in low and middle income countries: a systematic review. Soc. Sci. Med. 71 (3), 517–528. https://doi.org/10.1016/j.socscimed.2010.04.027.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57 (3), 519–530. https://doi.org/10.1093/biomet/57.3.519.

Marsh, H., Wen, Z., Hau, K.-T., 2004. Structural equation models of latent interactions: evaluation of alternative estimation strategies and Indicator construction. Psychol. Methods 9, 275–300. https://doi.org/10.1037/1082-989X.9.3.275.

McDonald, R.P., 1970. The theoretical foundations of principal component analysis, canonical factor analysis, and alpha factor analysis. Br. J. Math. Stat. Psychol. 23 (1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x.

Meade, A.W., Johnson, E.C., Braddy, P.W., 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. J. Appl. Psychol. 93 (3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568.

Muthén, L.K., Muthén, B.O., 2017. Mplus user's guide, 8th ed. Muthén & Muthén.

Noguera, A., Centeno, C., Carvajal, A., Tejedor, M.A.P., Urdiroz, J., Martinez, M., 2009. Spanish "fine tuning" of language to describe depression and anxiety. J. Palliat. Med. 12 (8), 707–712. https://doi.org/10.1089/jpm.2009.0024.

Ploderl, M., Tremblay, P., 2015. Mental health of sexual minorities. A systematic review. Int. Rev. Psychiatry 27 (5), 367–385. https://doi.org/10.3109/09540261.2015.1083949.

Prolific. (2023). What are the advantages and limitations of an online sample? https://researcher-help.prolific.co/hc/en-gb/articles/360009501473-What-are-the-advantages-and-limitations-of-an-online-sample-.

Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

Rea, L.M., Parker, R.A., 1992. Designing and Conducting Survey Research. Jossey-Boss.

Revelle, W., Zinbarg, R.E., 2009. Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. Psychometrika 74, 145–154. https://doi.org/10.1007/s11336-008-9102-z.

Recklitis, C.J., Blackmon, J.E., Chang, G., 2017. Validity of the brief symptom Inventory-18 (BSI-18) for identifying depression and anxiety in young adult cancer survivors: comparison with a structured clinical diagnostic interview. Psychol. Assess. 29 (10), 1189–1200. https://doi.org/10.1037/pas0000427.

Ridley, M., Rao, G., Schilbach, F., Patel, V., 2020. Poverty, depression, and anxiety: causal evidence and mechanisms. *Science, 370*(6522), 1289−+, article eaay0214. https://doi.org/10.1126/science.aay0214.

Ross, L. E., Salway, T., Tarasoff, L. A., MacKay, J. M., Hawkins, B. W., & Fehr, C. P. (2018). Prevalence of depression and anxiety among bisexual people compared to gay, lesbian, and heterosexual individuals: a systematic review and Meta-analysis. J. Sex Res., 55(4–5), 435–456. doi:https://doi.org/10.1080/00224499.2017.1387755.

Saha, S., Lim, C.C.W., Cannon, D.L., Burton, L., Bremner, M., Cosgrove, P., Huo, Y., J, J. M, 2021. Co-morbidity between mood and anxiety disorders: a systematic review and meta-analysis. Depress. Anxiety 38 (3), 286–306. https://doi.org/10.1002/da.23113.

Smirnova, D., Cumming, P., Sloeva, E., Kuvshinova, N., Romanov, D., Nosachev, G., 2018. Language patterns discriminate mild depression from Normal sadness and euthymic state. Front. Psych. 9, 105. https://doi.org/10.3389/fpsyt.2018.00105.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J., Patel, V., Silove, D., 2014. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int. J. Epidemiol. 43 (2), 476–493. https://doi.org/10.1093/ije/dyu038.

Stevanovic, D., Jafari, P., Knez, R., Franic, T., Atilola, O., Davidovic, N., Bagheri, Z., Lakic, A., 2017. Can we really use available scales for child and adolescent

psychopathology across cultures? A systematic review of cross-cultural measurement invariance data [review]. Transcult. Psychiatry 54 (1), 125–152. https://doi.org/10.1177/1363461516689215.

Svetina, D., Rutkowski, L., 2017. Multidimensional Measurement Invariance in an International Context: Fit Measure Performance with Many Groups. Sage Publications. https://doi.org/10.1177/0022022117717028.

