

# Adaptation and Validation of the Patient Health Questionnaire for Depression and Anxiety (PHQ-4) in a sample of Quechua-speaking Peruvians

## Adaptación y validación de la Patient Health Questionnaire for Depression and Anxiety (PHQ-4) en una muestra de quechua hablantes peruanos

Renzo Felipe Carranza Esteban<sup>1</sup>, Oscar Mamani-Benito<sup>2</sup>, Julio Cjuno<sup>3</sup>, Madona Tito-Betancur<sup>4</sup>, Tomás Caycho-Rodríguez<sup>5</sup>, Lindsey W. Vilca<sup>3</sup>, Julio Torales<sup>6,7,8</sup>, Iván Barrios<sup>6,8</sup>

<sup>1</sup>Universidad San Ignacio de Loyola, Lima, Perú

<sup>2</sup>Universidad Señor Sipán, Chiclayo, Peru

<sup>3</sup>Universidad Norbert Wiener, Lima, Perú

<sup>4</sup>Universidad Tecnológica del Perú, Arequipa, Perú

<sup>5</sup>Universidad Científica del Sur, Lima, Perú

<sup>6</sup>Universidad Nacional de Asunción, Facultad de Ciencias Médicas, Cátedra de Psicología Médica, San Lorenzo, Paraguay

<sup>7</sup>Universidad Nacional de Caaguazú, Instituto Regional de Investigación en Salud, Coronel Oviedo, Paraguay

<sup>8</sup>Universidad Sudamericana, Facultad de Ciencias de la Salud, Pedro Juan Caballero, Paraguay



Received: 24/10/2023

Revised: 02/11/2023

Accepted: 22/12/2023

### Corresponding author

Tomás Caycho-Rodríguez  
Universidad Científica del Sur,  
Perú  
[tcaycho@cientifica.edu.pe](mailto:tcaycho@cientifica.edu.pe)

### Responsible editor

Dr. Israel González

### Conflicts of interests

The authors declare that there is no conflict of interest.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

This article is published under [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



## ABSTRACT

**Introduction:** The Patient Health Questionnaire for Depression and Anxiety (PHQ-4) is an ultra-brief self-report measure of depression and anxiety symptoms. The PHQ-4 has not been adapted or validated in the Quechua-speaking population. **Objective:** This study aimed to adapt and validate the PHQ-4 in a sample of the Peruvian Quechua-speaking population with its Collao variant from Puno-Peru. **Methods:** A total of 221 Quechua speakers of the Collao variant (mean age 31.2 years, SD = 11.7, 47.1% male and 52.9% female) participated. Evidence of validity based on internal structure was evaluated using confirmatory factor analysis (CFA), reliability using the internal consistency method, and item difficulty and discrimination characteristics based on Item Response Theory (IRT). **Results:** The results indicated that all items of the PHQ-4 were relevant, consistent, and clear enough to adequately represent anxiety and depressive symptoms. The CFA results supported a one-factorial model, in which anxiety and depression items form a single latent variable with high reliability. All items of the PHQ-4 had good discrimination parameters, and the difficulty parameters indicated that the items were comparatively difficult. Therefore, a higher presence of the latent trait (anxiety and depression) is needed to answer higher response categories. **Discussion:** In conclusion, the PHQ-4 presented good psychometric properties as a rapid, reliable, and valid primary screening measure for Quechua speakers in need of in-depth assessment and symptom monitoring for the diagnosis and treatment of anxiety and depression.

**Keywords:** anxiety; depression; Quechua speaker; validity.

## RESUMEN

**Introducción:** El Patient Health Questionnaire for Depression and Anxiety (PHQ-4) es una medida de autoinforme ultrabreve de síntomas depresivos y ansiosos. El PHQ-4 no ha sido adaptado ni validado en la población quechua hablante. **Objetivo:** Este estudio tuvo como objetivo adaptar y validar el PHQ-4 en una muestra de población peruana quechua hablante en su variante Collao de Puno-Perú. **Metodología:** Participaron 221 personas hablantes de quechua variante collao (promedio de edad de 31,2 años, SD = 11,7, 47,1 % hombres y 52,9 % mujeres). Se evaluó la evidencia de validez basada en la estructura interna, mediante el análisis factorial confirmatorio (CFA), la fiabilidad con el método de consistencia interna y las características de dificultad y discriminación de los ítems en base a la Teoría de Respuesta al Ítem (IRT). **Resultados:** Los resultados indicaron que todos los ítems del PHQ-4 son relevantes, coherentes y claros para representar adecuadamente los síntomas de ansiedad y depresión. El resultado del AFC apoyó un modelo unifactorial, donde los ítems de ansiedad y depresión conforman una sola variable latente con una fiabilidad alta. Todos los ítems del PHQ-4 presentan buenos parámetros de discriminación y los parámetros de dificultad indican que los ítems son comparativamente difíciles. Por lo tanto, se necesita una mayor presencia del rasgo latente (ansiedad y depresión) para responder las categorías de respuesta más altas. **Discusión:** En conclusión, la PHQ-4 presentó buenas propiedades psicométricas como una medida de detección primaria rápida, confiable y válida para personas quechua hablantes que necesitan una evaluación profunda, monitoreo de los síntomas para diagnóstico y tratamiento para la ansiedad o depresión.

