



# Dimensionality and reliability of the online version of the Patient Health Questionnaire-4 (PHQ-4) in a large Colombian sample: Results from the PSY-COVID study

Juan P. Sanabria-Mazo<sup>1,2,3</sup> · Andrés Gómez-Acosta<sup>4</sup> · John Alexander Castro-Muñoz<sup>5</sup> · Yonatan Ferney Rojas<sup>5</sup> · Albert Feliu-Soler<sup>3,6</sup> · Juan V. Luciano<sup>2,3,6</sup> · Antoni Sanz<sup>1</sup>

Accepted: 28 August 2023  
© The Author(s) 2023

## Abstract

The Patient Health Questionnaire-4 (PHQ-4) is an ultra-brief screening instrument to measure depressive and anxiety symptoms. This study evaluated the dimensionality and reliability of the online version of the PHQ-4 in a large sample of the general population in Colombia. Data were collected during the first phase of lockdown measures occasioned by the COVID-19 pandemic. A total of 18,061 adult participants completed the online version of the PHQ-4. The characteristics of the items and subscales were explored. Dimensionality was examined using confirmatory factor analysis (CFA), including an examination of invariance (configural, metric, and scalar) across socio-demographic characteristics. Reliability indices were computed and known-groups validity was addressed by estimating associations between PHQ-4 scores and socio-demographic characteristics. The CFA showed significantly adequate fit indices for the expected two-factor structure, being invariant across gender, age, income level, education level, and region. Internal consistency was satisfactory for the PHQ-2 ( $\alpha = .83$ ), the GAD-2 ( $\alpha = .79$ ), and the PHQ-4 ( $\alpha = .86$ ). Higher scores on depressive (PHQ-2), anxiety (GAD-2), and psychological distress (PHQ-4) symptoms in females and young people, and those respondents with lower income, unemployed, and lower level of education were observed. The findings indicate that the PHQ-4 is a reliable scale for depressive and anxiety symptoms among adult Colombian people, being recommendable this tool for online surveys.

**Keywords** Patient Health Questionnaire-4 · Anxiety · Depression · COVID-19 pandemic · Psychometric properties

## Introduction

Anxiety and depression are the most prevalent emotional disturbances (Wittchen et al., 2011). The high comorbidity between them, close to 50% of cases (Kessler et al., 2015), as well as the impact of both mood disorders on the functioning and quality of life of affected individuals are priority problems for public health systems (Kroenke et al., 2009). The prevalence of anxiety disorders increased to 11% from 1990 to 2010, growing from 200 to 272 million reported cases worldwide (Baxter et al., 2014). A meta-analysis that examined 68 studies conducted in 30 countries, between 1994 and 2014, reported a prevalence of depression of around 13% in the general population (Lim et al., 2018). According to World Health Organization (2011) population reports estimate that by 2030 emotional disorders will be the most disabling mental health conditions worldwide. The COVID-19 pandemic has increased the prevalence of mood disorders in the clinical and non-clinical population (Luo

---

Juan V. Luciano and Antoni Sanz share senior authorship.

✉ Juan P. Sanabria-Mazo  
juanpablo.sanabria@uab.cat

- <sup>1</sup> Department of Basic, Developmental, and Educational Psychology, Autonomous University of Barcelona, Cerdanyola del Vallès, Spain
- <sup>2</sup> Teaching, Research, & Innovation Unit, Parc Sanitari Sant Joan de Déu, St. Boi de Llobregat, Spain
- <sup>3</sup> Centre for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain
- <sup>4</sup> Department of Psychology, Universidad de Pamplona, Pamplona, Colombia
- <sup>5</sup> Department of Psychology, Fundación Universitaria Sanitas, Bogotá, Colombia
- <sup>6</sup> Department of Clinical and Health Psychology, Autonomous University of Barcelona, Cerdanyola del Vallès, Spain

et al., 2020; Xiong et al., 2020), pointing to the relevance of reliable screening tools for the assessment and monitoring of symptoms severity.

In recent years, it has been highlighted the importance of brief, rapid, and reliable screening tools to facilitate diagnosis of mood disorders in healthcare settings (Olariu et al., 2015). Different authors have proposed the use of brief screening tools to reduce misdiagnosis (Castro-Rodríguez et al., 2015), to optimize health system resources (Cano-Vindel et al., 2018), and to improve clinical outcomes (Goldberg et al., 2017). Specifically, the Patient Health Questionnaire (PHQ-4) is one of the most widely used an ultra-brief screening instruments to measure depressive and anxiety symptoms (Kroenke et al., 2009). This ultra-brief self-report instrument combines two items of the PHQ-9 (Kroenke et al., 2001) and two items of the Generalized Anxiety Disorder (GAD-7; Spitzer et al., 2006). The psychometric properties of the PHQ-4 has been explored in clinical (Cano-Vindel et al., 2018; Ghaheri et al., 2020; Kroenke et al., 2009; Renovanz et al., 2019; Weihs et al., 2018) and non-clinical samples (Fong et al., 2023; Kazlauskas et al., 2023; Khubchandani et al., 2016; Kocalevent et al., 2014; Larionow & Mudło, 2023; Löwe et al., 2010; Meidl et al., 2023; Mendoza et al., 2022; Mills et al., 2015) in several countries (i.e., Colombia, Denmark, Ecuador, Germany, Greek, Iran, Korea, Philippines, Poland, Spain, and United States, among others), but mainly in paper-and-pencil format.

Online screening tools are reliable for the detection of mood disorders (Muñoz-Navarro et al., 2017a, b). These questionnaires facilitate data collection and help to avoid limitations of data loss in the classic paper-and-pencil format or response bias in face-to-face interviews. The evaluation of the psychometric properties of an online version is necessary, even if the paper version has been explored (Campbell et al., 2015; Coons et al., 2009; Mendoza et al., 2022). Recently, Cano-Vindel et al. (2018) tested the dimensionality, reliability, and validity of a computerized version of the PHQ-4 in a Spanish sample of 1052 patients from 28 primary care centers. Results indicated adequate internal consistency for depressive ( $\alpha = .86$ ) and anxiety ( $\alpha = .76$ ) symptoms. Even though the PHQ-4 has been standardized on a representative sample of 1500 people from the general Colombian population through face-to-face interviews (Kocalevent et al., 2014) –also with adequate properties–, there is no evidence of the psychometric properties of an online version of this instrument in other Spanish-speaking countries apart from Spain. This is the first study that provides information on the online version of the PHQ-4 in a large sample of the Colombian population during the COVID-19 pandemic.