Svetina, D., Rutkowski, L., Rutkowski, D., 2019. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semTools packages. Struct. Equ. Model. Multidiscip. J. 27, 1–20. https://doi.org/10.1080/10705511.2019.1602776.

Torres, L., Miller, M.J., Moore, K.M., 2013. Factorial invariance of the brief symptom Inventory-18 (BSI-18) for adults of Mexican descent across nativity status, language format, and gender. Psychol. Assess. 25 (1), 300–305. https://doi.org/10.1037/a0030436.

Twenge, J.M., 2011. Generational differences in mental health: are children and adolescents suffering more, or less? Am. J. Orthopsychiatry 81 (4), 469–472. https://doi.org/10.1111/j.1939-0025.2011.01115.x.

Twenge, J.M., 2015. Time period and birth cohort differences in depressive symptoms in the U.S., 1982–2013. Soc. Indic. Res. 121 (2), 437–454. https://doi.org/10.1007/s11205-014-0647-1.

UNDP. (2023). Human Development Index (HDI). https://hdr.undp.org/data-center/human-development-index#/indicies/HDI.

Uysal-Bozkir, Ö., Parlevliet, J.L., De Rooij, S.E., 2013. Insufficient cross-cultural adaptations and psychometric properties for many translated health assessment scales: a systematic review [review]. J. Clin. Epidemiol. 66 (6), 608–618. https://doi.org/10.1016/j.jclinepi.2012.12.004.

van de Schoot, R., Lugtig, P., Hox, J., 2012. A checklist for testing measurement invariance. Eur. J. Dev. Psychol. 9 (4), 486–492. https://doi.org/10.1080/17405629.2012.686740.

Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ. Res. Methods 3 (1), 4–70. https://doi.org/10.1177/109442810031002.

Wang, J., Kelly, B.C., Booth, B.M., Falck, R.S., Leukefeld, C., Carlson, R.G., 2010. Examining factorial structure and measurement invariance of the brief symptom inventory (BSI)-18 among drug users [article]. Addict. Behav. 35 (1), 23–29. https://doi.org/10.1016/j.addbeh.2009.08.003.

Weersing, V., Gonzalez, A., Hatch, B., Lynch, F., 2022. Promoting racial/ethnic equity in psychosocial treatment outcomes for child and adolescent anxiety and depression. Psychiatric Research and Clinical Practice 4 (3), 80–88. https://doi.org/10.1176/appi.prcp.20210044.

Weinandy, J.T.G., Lee, B., Hoagland, K.C., Grubbs, J.B., Bothe, B., 2023. Anxiety and compulsive sexual behavior disorder: a systematic review. J. Sex Res. 60 (4), 545–557. https://doi.org/10.1080/00224499.2022.2066616.

WHO. (2017). *Depression and other common mental disorders: global health estimates* (World Health Organization, Issue CC BY-NC-SA 3.0 IGO). https://apps.who.int/iris/handle/10665/254610.

WHO. (2022a). *ICD-11: International classification of diseases* (11 ed.). https://icd.who.int/.

WHO. (2022b). *World mental health report: Transforming mental health for all* (World Health Organization, Issue. https://www.who.int/publications/i/item/9789240049338.

Wiesner, M., Chen, V., Windle, M., Elliott, M.N., Grunbaum, J.A., Kanouse, D.E., Schuster, M.A., 2010. Factor structure and psychometric properties of the brief symptom Inventory-18 in women: a MACS approach to testing for invariance across racial/ethnic groups. Psychol. Assess. 22 (4), 912–922. https://doi.org/10.1037/a0020704.

Wittgens, C., Fischer, M.M., Buspavanich, P., Theobald, S., Schweizer, K., Trautmann, S., 2022. Mental health in people with minority sexual orientations: a meta-analysis of population-based studies. Acta Psychiatr. Scand. 145 (4), 357–372. https://doi.org/10.1111/acps.13405.

Yeung, A.Y., Yuliawati, L., Cheung, S.-H., 2020. A systematic review and meta-analytic factor analysis of the depression anxiety stress scales. Clin. Psychol. Sci. Pract. 27 (4), e12362 https://doi.org/10.1111/cpsp.12362.