**Palabras clave:** ansiedad; depresión; quechua hablante; validez.

**How to cite this article:** Carranza Esteban RF, Mamani-Benito O, Cjuno J, Tito-Betancur M, Caycho-Rodríguez T, Vilca LW, et al. Adaptation and Validation of the Patient Health Questionnaire for Depression and Anxiety (PHQ-4) in a sample of Quechua-speaking Peruvians. Med. clín. soc. 2024;8(1):63-74

## INTRODUCTION

Mental disorders are one of the leading causes of the global health burden (1). In 2017, it was estimated that approximately 264 million and 322 million people worldwide were living with anxiety disorders and depression, respectively (2). In 2019, depression and anxiety were considered the two most disabling mental disorders and were among the top 25 causes of health-related burdens worldwide (3). With the onset of the COVID-19 pandemic, it was estimated that there was a 25% increase in the number of cases of anxiety disorder and depression globally, with more than one billion people suffering from a mental disorder prior to the pandemic (4). This situation was expected because the prevalence of mental health problems tends to increase during disease outbreaks (5).

In Peru, during the pandemic, 19.7% and 14.5% of individuals were reported to have clinically relevant symptoms of depression and anxiety, respectively (6). For example, an average increase of 0.17% per quarter in the prevalence of moderate depressive symptoms was observed after the onset of the pandemic, that is, an increase of 1583 new cases of moderate depressive symptoms approximately every three months (7). However, previous studies or others that evaluated different mental health symptoms in different population groups in Peru (8, 9) did not consider the participation of Quechua-speaking groups.

Peru is a multicultural and multilingual country where 3 million 375 thousand 682 people (13.9%) have Quechua as their mother tongue, extending to almost all departments of the country (10). Little information is available on the prevalence of mental health symptoms in the Quechua-speaking population (11). Pre-pandemic studies have reported that 38.9% of Quechua speakers in Ayacucho presented symptoms of depression (12). Quechua people tend to perceive and express mental health symptoms differently than Spanish speakers do. It has been reported that the high Andean Quechua people construct and experience expressions of distress and suffering, such as *pinsamientuwan* (troubling thoughts), *ñakary* (suffering), and *llaki* (sadness), in a context characterized by the presence of social inequalities, social exclusion, and political violence (13). Many of these mental health problems can lead to suicide in quechua-speaking populations (14). Despite the above, no studies have reported the prevalence of mental health symptoms in the Peruvian Quechua-speaking population during the pandemic. Although we are in the post-pandemic period, it is important to monitor symptoms related to anxiety and depression that may be frequently reported after the pandemic (15). The

Quechua-speaking population has distinct cultural and linguistic characteristics that hinder access to quality mental health care because of the absence of trained personnel with adequate language proficiency (11, 16). This gap in mental health care means that when a Quechua-speaking population has an episode of sadness or *llaqui*, they go to *Yachactaita*, who is the person who heals using plants and rituals, because of the trust they have in him and the proximity to their communities, and do not go to a mental health professional (17).

The presence of mental health problems in the Quechua-speaking population has generated a need for instruments with adequate psychometric evidence to measure depression and anxiety. However, to the best of our knowledge, no instrument has been adapted to the Quechua language that would contribute to the objective diagnosis of depression and anxiety at the individual or group level. In Peru, most health personnel are not fluent in Quechua, either as a first or a second language. Therefore, to assess Quechua-speaking populations, bilingual people, usually younger family members who have a different worldview than the person being assessed, were used to translate the questions and answers of the questionnaire from Spanish to Quechua. However, this is a poorly recommended practice as it introduces bias in the assessment.

There are different instruments that assess anxiety and depression adapted and validated in Peru, such as the Reynolds Adolescent Depression Scale (RADS-2) (18), the Patient Health Questionnaire-9 (PHQ-9) (19), the two-item version of the PHQ (PHQ-2) (20), Coronavirus Anxiety Scale (21,22), Generalized Anxiety Disorder - 7 (GAD-7) (23), and the Depression, Anxiety and Stress Scales (DASS-21) (24). However, these instruments have been adapted in urban areas and in the Spanish language, leaving aside the population groups of the high Andean areas. This shows the lack of diagnostic tools in the Quechua language that have an adequate cultural adaptation process. It is important to have screening instruments that allow the early detection of depressive and anxiety symptoms in individuals or populations at risk (20).

The Patient Health Questionnaire for Depression and Anxiety (PHQ-4) (25) has proven useful in research and clinical practice (26). The PHQ-4 comprises the first two items of the PHQ-9 and the first two items of the GAD-7. These items are based on the DSM-IV criteria for the diagnosis of depression and generalized anxiety disorder (27). The PHQ-4 was developed based on the assumption that symptoms of depression and anxiety frequently coexist and the intent to be able to identify

individuals who may be suffering from one or both sets of symptoms (28). A composite measure of anxiety and depression based on a single construct is not in accordance with the current nosological science, as suggested in the Diagnostic and Statistical Manual of Mental Disorders (29). However, there is evidence of comorbidity between anxiety and depression across the lifespan (30, 31).