As far as it is known, the goodness-of-fit of a bifactorial structure for the PHQ-4 has not been tested. Recently, Tibubos et al. (2021) evaluated the internal structure of

the PHQ-9 using confirmatory factor analysis (CFA). The bifactor model yielded an excellent fit to the data, being superior to that obtained in the one- and two-factor models. Two types of latent factors are defined in bifactor models. The first is a general factor in which all items are allowed to load (i.e., PHQ-4) and the second is composed of specific factors in which the items are distributed by their content (i.e., PHQ-2 and GAD-2). In the case of the PHQ-4 –and following the Clark and Watson's tripartite model (1991)–, the general factor reflects the shared component of depression and anxiety (i.e., psychological distress; Drapeau et al., 2012), whereas the specific factors (depression and anxiety after controlling for the general negative affect factor) represent low positive affect (for depression; Kroenke et al., 2009) and hyperarousal (for anxiety; Kroenke et al., 2009).

For all these reasons, it seems relevant to study the psychometric properties of the PHQ-4. Particularly, this study evaluated the dimensionality and reliability of the online version of the PHQ-4 in a large sample of the general population in Colombia collected during the first phase of lockdown measures occasioned by the COVID-19 pandemic (i.e., May to June 20, 2020). The four objectives and hypotheses explored in this research are presented below:

1. Firstly, to examine the goodness-of-fit of the one-, two-, and bifactor model of the PHQ-4. In line with the evidence reported in previous studies (Cano-Vindel et al., 2018; Fong et al., 2023; Kocalevent et al., 2014; Kroenke et al., 2009; Löwe et al., 2010; Meidl et al., 2023; Mendoza et al., 2022), it was speculated that the two-factor correlated model would have a significantly better fit to the data than the other models (hypothesis 1).
2. Secondly, to test invariance (configural, metric, and scalar) of the best-fitting model across socio-demographic characteristics. As in previous validations (Kocalevent et al., 2014; Larionow & Mudło, 2023; Löwe et al., 2010; Mendoza et al., 2022), it was expected that the dimensions were invariant across gender, age, income level, education level, and region (hypothesis 2).
3. Thirdly, to explore the reliability of the PHQ-4 subscales through different reliability indexes (i.e., Cronbach's  $\alpha$ , McDonald's  $\omega$ , and Guttman's  $\lambda^2$ ). The PHQ-2, the GAD-2, and the PHQ-4 subscales were expected to have the capacity to reliably measure anxiety, depression, and psychological distress beyond the reliability index examined (hypothesis 3).
4. Fourthly, to explore the relationship between the PHQ-4 scores with socio-demographic characteristics of this sample. Based on the results from previous psychometric studies (Cano-Vindel et al., 2018; Fong et al., 2023; Kocalevent et al., 2014; Löwe et al., 2010; Meidl et al., 2023; Mendoza et al., 2022), it was expected that

females, older individuals, those with lower incomes, employed, or with lower levels of education would exhibit higher depressive, anxiety, and psychological distress symptoms (hypothesis 4).

## Method

### Study design

Data analyses of the online version of the PHQ-4 were conducted using the database of the PSY-COVID study in Colombia (Sanabria-Mazo & Sanz, 2021). PSY-COVID is a cross-sectional study that aimed to assess the psychosocial impact of the COVID-19 pandemic in 30 countries. Specifically, this article explored data from the general population residing in Colombia during the first phase of the lockdown measures. Using the database of PSY-COVID, two previous studies have been published on the impact of COVID-19 lockdown measures on the mental health in the Colombian population (see Sanabria-Mazo et al., 2021a, b).

### Participants

In total, 18,833 people completed the online questionnaire in Colombia, of which 772 were excluded from this analysis because they resided in other countries during the first wave of COVID-19 pandemic. Finally, the sample consisted of 18,061 participants from all regions of the country. As shown in Table 1, majority of the participants were female (75%), adults between 25–34 years old (30%), with medium income levels (62%), with university education level (90%), and resided in the Andean region (52%). Inclusion criteria were adults ( $\geq 18$  years old) residing in Colombia during the period in which the data were collected (see Table 1). No participants who met the eligibility criteria were excluded from the analyses.

### Procedure

Administration of an anonymous online questionnaire generated with Google Forms® was carried out using a non-probabilistic sampling (snowball method) from May 20<sup>th</sup> to 20 June 20<sup>th</sup>, 2020. The survey was distributed through social networks, media, and institutional contacts. A panel of 30 international experts in clinical and health psychology validated the online questionnaire. The instruments used in this online survey were piloted prior to administration. No economic incentives were offered to participants for responding to this anonymous survey. The time to complete the socio-demographic and PHQ-4 items were approximately 3 min. More information on the procedure of the PSY-COVID study in Colombia is available in Sanabria-Mazo et al., (2021a, b).

**Table 1** Socio-demographic characteristics of the sample

Variables, <i>n</i> (%)	Sample ( <i>n</i> = 18,061)
Gender	
Female	13,491 (75.0)
Male	4,495 (25.0)
Age groups	
18–24 years	4,564 (25.3)
25–34 years	5,426 (30.0)
35–44 years	3,865 (21.4)
45–54 years	2,411 (13.3)
$\geq 55$ years	1,795 (9.9)
Income level	
Low	5,173 (28.6)
Medium	11,186 (61.9)
High	1,702 (9.4)
Work status	
Employed	10,872 (60.3)
Unemployed	7,162 (39.7)
Education level	
Primary	322 (1.8)
Secondary	1,536 (8.5)
University	16,185 (89.7)
Region	
Amazon	285 (1.6)
Andean	9,347 (51.9)
Caribbean	1,395 (7.8)
Orinoco	557 (3.1)
Pacific	6,420 (35.7)

*n* = frequency, % = percentage

This research was approved by the Ethical Committee on Animal and Human Experimentation of the Autonomous University of Barcelona (CEEAH-5197).

### Measures

The *Patient Health Questionnaire-4* (PHQ-4) was used to measure depressive and anxiety symptoms (Löwe et al., 2010). The two items of PHQ-2 correspond to the symptoms of the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) diagnostic criteria for major depressive disorder (i.e., loss of interest and depressive mood) and the two items of GAD-2 to the symptoms of generalized anxiety disorder (i.e., nervousness and worries). This version contains 4 items with a 4-point Likert response format, where 0 corresponds to "not at all" and 3 to "nearly every day", and questions are asked in the time frame of the last two weeks. The total score of the PHQ-4 (psychological distress) ranges from 0 to 12 and the specific score of its two subscales (PHQ-2 and GAD-2) ranges from 0 to 6. The cut-off points for detecting probable cases of depression (PHQ-2) or anxiety (GAD-2) is 3 or more for each subscale; and for

probable cases of psychological distress (PHQ-4) is 6 or more for total scale (Cano-Vindel et al., 2018; Kocalevent et al., 2014; Löwe et al., 2010; Mendoza et al., 2022).