Previous studies have shown adequate evidence of the construct validity of the PHQ-4, significant relationships with the degree of functional disability, and the ability to discriminate between groups with worse anxiety and depressive symptoms (25,26,32). For example, in the original study conducted on a sample of 2149 patients from 15 primary care clinics in 12 U.S. states, Kroenke et al. (25) used the principal component analysis method to suggest the presence of two subscales, one for depression and one for anxiety, consisting of two items each. The same study indicated that the reliability estimates for anxiety, depression, and total scale subscales were good (0.85, 0.82, and 0.81, respectively). Another more recent study, involving 5030 participants from the general population of Germany, showed evidence of the presence of a two-factor model of the PHQ-4, with good reliability and invariance across different age and gender groups, and showed significant relationships with demographic risk factors for depression and anxiety (32). A study conducted in South Korea with outpatients at the Department of Psychiatry, Ansan Hospital, Korea University (33) indicated that the Korean version of the PHQ-4 has good internal consistency reliability, test-retest reliability, and significant relationships with other depression/anxiety scales. Regarding its internal structure, a Korean study indicated that although the two-factor model had better fit indices than the one-factor model, the two-factor structure may not be completely adequate. Another study with Midwestern college students (34) indicated the presence of two factors (anxiety and depression) using component analysis and varimax rotation, with good reliability. In addition, it was reported that people diagnosed with depression or anxiety by a mental health professional had significantly higher scores on the PHQ-4, providing evidence of criterion validity. One study indicated that the PHQ-4 has a two-dimensional structure, adequate reliability, and can adequately assess the dimensions of anxiety and depression during pregnancy (35). The PHQ-4 also presented a good fit for a two-factor model, good reliability, and significant relationships with perceived stress in a sample of English- and Spanish-speaking Spanish Americans (36). In South America, the psychometric properties of the PHQ-4 have been

evaluated in the general population of Colombia (37), with a Cronbach's alpha coefficient value of 0.84. The presence of a two-factor model, which was invariant across different age and sex groups, evidence of convergent validity with another measure of hospital anxiety, depression, and general health, and evidence of divergent validity with measures of self-efficacy and life satisfaction were reported.

Compared to the other instruments mentioned above that measure symptoms of depression and anxiety, the PHQ-4 has several advantages. First, long instruments pose challenges for different groups of people, such as those with deficits in cognitive functioning (38). In this regard, ultra-short measures, such as the PHQ-4, comprising simply worded items, can be completed within a few minutes (39). Ultra-brief screening instruments are defined as measures that have one to four items and take less than four minutes to complete (39). Ultrashort measures are useful in primary care settings with limited time and resources (40). Second, two- or three-item measures have been shown to perform better than single-item measures (39). The PHQ-4 comprises items that measure the core characteristics of depression and anxiety, which are common psychiatric syndromes in many Latin American countries. Third, rather than being a definitive diagnostic measure, the PHQ-4 is a preliminary measure that suggests further research on anxiety and depression, which promotes discussion and further treatment planning and evaluation (25).

The PHQ-4 has been used to detect symptoms of depression and anxiety in different groups, such as oncology patients (41), heart disease patients (42), and recently in healthcare personnel during the COVID-19 pandemic (43). However, unlike other similar measures such as the PHQ-9 and GAD-7, psychometric evidence for the PHQ-4 has not been widely examined in Latin American cultural contexts, and less so in Quechua speakers. This is despite the fact that, as reported, different studies have provided evidence for the widespread use of the PHQ-4 in medical and psychological research (28).

Therefore, the aim of this study was to adapt and validate the PHQ-4 (25, 32) in the Quechua-speaking population of the Collao variant of Puno-Peru. Specifically, evidence of validity based on internal structure was evaluated by confirmatory factor analysis (CFA), reliability with the internal consistency method, and the difficulty and discrimination characteristics of the items based on Item Response Theory (IRT).

Generally, CFA and internal consistency reliability tests

have been used to validate mental health measurement instruments for use in research and clinical practice (44), whereas IRT-based procedures have rarely been used (45). However, the development and validation of measurement instruments can benefit from the complementary use of the two methods. CFA focuses on the study of the structure of an instrument, such as the number of factors, factor loadings of items, presence of correlated errors, and presence of higher-order factors, among others. This provides evidence to indicate whether the set of items can be used to measure a construct for research or practice (46). On the other hand, IRT brings together a set of model-based techniques that examine the process by which people respond to the items of a measurement instrument (47). IRT methods focus on the items and, therefore, assess their individual characteristics and the specific relationships of the item with the latent trait. In general, CFA allows for a summative view of the level of anxiety and depression in individuals, allowing the calculation of a total score, whereas IRT analyses identify items that are more likely to be selected by people with different levels of anxiety and depression (48). The combination of SEM and IRT allows a more complete overview of an instrument (48). This would help develop more accurate profiles of individuals, allow the identification of anxiety and depression levels in different population groups, and help to design more effective interventions (49). Having an instrument with adequate psychometric evidence is important because the detection of anxiety and depression symptoms in the Quechua-speaking population is a necessary first step in improving the outcomes of patients diagnosed with anxiety and depression disorders (32). This is even more important if one considers that indigenous peoples are vulnerable groups because of the difficulties in developing prevention and mental health promotion strategies if their idiosyncratic characteristics and language are not taken into account (50,51). This will have greater adverse impacts on mental health problems in this population (52).

## METHODS

### Participants

Participants were speakers of the Collao variant Quechua living in the northeast of the high Andean city of Puno, mainly in the provinces of San Román, Azángaro, and Melgar. Participants were selected by non-probabilistic convenience sampling according to the following inclusion criteria: 1) being of legal age; 2) being able to communicate in Quechua (read, write,

and understand); and 3) providing informed consent. Exclusion criteria were those who understood only Quechua, but could not speak or write. The number of participants was calculated using Soper software (2023) based on the number of observed ( $n = 4$ ) and latent ( $n = 1$ ) variables in the model, the anticipated effect size ( $\lambda = 0.3$ ), the desired probability ( $\alpha = 0.05$ ), and the statistical power level ( $1 - \beta = 0.95$ ). The minimum recommended sample size was 100 individuals. The final number of participants in this study was higher than the recommended minimum.

A total of 221 speakers of the Collao variant of Quechua language with ages ranging from 18 to 66 years, with an average age of 31.2 years ( $SD = 11.7$ ) participated. Regarding sex, 104 were men (47.1%) and 117 were women (52.9%).

### Measures

Sociodemographic survey. This survey was constructed specifically for the present study with the objective of obtaining information on sex and age.

The Patient Health Questionnaire-4 (PHQ-4) (25). The PHQ-4 detects the frequency of the main symptoms of a depression and anxiety disorder during the last 2 weeks. It consists of four items, combining the two items of the PHQ-2 (53) and two items of Generalized Anxiety Disorder Scale- 2 ítems (GAD-2) (54). The first two items measure the frequency of depressive symptoms ("Little interest or pleasure in doing things" and "Feeling depressed or hopeless") and the last two items measure the frequency of anxiety symptoms ("Feeling nervous, anxious or on edge" and "Not being able to stop or control worry"). Each of the PHQ-4 items has four Likert-type response options, ranging from 0 = not at all to 3 = almost every day. The total score ranges from 0 to 12, where higher scores indicate a higher frequency of depressive and anxiety symptoms. For this study, the Spanish version was used, validated in Ecuador by López Guerra et al. (55). The purpose of the PHQ-4 is not to make a final diagnosis or to monitor the severity of depression and anxiety but to perform a preliminary screening for depressive and anxiety symptoms.

The translation of the items of the PHQ-4 from Spanish to Quechua in its Collao variant from Puno (Peru) was conducted following the indications of Hernández et al. (56). This included an initial translation performed by a group of translators, followed by back-translation performed by other translators without knowledge of the original questionnaire. In the initial translation, professional support from two translators was requested to perform the translation from Peruvian Spanish to Quechua in its Collao variant from Puno

(Peru). The translators were Peruvian and had a command of Peruvian Spanish and an expert level in the Collao de Puno variant of Quechua. Independently, each translator presented his or her report together with a Peruvian psychologist whose native language was Quechua. They met to record approval from the review committee and to leave a written version of the items of the PHQ-4 in Collao de Puno Quechua. In the back-translation, two translators, different from those assigned in the first stage, were in charge of the back-translation (from Quechua into Peruvian Spanish). Each one presented his or her translation report, and together with the review committee led by a psychologist who spoke Spanish and Quechua in the Collao variant, they reviewed and compared the coincidences between the translated versions. It was verified that the items complied with the measurement of the constructs in the back-translated version; therefore, the final version of the items of the PHQ-4 in

Quechua Collao de Puno was approved. Subsequently, a Quechua-speaking psychologist proficient in qualitative focus group data collection organized a face-to-face meeting with seven Quechua speakers of the Collao variant of Puno. These people were over 18 years of age, with five males and three females, all of whom had completed secondary school and could read Quechua and Spanish (bilingual). Initially, the participants were asked to read the written version of the survey and answer the two items as appropriate. Following the application of the items of the Quechua PHQ-4, the psychologist moderator invited participants to comment on the clarity of the items and their comprehension of everyday Quechua. There were no suggestions for changes or modifications of the Quechua version of the items. Table 1 shows the original English, Spanish, and Quechua versions of the Collao de Puno variant of the PHQ-4.

**TABLE 1. FINAL VERSION OF THE PHQ-4 IN QUECHUA COLLAO VARIANT OF PUNO-PERU**

N°	Spanish version of PHQ-4 in Spanish ( <i>English version in italics</i> )	Quechua version of PHQ-4
1	Sentirse nervioso/a, ansioso/a, o con los nervios de punta ( <i>Feeling nervous, anxious or on edge</i> ).	Yanqa nishu manchariska kasharani
2	No poder dejar de preocuparse o no poder controlar la preocupación ( <i>Not being able to stop or control worrying</i> ).	Manan llakiyta atipayta atinichu
3	Sentirse desanimado/a, deprimido/a o sin esperanzas ( <i>Feeling down, depressed or hopeless</i> ).	Manan imatapas ruwayta munanichu. Sinchi llakisqan kashani. Manan imatapas suyanichu kay kausaypi.
4	Poco interés o placer en hacer las cosas ( <i>Little interest or pleasure in doing things</i> ).	Manan imatapas ruwanayawanchu

## Procedure

An online form was prepared using the Google Forms platform, which included an informed consent form, a sociodemographic form, and items of the scales. The objective of the study and informed consent were obtained in the initial part of the form. In addition, it was clearly indicated that only bilingual people with a minimum reading level in Quechua were allowed to complete the survey. The online form was administered between February and April, 2022. Two members of the research team monitored the application of the form for data collection. We collaborated with university students, fluent in Collao of the Puno Quechua variant, who were trained in the use and application of the instrument. The surveyors identified WhatsApp groups of parents of educational institutions, Christian churches, and groups of

associations in Quechua-speaking rural communities to whom they presented the survey.

## Ethical considerations

The Research Ethics Committee of the Unión Universidad Peruana approved this study (approval number: 2022-CEEPG-0000190). The study followed all the ethical principles of the Declaration of Helsinki for research in humans. All participants provided informed consent before starting the study. In addition, they were informed that participation was anonymous and voluntary, and that the information collected would be used only for research purposes. Participants could withdraw from the study at any time they saw a fit.