In addition, a socio-demographic information questionnaire was included to collect data about gender (female and male), age, income level (low, medium, and high), work status (employed and unemployed), educational level (primary, secondary, and university), and region of residence (Amazon, Andean, Caribbean, Orinoco, and Pacific).

## Data analyses

Initially, descriptive statistics were conducted for socio-demographic variables as frequencies ( $n$ ) and percentages (%). Characteristics of the PHQ-4 were explored, including item means and standard deviations (SD), skewness and kurtosis, corrected item-total correlations, among items in each subscale, and between items of different subscales. Given the brevity of the scales, these correlations were analyzed using the Spearman-Brown correction. No outliers were identified in the analyses and none of the participants were excluded due to missing data. The dimensionality of the PHQ-4 was examined through CFA, using maximum likelihood as the estimation method. Regarding dimensionality, it tested a (1) one-factor model with the four items loading on the latent factor; (2) two-factors model including two correlated dimensions; and (3) bifactor model with the four items saturated with a global latent factor of psychological distress plus two uncorrelated specific factors of anxiety and depression.

The Tucker–Lewis's Index (TLI), Normed Fit Index (NFI), and Comparative Fit Index (CFI) were used to evaluate goodness-of-fit, with  $> .90$  confidence intervals and the Root Mean Square Error of Approximation (RSMEA)  $< .08$ , according to Schermelleh-Engel et al. (2003). The invariance (configural, metric, and scalar) of the models were tested by gender, age, income level, education level, and region in

comparable subsamples with random assignment. The configural invariance provides evidence on the consistency of the factor structure of the model across groups, the metric invariance on the factor loadings of the items across groups, and the scalar invariance on the equality of the mean scores across groups (Van de Schoot et al., 2012). In addition, to determine measurement invariance, the multigroup CFA was conducted, observing a change of  $\Delta CFI$  that is less than or equal to .01, according with Chen's (2007).

To estimate reliability, both total internal consistency of the scale (PHQ-4) and the subscales (PHQ-2 and GAD-2) were assessed through Cronbach's  $\alpha$ , McDonald's  $\omega$ , and Guttman's  $\lambda^2$ . Furthermore, a known-groups validity approach was used to estimate associations between PHQ-4 scores and socio-demographic characteristics that have been reported in the literature as risk factors for depression and anxiety. For this purpose, univariate group comparisons were performed with the PHQ-2, the GAD-2, and the PHQ-4 scores as dependent variables through  $t$ -tests and analysis of variance (ANOVA), considering the Bonferroni adjustment for multiple testing. Statistical analyses were performed with SPSS-26®, AMOS-5, R Studio, and JASP®.

## Results

### Item and scale characteristics

Table 2 shows descriptive analyses of the items, subscales (PHQ-2 and GAD-2) and total scale (PHQ-4). Mean (SD) score of PHQ-2 was 2.28 (1.61), GAD-2 was 2.01 (1.67), and PHQ-4 was 4.29 (3.01). Corrected item-total correlations ranged from  $r = .62$  to  $r = .77$ . Correlation between the two items of PHQ-2 was  $r = .64$  and between the two items of GAD-2 was  $.68$ , while correlation of the items with the items of the other subscale ranged from  $r = .46$  to  $r = .68$ . PHQ-2 and GAD-2 had a correlation of  $r = .64$ ,

**Table 2** Characteristics of the items and subscales of the PHQ-4

Items	$M$ (95% CI)	$SD$	Skewness	Kurtosis	Reliability		
					$\alpha$	$\omega$	$\lambda^2$
Depression (PHQ-2)							
1. Little interest or pleasure in doing things	1.24 (1.23–1.25)	0.88	.50	-.37			
2. Feeling down, depressed, or hopeless	1.04 (1.03–1.05)	0.89	.66	-.21			
PHQ-2 total score	2.28 (2.26–2.30)	1.61	.61	-.02	.79	.81	.80
Anxiety (GAD-2)							
1. Feeling nervous, anxious, or on edge	1.04 (1.03–1.05)	0.89	.64	-.23			
2. Not being able to stop or control worrying	0.97 (0.95–0.98)	.92	.72	-.30			
GAD-2 total score	2.01 (1.98–2.03)	1.67	.74	-.10	.83	.83	.82
Distress (PHQ-4)							
PHQ-4 total score	4.29 (4.24–4.33)	3.01	.73	-.03	.86	.86	.86

indicating high overlap between subscales. All the above correlations were statistically significant ( $p < .01$ ).

### Dimensionality

As shown in Table 3, the fit indices for the correlated two-factor model were significantly better ( $p < .001$ ;  $X^2 = 938.24$ ) than those obtained for the one-factor model [CFI (0.99 vs. 0.94), TLI (0.99 vs. 0.83), NFI (0.99 vs. .94), and RMSEA (0.04 vs. 0.23)], which provides strong support for the adequacy of the original model proposed by Kroenke et al. (2009). The bifactor structure was tested, although no convergence was found for this model.

Regarding factor loadings of the tested factor models, in the two-factor model ranged between .71 and .92, and those of the one-factor model between .68 and .83 (see Fig. 1). The results slightly differed between the specific factors of anxiety and depression. In line with hypothesis 1, these results confirm that the two-factor correlated model have a significantly better fit to the data than the other models.

Comparable subsamples with random assignment were used to test the invariance (configural, metric, and scalar) of the two-factor correlated model by gender (female:  $n = 4,305$ ; male:  $n = 4,295$ ), age ( $\leq 32$  years:  $n = 9,169$ ;  $> 32$  years:  $n = 8,892$ ), income level (low:  $n = 1,173$ ; medium:  $n = 1,186$ ; high:  $n = 1,102$ ), education level (primary:  $n = 340$ ; secondary:  $n = 436$ ; university:  $n = 485$ ), and region (Amazon:  $n = 285$ ; Andean:  $n = 447$ ; Caribbean:  $n = 345$ ; Orinoco:  $n = 328$ ; Pacific:  $n = 420$ ). Table 4 shows that no structural differences were identified the best-fitting model according to gender, age, income level, education level, and region with a  $\Delta$  CFI lower than .01, which confirm hypothesis 2.