## Data analysis

First, expert judges evaluated the evidence of content

validity. Three psychologists (with a master's degree) participated, two of whom had at least two years of work experience in the clinical area and one in university teaching. All had experience in the treatment of depression in bilingual Quechua-speaking adults (Quechua and Spanish). The review of these experts had a qualitative component for each item, where they could provide suggestions on the cultural equivalence, contextualization, and relevance of the words in Quechua. Additionally, it has a quantitative component that consists of assigning a certain score for each item, based on which content validity was determined through the relevance, representativeness, and clarity of the items, and the Aiken V value was calculated for each item (57).

In the Confirmatory Factor Analysis (CFA), because the items presented four response categories, Diagonally Weighted Least Squares with Mean and Variance corrected (WLSMV) estimator was used (58). The criteria used to assess model fit were RMSEA ( $< .08$ ), SRMR ( $< .08$ ), CFI ( $> .95$ ), and TLI ( $> .95$ ) (59,60). The internal consistency of the scale was assessed using Cronbach's alpha coefficient (61) and the omega coefficient (62). Values greater than .70 were considered adequate (63).

Item Response Theory (IRT) is an extension of the 2-parameter logistic model (2-PLM) for ordered polytomous items (64), which is based on the Graded Response Model (GRM) (65). Before assessing the item

parameters, the model fit was estimated using the C2 test developed for ordinal items (66). The fit criteria for assessing the GRM model were as follows: RMSEA  $\leq .05$  (67) and SRMSR  $\leq .05$  (68). For CFI and TLI values, criteria similar to those used in the SEM models ( $\geq .95$ ) were used (69). Regarding item parameters, discrimination (a) and difficulty (b) parameters were used. Information Curves for the Items and scale (IIC and TIC, respectively) were also calculated.

All statistical analyses were performed using the "lavaan" package (70) for the CFA and the "mirt" package for GRM (47). The RStudio environment for R was used in all cases.

## RESULTS

### Content validity

Table 2 shows that the four items of the PHQ-4 are clear, relevant, and consistent, both at the sample ( $V > 0.70$ ) and population levels ( $Li > 0.59$ ).

### Descriptive analysis

Table 3 shows that most of the participants had a low average score ( $< 1$ ); that is, most of them chose the lower response categories of the items. The items also presented adequate Asymmetry and Kurtosis indices ( $As < \pm 2$ ;  $Ku < \pm 7$ ) (71). In the polychoric correlation matrix, all items presented moderate relationships.

TABLE 2. AIKEN'S V FOR ASSESSING THE RELEVANCE, REPRESENTATIVENESS AND CLARITY OF PHQ-4 ITEMS

Items	Relevance (n = 3)				Representativeness (n = 3)				Clarity (n = 3)			
	M	DE	V	IC 95%	M	DE	V	IC 95%	M	DE	V	IC 95%
Item 1	3.00	0.00	1.00	0.70 – 1.00	3.00	0.00	1.00	0.70 – 1.00	2.67	0.58	0.89	0.56 – 0.98
Item 2	3.00	0.00	1.00	0.70 – 1.00	3.00	0.00	1.00	0.70 – 1.00	2.67	0.58	0.89	0.56 – 0.98
Item 3	3.00	0.00	1.00	0.70 – 1.00	2.67	0.58	0.89	0.56 – 0.98	2.67	0.58	0.89	0.56 – 0.98
Item 4	3.00	0.00	1.00	0.70 – 1.00	3.00	0.00	1.00	0.70 – 1.00	3.00	0.00	1.00	0.70 – 1.00

TABLE 3. DESCRIPTIVE ANALYSIS OF THE SCALE ITEMS

Anxiety and depression items	Descriptive analysis				Correlation matrix			
	M	SD	g1	g2	1	2	3	4
A1. Sentirse nervioso/a, ansioso/a, o con los nervios de punta ( <i>Feeling nervous, anxious or on edge</i> )	.64	.78	1.13	.81	1			
A.2. No poder dejar de preocuparse o no poder controlar la preocupación ( <i>Not being able to stop or control worrying</i> )	.87	.86	.72	-	.56	1		
D3. Sentirse desanimado/a, deprimido/a o sin esperanzas ( <i>Feeling down, depressed or hopeless</i> )	.63	.79	1.13	.68	.55	.65	1	
D4. Poco interés o placer en hacer las cosas ( <i>Little interest or pleasure in doing things</i> ).	.76	.85	.98	.32	.62	.55	.68	1

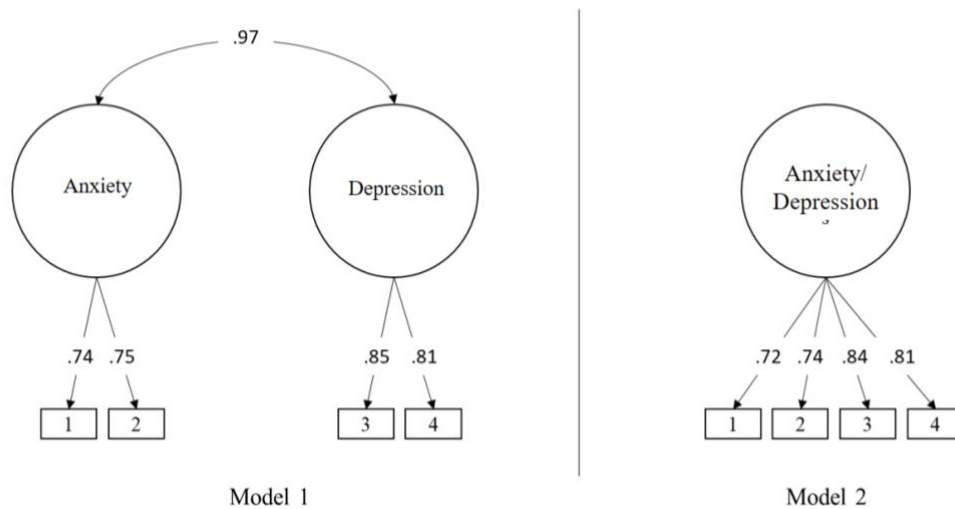
Note: M=Mean; SD=Standard Deviation; g1= Skewness; g2= Kurtosis