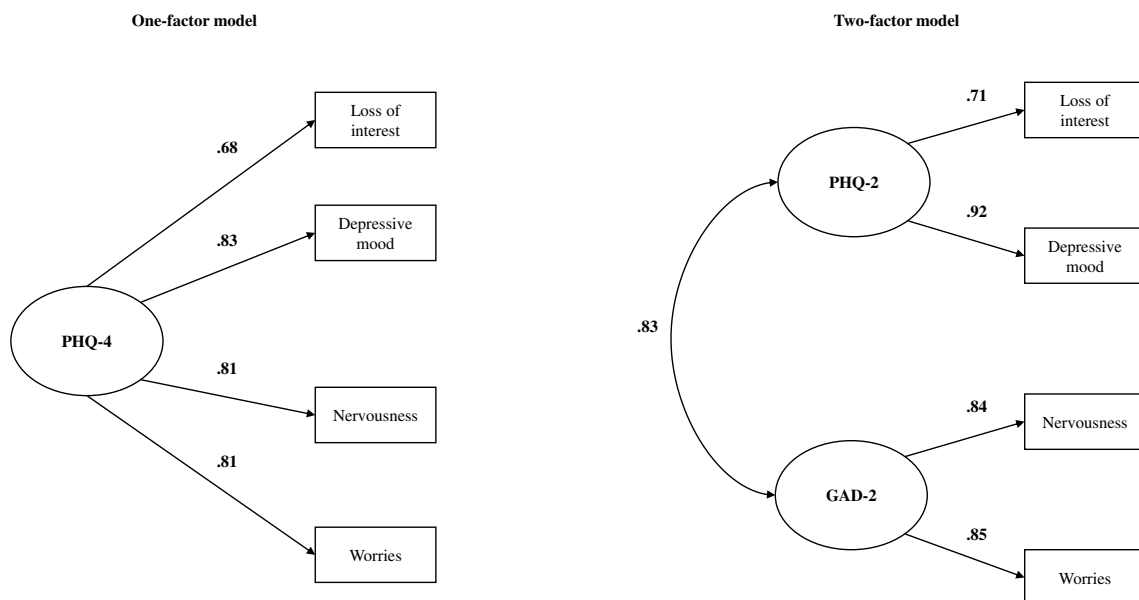
### Reliability

Reliability of the PHQ-2 ( $\alpha = .79$ ,  $\omega = .81$ , and  $\lambda^2 = .80$ ), the GAD-2 ( $\alpha = .83$ ,  $\omega = .83$ , and  $\lambda^2 = .82$ ), and the PHQ-4 ( $\alpha = .86$ ,  $\omega = .86$ , and  $\lambda^2 = .86$ ) was above .78 on all calculated indicators. The adequate reliability indices for depression, anxiety, and psychological distress confirms hypothesis 3.

**Table 3** Confirmatory factor analysis (CFA) comparing fit indices of the one-factor and two-factor model of PHQ-4

	SB $X^2$	df	CFI	TLI	NFI	RMSEA	$X^2$	$p$
One-factor model	1960.335	1	.94	.83	.94	.23	938.24	< .001
Two-factor model	33.528	6	.99	.99	.99	.04		

The bifactor structure was tested, although no convergence was found for this model



**Fig. 1** Factor loadings for the one- and two-factor models of the PHQ-4

**Table 4** Test for configural invariance across gender, age, income level, education level, and region using multi-group CFA

	X <sup>2</sup> (df)	ΔP	CMIN/df	CFI	Δ CFI	RMSEA	Δ RMSEA	Invariance
Across gender								
Groups								
Female ( <i>n</i> = 4,305)	13.772 (2)	–	6.886	.999	–	.036	–	–
Male ( <i>n</i> = 4,295)	12.054 (2)	–	6.027	.999	–	.033	–	–
Multigroup analysis								
Configural model	32.289 (2)	–	–	.999	–	.029	–	Invariant
Metric model	33.688 (4)	< .01	8.422	.999	.000	.020	.009	Invariant
Scalar model	175.090 (8)	< .01	21.886	.995	.004	.034	.014	Invariant
Across age								
Groups								
≤ 32 years ( <i>n</i> = 9,169)	15.143 (1)	–	15.143	.999	–	.039	–	–
> 32 years ( <i>n</i> = 8,892)	13.543 (1)	–	13.543	.999	–	.038	–	–
Multigroup analysis								
Configural model	28.685 (2)	–	14.343	.999	–	.027	–	Invariant
Metric model	34.327 (4)	< .01	8.582	.999	.000	.020	.007	Invariant
Scalar model	1444.32 (8)	< .01	180.540	.991	.008	.035	.015	Invariant
Across income level								
Groups								
Low ( <i>n</i> = 1,173)	17.424 (1)	–	17.424	.998	–	.056	–	–
Medium ( <i>n</i> = 1,186)	11.049 (1)	–	11.049	1.00	–	.030	–	–
High ( <i>n</i> = 1,102)	2.900 (1)	–	2.900	.999	–	.033	–	–
Multigroup analysis								
Configural model	31.370 (3)	–	–	.999	–	.023	–	Invariant
Metric model	51.474 (7)	< .01	7.35	.999	.000	.019	.004	Invariant
Scalar model	586.068 (15)	< .01	39.5071	.994	.005	.046	.027	Invariant
Across education level								
Groups								
Primary ( <i>n</i> = 340)	17.424 (1)	–	17.424	.998	–	.056	–	–
Secondary ( <i>n</i> = 436)	11.049 (1)	–	11.049	1.00	–	.030	–	–
University ( <i>n</i> = 485)	2.900 (1)	–	2.900	.999	–	.033	–	–
Multigroup analysis								
Configural model	31.370 (3)	–	–	.999	–	.023	–	Invariant
Metric model	51.474 (7)	< .01	7.35	.999	.000	.019	.004	Invariant
Scalar model	586.068 (15)	< .01	39.5071	.989	.010	.046	.027	Invariant
Across region								
Groups								
Amazon ( <i>n</i> = 285)	.005 (1)	–	0.005	1.00	–	.000	–	–
Andean ( <i>n</i> = 447)	5.609 (1)	–	5.609	1.00	–	.022	–	–
Caribbean ( <i>n</i> = 345)	8.369 (1)	–	8.369	.997	–	.073	–	–
Orinoco ( <i>n</i> = 328)	.123 (1)	–	.123	1.00	–	.000	–	–
Pacific ( <i>n</i> = 420)	28.147 (1)	–	28.147	.998	–	.065	–	–
Multigroup analysis								
Configural model	42.253 (5)	–	–	.999	–	.020	–	Invariant
Metric model	55.012 (13)	< .01	4.232	.999	.000	.013	.007	Invariant
Scalar model	157.242 (29)	< .01	5.422	.996	.003	.016	.003	Invariant

### Known groups validity

As shown in Table 5, statistically significant differences were found in the PHQ-2, the GAD-2, and the PHQ-4 scores according to gender, age, income level, work status, and educational level, but with small effect sizes ( $d < 0.2$  and  $\eta^2 < 0.12$ ). Females, younger age, unemployed, and those with lower incomes and educational levels reported the higher depression (PHQ-2), anxiety (GAD-2), and psychological distress (PHQ-4) scores. The higher scores for females, and those with lower incomes and educational levels were consistent with hypothesis 4. However, they were inconsistent with the higher scores expected for older age and employed. For more information, the prevalence of depressive (35%) and anxiety (29%) symptoms can be read in detail in Sanabria-Mazo et al. (2021a).

### Discussion

The findings of this study provide evidence that the online version of PHQ-4 is a reliable ultra-brief self-administered instrument for measuring depressive and anxiety symptoms in the general population in Colombia. Previous studies have demonstrated the validity of the classic

paper-and-pencil format of the PHQ-4 in clinical (Ghaheri et al., 2020; Kroenke et al., 2009; Renovanz et al., 2019; Weihs et al., 2018) and non-clinical samples (Khubchandani et al., 2016; Kocalevent et al., 2014; Löwe et al., 2010; Mills et al., 2015), as well as non-clinical samples in online version during the first few months of the COVID-19 outbreak in Philippines (Mendoza et al., 2022). However, as far as it is known, this is the first study to evaluate the dimensionality and reliability of an online version of the PHQ-4 in a large sample of the general population in Colombia during the COVID-19 pandemic.

Consistent with previous research carried out in the classic paper-and-pencil and face-to-face interviews (Kocalevent et al., 2014; Kroenke et al., 2009; Löwe et al., 2010), CFA indicates that the two-factor structure (i.e., depression and anxiety) of the online PHQ-4 performs significantly better than the one-factor structure (i.e., psychological distress), with excellent fit indices on all parameters (CFI = .99, TLI = .99, NFI = .99, RMSEA = .04). Furthermore, these findings were consistent with those reported in another recent study of the PHQ-4 administered online (Mendoza et al., 2022). The bifactorial model did not converge probably due to the small number of indicators per latent variable. Therefore, these results demonstrate that the two-factor

**Table 5** Association PHQ-4 scores and socio-demographic characteristics

Variables	PHQ-2 (0–6)		GAD-2 (0–6)		PHQ-4 (0–12)	
	Score <i>M</i> ( <i>SD</i> )	Difference groups <i>p</i> value (effect size)	Score <i>M</i> ( <i>SD</i> )	Difference groups <i>p</i> value (effect size)	Score <i>M</i> ( <i>SD</i> )	Difference groups <i>p</i> value (effect size)
Gender		< .001 ( $d=0.15$ )		< .001 ( $d=0.19$ )		< .001 ( $d=0.19$ )
Female	2.34 (1.60)		2.08 (1.68)		4.42 (3.00)	
Male	2.09 (1.63)		1.77 (1.63)		3.86 (2.99)	
Age groups		< .001 ( $\eta^2=0.09$ )		< .001 ( $\eta^2=0.03$ )		< .001 ( $\eta^2=0.07$ )
1. 18–24 years	2.98 (1.65)	1 > 2 > 3 > 4 > 5	2.43 (1.81)	1 > 2 > 3 > 4 > 5	5.41 (3.14)	1 > 2 > 3 > 4 > 5
2. 25–34 years	2.38 (1.55)		2.04 (1.68)		4.42 (2.96)	
3. 35–44 years	1.99 (1.50)		1.86 (1.57)		3.85 (2.82)	
4. 45–54 years	1.76 (1.45)		1.70 (1.54)		3.46 (2.75)	
5. ≥ 55 years	1.53 (1.37)		1.55 (1.40)		3.08 (2.51)	
Income level		< .001 ( $\eta^2=0.03$ )		< .001 ( $\eta^2=0.01$ )		< .001 ( $\eta^2=0.02$ )
1. Low	2.69 (1.70)	1 > 2 > 3	2.27 (1.80)	1 > 2 > 3	4.95 (3.23)	1 > 2 > 3
2. Medium	2.15 (1.54)		1.92 (1.61)		4.08 (2.88)	
3. High	1.88 (1.54)		1.76 (1.55)		3.63 (2.81)	
Work status		< .001 ( $d=0.15$ )		< .001 ( $d=0.08$ )		< .001 ( $d=0.13$ )
Employed	2.18 (1.57)		1.95 (1.64)		4.13 (2.94)	
Unemployed	2.43 (1.65)		2.09 (1.72)		4.52 (3.09)	
Education level		= .002 ( $\eta^2 < 0.01$ )		= .004 ( $\eta^2 < 0.01$ )		= .001 ( $\eta^2 = 0.01$ )
1. Primary	2.25 (1.74)		2.06 (1.80)		4.32 (3.28)	
2. Secondary	2.42 (1.71)	2 > 3 > 1	2.14 (1.82)	2 > 3 > 1	4.55 (3.26)	2 > 3 > 1
3. University	2.27 (1.60)		1.99 (1.65)		4.26 (2.98)	

Between-group differences were calculated through t-test and ANOVA, considering the Bonferroni adjustment for multiple testing

correlated model is the best fit to the data, confirming hypothesis 1.

The high correlation between the depression (PHQ-2) and anxiety (GAD-2) subscales was like those reported in previous studies (Kroenke et al., 2009; Löwe et al., 2010; Mendoza et al., 2022). Comorbidity between these mood disorders, close to 50% of cases (Kessler et al., 2015), theoretically explains the high correlation identified between both subscales of the PHQ-4 (Kocalevent et al., 2014; Kroenke et al., 2009; Löwe et al., 2010; Mendoza et al., 2022). The structure of two factors (depression and anxiety) that share a higher order factor (psychological distress) is consistent with the conception of two nosological entities clearly differentiated by the causal cognitive processes and their clinical manifestations (Clark & Watson, 1991) and with the extensive empirical evidence of the high comorbidity of both disorders (Kessler et al., 2015).

As in previous validations (Fong et al., 2023; Kocalevent et al., 2014; Löwe et al., 2010; Meidl et al., 2023; Mendoza et al., 2022), the two-factor structure of the PHQ-4 was invariant (configural, metric, and scalar) across gender, age, income level, education level, and region in this study, supporting hypothesis 2. Given the geographically based cultural heterogeneity of Colombia, regional invariance of PHQ-4 seems to be a particularly relevant finding. Despite the criticisms about the real relevance of the invariance of psychological assessment instruments (Welzel et al., 2021), there is certainly a great consensus that in cross-cultural studies it is necessary to guarantee its homogeneous behavior. Recently, the results of a cross-cultural research conducted in 7 European countries (Austria, Croatia, Georgia, Germany, Lithuania, Portugal, and Sweden) have shown that the PHQ-4 could be generalized to other countries and cultures (Kazlauskas et al., 2023).