Validity based on internal structure

In the present study, the model of two related dimensions was found to present acceptable fit indices for the data ( $\chi^2 = 8.72$ ;  $df = 1$ ;  $p = .003$ ;  $RMSEA = .187$  [CI90% .088 - .310];  $SRMR = .028$ ;  $CFI=.99$ ;  $TLI=.93$ ). Nevertheless, the relationship between the dimensions was high (.97), which suggests that the items in the sample studied do not differentiate between the two separate constructs of anxiety and

depression, but would measure a single construct. Following this line of thought, the existence of a one-dimensional model was evaluated, which presented slightly higher fit indices than the previous model ( $\chi^2 = 8.41$ ;  $df = 2$ ;  $p = .015$ ;  $RMSEA = .121$  [CI90% .045 - .210];  $SRMR = .029$ ;  $CFI=.99$ ;  $TLI=.97$ ). In addition, those of the one-dimensional model had high factor weights (Figure 1). Considering this evidence, a one-dimensional model was used for subsequent analyses.

FIGURE 1. CONFIRMATORY FACTOR ANALYSIS OF THE SCALES



Reliability of the scale

The study found that the brief scale presented adequate reliability indices ( $\alpha = .86$ ;  $\omega = .81$ ), evidencing that the items presented adequate precision in measuring the construct.

Item Response Theory Model: Graded Response Model (GRM)

The results of the Confirmatory Factor Analysis (CFA) allow two main assumptions to be met: the existence of one-dimensionality and consequently local independence. Table 4 shows that the GRM model presents acceptable fit indices ( $C2[df] = 5.41[2]$ ;  $p =$

.066;  $RMSEA = .088$ ;  $SRMSR = .036$ ;  $TLI = .97$ ;  $CFI = .99$ ). It is also appreciated that all item discrimination parameters are above the value of 1, generally considered good discrimination (Zickar, et al., 2002). Regarding the difficulty parameters, all the threshold estimators increased monotonically.

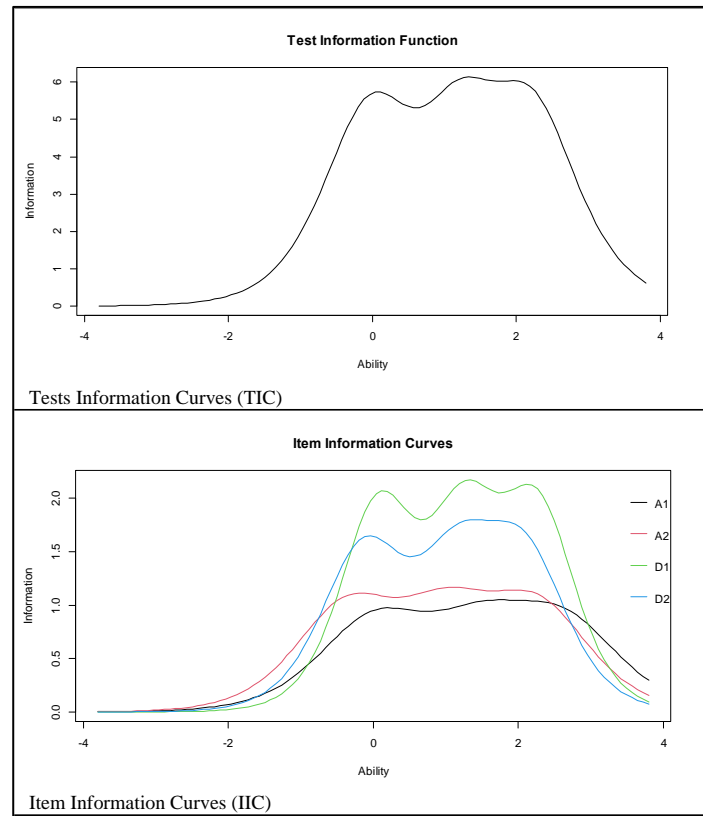
Figure 2 shows the Information Curves for the four items and the scale (IIC and TIC). The IIC shows that items D1 and D2 are the most accurate items of the scale for assessing latent traits. In addition, the TIC shows that the test is the most reliable (accurate) in the scale between -.5 and 2.5.

TABLE 4. GRM MODEL FIT INDICES, DISCRIMINATION AND DIFFICULTY PARAMETERS

Item	Item parameters			
	a	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>
Item 1	1.91	.04	1.49	2.58
Item 2	2.04	-.33	1.03	2.23
Item 3	2.83	.08	1.26	2.24
Item 4	2.51	-.11	1.18	2.05

Note: a= discrimination parameters; b= difficulty parameters

FIGURE 2. ITEM AND TEST INFORMATION CURVES FOR THE SCALE



## DISCUSSION

Developing instruments with adequate characteristics to measure constructs involves a process that involves adapting them to different social and cultural conditions. Therefore, the present study aimed to adapt and validate the PHQ-4 in the Quechua-speaking population in its Collao variant in Puno-Peru. To the best of our knowledge, this is the first study to evaluate psychometric evidence of the PHQ-4 in a Quechua-speaking population.