With regard to reliability, internal consistency values were slightly higher than those reported in other psychometric studies (Cano-Vindel et al., 2018; Ghaheri et al., 2020; Khubchandani et al., 2016; Kocalevent et al., 2014; Löwe et al., 2010; Mendoza et al., 2022), with values close to  $\alpha = .83$  for depression (PHQ-2),  $\alpha = .79$  for anxiety (GAD-2), and  $\alpha = .86$  for psychological distress (PHQ-4), which supports hypothesis 3. Although small effect sizes were obtained, the findings of this study provide further evidence about gender, age, income level, work status, and educational level role as risk factors for depression and anxiety. In line with other studies and hypothesis 4, it was identified that people of female gender, low income, and low education levels reported higher scores on depression, anxiety, and psychological distress (Khubchandani et al., 2016; Kocalevent et al., 2014; Löwe et al., 2010; Mendoza et al., 2022; Mills et al., 2015). In contrast, it was found that younger people and unemployed reported higher scores than people who were older (Löwe et al., 2010) and employed (Kocalevent

et al., 2014), results in line a vast amount of empirical evidence reflecting the negative effect of the COVID-19 pandemic on the mental health (Hossain et al., 2020), a phenomenon also observed in other studies on Colombian population (Caballero-Domínguez et al., 2022; Cénat et al., 2022; Gómez-Restrepo et al., 2022).

Regarding prevalence, 35% of the participants in this sample showed depressive symptoms and 29% anxiety symptoms (see Sanabria-Mazo et al., 2021a, b). Consistent with other research in the general population during the COVID-19 pandemic (Luo et al., 2020; Xiong et al., 2020), it was found that about one third of this sample showed depressive and anxiety symptoms. Specifically, the population groups most affected by the COVID-19 pandemic in Colombia were low-income people, students, and young adults, with depressive symptoms between 46 and 56% and anxiety symptoms between 36 and 40% (Sanabria-Mazo et al., 2021a). Compared with the results of a national survey, this study provided evidence that there were 2.5 to 2.8 times more people with risk of anxiety and 1.5 to 1.9 times more with risk of depression in the first wave of the COVID-19 outbreak (Sanabria-Mazo et al., 2021a). Although the above comparison is from a non-representative sample of the Colombian population, the reported differences highlight the need to prioritize prevention, intervention, and monitoring of symptoms related to emotional disorders.

## Limitations

These findings should be interpreted considering the following limitations. First, the analyses were conducted based on a non-representative sample, which impedes the generalizability of the results to the general population of Colombia or other Spanish-speaking languages. Second, due to the cross-sectional design, it was not possible to calculate the test-retest reliability of the instrument. Third, although the convergent and divergent of the PHQ-4 has been demonstrated in previous studies (Kocalevent et al., 2014; Kroenke et al., 2009; Löwe et al., 2010), no other instruments were used to provide further evidence of construct validity. Fourth, diagnostic interviews were not considered as a procedure to verify criterion validity, making it not possible to provide further evidence on specificity and sensitivity for the optimal cut-off point (Mitchell et al., 2016; Plummer et al., 2016). Fifth, it was not possible to examine the responsiveness, the smallest detectable change, or the minimal clinically important difference for scoring the PHQ-4. Sixth, online data collection can have a negative impact on the representation of population groups with internet connection difficulties, lack of knowledge in the use of new technologies, and low literacy.

## Conclusions

In summary, this study provides further evidence on the dimensionality and reliability of an ultra-brief online screening instruments for the detection of depressive and anxiety symptoms. It also shows that presentation in its online format does not alter its psychometric properties and it is invariant across gender, age, income level, education level, and region. The existing results from the PHQ-2 and GAD-2 denote similar psychometric behavior to the full versions of the PHQ-9 and GAD-2. Although the PHQ-2 and GAD-2 are reliable subscales for rapid screening of depression and anxiety, the use of their full versions is recommended when all DSM-IV diagnostic criteria need to be assessed. In line with the proposal by Löwe et al. (2010), it is suggested to use the total scale as a global screening tool for psychological distress (PHQ-4), and the depression (PHQ-2) and anxiety (GAD-2) subscales for their discriminated detection. Finally, PHQ-4 is an ultra-brief screening tool that can help to optimize the time resources of health systems, especially during health crises.

**Acknowledgements** JPS-M has a PFIS predoctoral contract from the Institute of Health Carlos III (ISCIII; FI20/00034). AF-S acknowledges the funding from the Serra Hünter program (Generalitat de Catalunya; reference number UAB-LE-8015). The authors are grateful to the CIBER of Epidemiology and Public Health (CIBERESP CB22/02/00052; ISCIII) for its support.

**Author contributions** Juan P. Sanabria-Mazo: conceptualization, data curation, software, formal analysis, methodology, validation, visualisation, and writing—original draft. Andrés Gómez-Acosta: conceptualization, formal analysis, methodology, validation, and writing—review & editing. John Alexander Castro-Muñoz: conceptualization, validation, and writing—review & editing. Yonatan Ferney Rojas: methodology, formal analysis, and writing—review & editing. Albert Feliu-Soler: methodology and writing – review & editing. Juan V. Luciano: methodology and writing – review & editing. Antoni Sanz: funding acquisition, project administration, methodology, and writing—review & editing.

**Funding** Open Access Funding provided by Autonomous University of Barcelona. This research was funded by the Agency for Management of University and Research Grants (AGAUR; 2020PANDE00025).

**Data availability** The database of this study is available in Sanabria-Mazo and Sanz (2021).

## Declarations

**Conflict of interest** All authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine*, *44*(11), 2363. <https://doi.org/10.1017/S0033291713003243>
- Caballero-Domínguez, C. C., Pedrozo-Pupo, J. C., & Campo-Arias, A. (2022). Depression among COVID-19 survivors in Colombia. *Psychology, Health & Medicine*, *28*(6), 1–8. <https://doi.org/10.1080/13548506.2022.2085879>
- Cano-Vindel, A., Muñoz-Navarro, R., Medrano, L. A., Ruiz-Rodríguez, P., González-Blanch, C., Gómez-Castillo, M. D., Capafons-Bonet, A., Chacón, F., Santolayay, F., & PsicAP Research Group. (2018). A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional disorders in primary care. *Journal of Affective Disorders*, *234*(247), 255. <https://doi.org/10.1016/j.jad.2018.01.030>
- Campbell, N., Ali, F., Finlay, A. Y., & Salek, S. S. (2015). Equivalence of electronic and paper-based patient-reported outcome measures. *Quality of Life Research: An International Journal of Quality-of-Life Aspects of Treatment, Care, and Rehabilitation*, *24*(8), 1949–1961. <https://doi.org/10.1007/s11136-015-0937-3>
- Castro-Rodríguez, J. I., Olariu, E., Garnier-Lacueva, C., Martín-López, L. M., Pérez-Solà, V., Alonso, J., Forero, C. G., & INSAYD Investigators. (2015). Diagnostic accuracy and adequacy of treatment of depressive and anxiety disorders: A comparison of primary care and specialized care patients. *Journal of Affective Disorders*, *172*, 462–471. <https://doi.org/10.1016/j.jad.2014.10.020>
- Cénat, J. M., Farahi, S. M. M., Dalexis, R. D., Darius, W. P., Bekarkhanechi, F. M., Poisson, H., Broussard, C., Ukwu, G., Auguste, E., Nguyen, D. D., Sehabi, G., Furyk, S. E., Gedeon, A. P., Onesi, O., El Aouame, A. M., Khodabocus, S. N., Shah, M. S., & Labelle, P. R. (2022). The global evolution of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis of longitudinal studies. *Journal of Affective Disorders*, *315*, 70–95. <https://doi.org/10.1016/j.jad.2022.07.011>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316–336. <https://doi.org/10.1037/0021-843X.100.3.316>
- Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderking, W. R., Cella, D., & Basch, E. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value in Health*, *12*(4), 419–429. <https://doi.org/10.1111/j.1524-4733.2008.00470.x>
- Drapeau, A., Marchand, A., & Beaulieu-Prévost, D. (2012). Epidemiology of psychological distress. In L. Labate (Ed.), *Mental illnesses - Understanding, prediction, and control* (pp. 105–133). <https://doi.org/10.5772/30872>
- Fong, T. C., Ho, R. T., & Yip, P. S. (2023). Psychometric properties of the Patient Health Questionnaire-4 among Hong Kong young adults in 2021: Associations with meaning in life and suicidal

- ideation. *Frontiers in Psychiatry*, 14, 1138755. <https://doi.org/10.3389/fpsy.2023.1138755>
- Ghaheeri, A., Omani-Samani, R., Sepidarkish, M., Hosseini, M., & Maroufzadeh, S. (2020). The four-item patient health questionnaire for anxiety and depression: A validation study in infertile patients. *International Journal of Fertility and Sterility*, 14(3), 234–239. <https://doi.org/10.22074/ijfs.2020.44412>
- Goldberg, D. P., Reed, G. M., Robles, R., Minhas, F., Razzaque, B., Fortes, S., de Jesus, J., Pong-Lam, T., García, J. A., Gask, L., Dowell, A. C., Rosendal, M., Mbatia, J. K., & Saxena, S. (2017). Screening for anxiety, depression, and anxious depression in primary care: A field study for ICD-11 PHC. *Journal of Affective Disorders*, 213, 199–206. <https://doi.org/10.1016/j.jad.2017.02.025>
- Gómez-Restrepo, C., Cepeda, M., Torrey, W. C., Suarez-Obando, F., Uribe-Restrepo, J. M., Park, S., Jassir, M. P., Martínez, P., Castro, S. M., Aguilera, J., González, L., Chaparro, N., Gómez, A. M., Bell, K., & Marsch, L. A. (2022). Perceived access to general and mental healthcare in primary care in Colombia during COVID-19: A cross-sectional study. *Frontiers in Public Health*, 10, 3124. <https://doi.org/10.3389/fpubh.2022.896318>
- Hossain, M. M., Sultana, A., & Purohit, N. (2020). Mental health outcomes of quarantine and isolation for infection prevention: A systematic umbrella review of the global evidence. *SSRN*, 42, 3561265. <https://doi.org/10.2139/ssrn.3561265>
- Kazlauskas, E., Gelezelyte, O., Kvedaraite, M., Ajdukovic, D., Johannesson, K. B., Böttche, M., Bondjers, K., Dragan, M., Figueiredo-Braga, M., Grajewski, P., Anastassiou-Hadjicharalambous, X., Javakhishvili, J. D., Lioupi, C., Lueger-Schuster, B., Mouthaan, J., Bagaric, I. R., Sales, L., Schäfer, I., Soydas, S., ... Lotzin, A. (2023). Psychometric properties of the Patient Health Questionnaire-4 (PHQ-4) in 9230 adults across seven European countries: Findings from the ESTSS ADJUST study. *Journal of Affective Disorders*, 335, 18–23. <https://doi.org/10.1016/j.jad.2023.05.007>
- Kessler, R. C., Sampson, N. A., Berglund, P., Gruber, M. J., Al-Hamzawi, A., Andrade, L., Bunting, B., Demyttenaere, K., Florescu, S., de Girolamo, G., Gureje, O., He, Y., Hu, C., Huang, Y., Karam, E., Kovess-Masfety, V., Lee, S., Levinson, D., Medina Mora, M. E., ... Wilcox, M. A. (2015). Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. *Epidemiology and Psychiatric Sciences*, 24(3), 210–226. <https://doi.org/10.1017/S2045796015000189>
- Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J. A., & Anderson, J. (2016). The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Archives of Psychiatric Nursing*, 30(4), 457–462. <https://doi.org/10.1016/j.apnu.2016.01.014>
- Kocalevent, R. D., Finck, C., Jimenez-Leal, W., Sautier, L., & Hinze, A. (2014). Standardization of the Colombian version of the PHQ-4 in the general population. *BMC Psychiatry*, 14(1), 1–8. <https://doi.org/10.1186/1471-244X-14-205>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics*, 50(6), 613–621. <https://doi.org/10.1176/appi.psy.50.6.613>
- Larionow, P., & Mudlo, K. (2023). The Patient Health Questionnaire-4: Factor structure, measurement invariance, latent profile analysis of anxiety and depressive symptoms and screening results in polish adults. *Advances in Cognitive Research*, 19(2), 123–137. <https://doi.org/10.5709/acp-0384-9>
- Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports*, 8(1), 2861. <https://doi.org/10.1038/s41598-018-21243-x>
- Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, 122(1–2), 86–95. <https://doi.org/10.1016/j.jad.2009.06.019>
- Luo, M., Guo, L., Yu, M., Jiang, W., & Wang, H. (2020). The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public: A systematic review and meta-analysis. *Psychiatry Research*, 291, 113190. <https://doi.org/10.1016/j.psychres.2020.113190>
- Meidl, V., Dallmann, P., Leonhart, R., Brethauer, B., Busch, A., Kubosch, E. J., Wrobel, N., & Hirschmüller, A. (2023). Validation of the Patient Health Questionnaire-4 for longitudinal mental health evaluation in elite Para athletes. *PM&R*. <https://doi.org/10.1002/pmrj.13011>
- Mendoza, N. B., Frondoza, C. E., Dizon, J. I. W. T., & Buenconsejo, J. U. (2022). The factor structure and measurement invariance of the PHQ-4 and the prevalence of depression and anxiety in a Southeast Asian context amid the COVID-19 pandemic. *Current Psychology*, 1, 1–10. <https://doi.org/10.1007/s12144-022-02833-5>
- Mills, S. D., Fox, R. S., Pan, T. M., Malcarne, V. L., Roesch, S. C., & Sadler, G. R. (2015). Psychometric evaluation of the patient health questionnaire-4 in Hispanic Americans. *Hispanic Journal of Behavioral Sciences*, 37(4), 560–571. <https://doi.org/10.1177/0739986315608126>
- Mitchell, A. J., Yadegarfar, M., Gill, J., & Stubbs, B. (2016). Case finding and screening clinical utility of the patient health questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic meta-analysis of 40 studies. *British Journal of Psychiatry Open*, 2(2), 127–138. <https://doi.org/10.1192/bjpo.bp.115.001685>
- Muñoz-Navarro, R., Cano-Vindel, A., Medrano, L. A., Schmitz, F., Ruiz-Rodríguez, P., Abellán-Maeso, C., Font-Payeras, M. A., & Hermosilla-Pasamar, A. M. (2017a). Utility of the PHQ-9 to identify major depressive disorder in adult patients in Spanish primary care centers. *BMC Psychiatry*, 17(1), 291. <https://doi.org/10.1186/s12888-017-1450-8>
- Muñoz-Navarro, R., Cano-Vindel, A., Moriana, J. A., Medrano, L. A., Ruiz-Rodríguez, P., Agüero, L., Rodríguez-Enríquez, M., Pizà, M. R., & Ramírez-Manent, I. (2017b). Screening for generalized anxiety disorder in Spanish primary care centers with the GAD-7. *Psychiatry Research*, 256, 312–317. <https://doi.org/10.1016/j.psychres.2017.06.023>
- Olariu, E., Forero, C. G., Castro-Rodríguez, J. I., Rodrigo-Calvo, M. T., Álvarez, P., Martín-López, L. M., Sánchez-Toto, A., Adroher, N. D., Blasco-Cubedo, M. J., Vilagut, G., Fullana, M. A., & Alonso, J. (2015). Detection of anxiety disorders in primary care: A meta-analysis of assisted and unassisted diagnoses. *Depression and Anxiety*, 32(7), 471–484. <https://doi.org/10.1002/da.22360>
- Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic metaanalysis. *General Hospital Psychiatry*, 39, 24–31. <https://doi.org/10.1016/j.genhosppsych.2015.11.005>
- Renovanz, M., Soebianto, S., Tsakmaklis, H., Keric, N., Nadji-Ohl, M., Beutel, M., Ringel, F., Wollschlager, D., & Hickmann, A. K. (2019). Evaluation of the psychological burden during the early disease trajectory in patients with intracranial tumors by the ultra-brief Patient Health Questionnaire for Depression and Anxiety (PHQ-4). *Supportive Care in Cancer*, 27(12), 4469–4477. <https://doi.org/10.1007/s00520-019-04718-z>
- Sanabria-Mazo, J. P., & Sanz, A. (2021). *PSY-COVID database 1 (Colombia)*. <https://doi.org/10.6084/m9.figshare.13951775.v3>