The results of the content validity analysis of the PHQ-4 items reported that all of them were relevant, coherent, and clear to adequately represent the symptoms of anxiety and depression. Thus, the contents of the four items are adequate for their application to a sample of Quechua speakers. Previous analyses of the internal structure of the PHQ-4 have reported the presence of two factors: anxiety and depression (25, 32-37). However, this factor structure was not confirmed in the present study, as the CFA result supported a one-dimensional model, where anxiety and depression items formed a single latent variable of emotional problems. A previous study indicated that the two-factor model may not be completely adequate, as the covariance between the depression and anxiety factors was also high, and the discriminant validity was insufficient, suggesting that

the factors are not sufficiently independent (33). The lack of support for the two-factor structure may be explained by the characteristics of the sample. In this study, the participants were Quechua-speaking individuals, whereas in other studies, the sample consisted of members of the general population from European, Asian, and Latin American countries, or patients seen in primary care centers or psychiatric departments. In the present study, the mean PHQ-4 score was 2.89 (SD = 2.62), which is higher than that reported in German (1.76, SD = 2.06) (32) and Colombian (1.27, SD = 2.01) general populations (37). However, it was similar to that reported in a study with a sample of patients from a primary care center in the United States (2.5, SD = 2.8) (25) and significantly lower than that reported in a study of psychiatric patients (6.52 SD = 3.45) (33). It appears that the symptom burden and comorbidity of depression and anxiety, similar to patients attending primary care centers, may make the distinction between the two factors unclear.

However, considering the brevity of the PHQ-4, the reliability was relatively high. The values of the reliability coefficients were similar to the values previously reported by studies conducted in different populations (25, 32-37). This would indicate that the PHQ-4 is as accurate for measuring symptoms of depression and anxiety in the Quechua-speaking population as in other population groups.



Having a one-dimensional measure, and consequently presenting local independence, are assumptions needed to perform analyses based on the IRT (64). One of the main results indicated that all PHQ-4 items had good discrimination ability. This indicates that the PHQ-4 items are useful for discriminating between Quechua-speaking participants with high and low levels of anxiety and depression. The difficulty parameters indicate that the items are comparatively difficult and that a higher presence of the latent trait (anxiety and depression) is needed to answer higher response categories (almost every day). The information curves allow a better understanding of the efficiency of the items in measuring symptoms of anxiety and depression in Quechua-speaking participants. In this sense, the PHQ-4 is accurate in assessing anxiety and depressive symptoms between .5 and 2.5 of the latent variables, where the items measuring depression (D1 and D2) are the most accurate in measuring emotional problems among the four items.

Despite these important results, this study has limitations that can be addressed in future studies. First, participants were selected through non-probabilistic convenience sampling. This means that the sample is not representative and the findings cannot be generalized to the total population of Quechua speakers in its Collao variant in Puno-Peru. Second, as the main objective was to establish a preliminary and brief screening measure for anxiety and depression, the study included a sample of the general population, and there were no diagnostic measures to assess criterion validity. Therefore, these findings cannot be generalized to clinical populations. Previous studies have suggested that clinical versus general population-derived samples do not confound the importance of emotional problems (72). Nevertheless, it is suggested that future studies incorporate clinical diagnostic standards to verify or modify what was reported in this study. Third, the study only analyzed data from speakers of the Quechua variant of Collao living in the northeastern high Andean city of Puno. However, it is not clear whether these findings can be generalized to other samples of speakers of other Quechua variants from the Amazonian, northern, central, and southern regions. It should be noted that the same item may present different nuances depending on the translation. This may generate different interpretations of the contents of anxiety and depression symptoms in different cultures and contexts. Therefore, future studies should examine the psychometric properties of the PHQ-4 in a broader set of Quechua-speaking populations. Fourth, other measures of depression and anxiety were not

used in this study; therefore, there was no evidence of the relative performance of the PHQ-4 in comparison with other measures of anxiety and depression. Fifth, the test-retest reliability of the PHQ-4 was not assessed. This leads to the possibility that future studies may apply the DAS-4 at different time points to the same sample to assess the stability of the scores. Sixth, the PHQ-4 is a self-report measure that can assess the presence of social desirability. Therefore, it is advisable to consider, in the future, the use of structured diagnostic interviews to have complementary information on the presence of anxiety and depressive symptoms.

In conclusion, despite these limitations, the results indicated that the PHQ-4 had good psychometric properties in a sample of Quechua Collao speakers from Puno. Thus, it demonstrates the usefulness of using the PHQ-4 as a rapid, reliable, and valid primary screening measure for Quechua speakers in need of in-depth assessment and symptom monitoring for the diagnosis and treatment of anxiety and depression. In this sense, the study is unique in that it focuses only on a particular group of people within a particular context (i.e., Quechua speakers in Peru). However, much more information is needed on this population, as they are very vulnerable to mental health problems and lack access to health care. The scale makes it possible to report outcomes achieved in mental health settings and objectively evaluate the effectiveness of mental health care actions. However, it has been suggested that the application of short questionnaires, such as the PHQ-4, should be accompanied by assessments using longer instruments and interviews conducted by clinical psychologists (73). This would provide a more useful systematic assessment to identify and eventually intervene in cases that are more prone to anxiety and depression. Caution should be exercised in using only brief screening questionnaires, such as the PHQ-4, without having reasonable plans to identify truly depressed or anxious individuals through a diagnosis confirmed by a qualified mental health professional (34). Using instruments such as the PHQ-4 without the possibility of confirming the diagnosis may generate false positives, indicating, for example, the presence of increased levels of distress and dysphoria but not clinical depression (74). However, in settings such as primary care, where time is a priority, the use of the PHQ-4 may be appropriate. In this sense, the speed of administration and ease of interpretation of PHQ-4 results represent important features for the implementation and evaluation of new screening procedures for depression and anxiety symptoms. A systematic review indicated that brief measures, such as the PHQ-4, may be suitable for screening and

identifying mental health disorders that can support the development of early intervention (75).