- Sanabria-Mazo, J. P., Useche-Aldana, B., Ochoa, P. P., Rojas-Gualdrón, D. F., Mateo-Canedo, C., Carmona-Cervelló, M., Crespo-Puig, N., Selva-Olíd, C., Muro, A., Méndez-Ulrich, J. L., Feliu-Soler, A., Luciano, J. V., & Sanz, A. (2021a). Social inequities in the impact of COVID-19 lockdown measures on the mental health of a large sample of the Colombian population (PSY-COVID study). *Journal of Clinical Medicine*, *10*(22), 5297. <https://doi.org/10.3390/jcm10225297>
- Sanabria-Mazo, J. P., Useche-Aldana, B., Ochoa, P. P., Rojas-Gualdrón, D. F., & Sanz, A. (2021b). *Impacto de la pandemia de COVID-19 en la salud mental en Colombia*. Editorial CES. <https://www.colpsic.org.co/wp-content/uploads/2021/09/Libro-Impacto-de-la-pandemia-de-COVID-19-en-la-salud-mental-en-Colombia.pdf>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Tibubos, A. N., Otten, D., Zöller, D., Binder, H., Wild, P. S., Fleischer, T., Johar, H., Atasoy, S., Schulze, L., Ladwig, H. K., Schomerus, G., Linkohr, B., Grabe, H. J., Kruse, J., Schmidt, C. O., Münzel, T., König, J., Brähler, E., & Beutel, M. E. (2021). Bidimensional structure and measurement equivalence of the Patient Health Questionnaire-9: Sex-sensitive assessment of depressive symptoms in three representative German cohort studies. *BMC Psychiatry*, *21*(1), 1–13. <https://doi.org/10.1186/s12888-021-03234-x>
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- Weihls, K. L., Wiley, J. F., Crespi, C. M., Krull, J. L., & Stanton, A. L. (2018). Predicting future major depression and persistent depressive symptoms: Development of a prognostic screener and PHQ-4 cutoffs in breast cancer patients. *Psycho-Oncology*, *27*(2), 420–426. <https://doi.org/10.1002/pon.4472>
- Welzel, C., Brunkert, L., & Inglehart, R.F. (2021). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, *52*(3), 1368–1400. <https://doi.org/10.1177/0049124121995521>
- World Health Organization. (2011). *Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level*. World Health Organization.
- Wittchen, H. U., Jacobi, F., Rehm, J., Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., ... Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*, 655–679. <https://doi.org/10.1016/j.euroneuro.2011.07.018>
- Xiong, J., Lipsitz, O., Nasri, F., Lui, L., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., & McIntyre, R. S. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders*, *277*, 55–64. <https://doi.org/10.1016/j.jad.2020.08.001>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.