## AUTHORS CONTRIBUTIONS

RFCE and TC-R: conceptualized, conducted the research, wrote the initial draft, and wrote and revised the final version. LWV: designed the methodology and analyzed the data, wrote and reviewed the final version; OM-B, JC, MT-B, JT and IB: data collection, critical review of the manuscript, managed the research activities, wrote and reviewed the final version. All authors assume responsibility for the article.

## REFERENCIAS

- Santomauro DF, Herrera AMM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet* 2021; 398(10312): 1700-12. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
- World Health Organization. Depression and other common mental disorders: global health estimates. Switzerland: World Health Organization: 2017.
- Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020; 396(10258): 1204-22. [https://doi.org/10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)
- World Health Organization. World mental health report: transforming mental health for all. Executive summary. Ginebra: World Health Organization, 2022.
- Tomfohr-Madsen LM, Racine N, Giesbrecht GF, Lebel C, Madigan S. Depression and anxiety in pregnancy during COVID-19: A rapid review and meta-analysis. *Psychiatry Res.* 2021; 300: 113912. <https://doi.org/10.1016/j.psychres.2021.113912>
- Villarreal-Zegarra D, Copez-Lonzoy A, Vilela-Estrada AL, Huarcaya-Victoria J. Depression, post-traumatic stress, anxiety, and fear of COVID-19 in the general population and health-care workers: prevalence, relationship, and explicative model in Peru. *BMC Psychiatry* 2021; 21: 1-14. <https://doi.org/10.1186/s12888-021-03456-z>
- Villarreal-Zegarra D, Reátegui-Rivera CM, Otazú-Alfaro S, Yantas-Alcantara G, Soto-Becerra P, Melendez-Torres GJ. Estimated impact of the COVID-19 pandemic on the prevalence and treatment of depressive symptoms in Peru: an interrupted time series analysis in 2014–2021. *Soc Psychiatry Psychiatr Epidemiol.* 2023; 1-11. <https://doi.org/10.1007/s00127-023-02446-8>
- Huarcaya-Victoria J, Elera-Fitzcarrald C, Crisol-Deza D, Villanueva-Zúñiga L, Pacherras A, Torres A, et al. Factors associated with mental health in Peruvian medical students during the COVID-19 pandemic: a multicentre quantitative study. *Rev Colomb Psiquiatr (Engl Ed).* 2021;50(34-7450(21)00108-6. <https://doi.org/10.1016/j.rcp.2021.06.002>
- Quispe-Sancho A, Chambi-Macedo KL, Laurel-Vargas V, Huamani-Merma E, Cuzcano-Gonzales KV, Huaita-Rocha MA, et al. Depression, Anxiety, and Stress in Health Professionals Working During the COVID-19 Pandemic in Peru: An Analytical Cross-Sectional Study. *Electron J Gen Med.* 2021;18(6):em319. <https://doi.org/10.29333/ejgm/11210>
- Instituto Nacional de Estadística e Informática. Perú. Estado de la población en el año del bicentenario, 2021. Lima: Instituto Nacional de Estadística e Informática; 2021.
- Cjuno J, Figueroa-Quiñones J, Marca-Dueñas GG, Carranza-Esteban RF. Producción científica sobre depresión en poblaciones Quechua hablantes: una revisión narrativa. *Ter Psicol.* 2021;39(2):163-74. <http://dx.doi.org/10.4067/s0718-48082021000200163>
- Tremblay J, Pedersen D, Errazuriz C. Assessing mental health outcomes of political violence and civil unrest in Peru. *Int J Soc Psychiatry.* 2009;55(5):449-63. <https://doi.org/10.1177/0020764009103214>
- Pedersen D, Kienzler H, Gamarra J. Llaki and nakay: idioms of distress and suffering among the highland Quechua in the Peruvian Andes. *Cult Med Psychiatry.* 2010;34:279-300. <https://doi.org/10.1007/s11013-010-9173-z>
- Espinosa Menéndez N, Galindo Rosero A, Bastidas Lopera W, Monsalve Rojas JE. Estado actual de la población víctima de desplazamiento forzado en cuanto al enfoque diferencial de su atención, el derecho a la integridad, la subsistencia mínima y al ingreso económico. *Ágora.* 2012;12(1):19-46. <https://doi.org/10.21500/16578031.220>
- Gallegos M, Martino P, Caycho-Rodríguez T, Calandra M, Razumovskiy A, Arias-Gallegos WL, et al. ¿Qué es el síndrome pos-COVID-19? Definición y actualización. *Gac Med Mex.* 2022; 158(6): 451-55. <https://doi.org/10.24875/gmm.22000144>
- Darghouth S, Pedersen D, Bibeau G, Rousseau C. Painful languages of the body: Experiences of headache among women in two Peruvian communities. *Cult Med Psychiatry.* 2006;30:271-97. <https://doi.org/10.1007/s11013-006-9021-3>
- Abe A-M. Psychiatrists and traditional healers: unwitting partners in global mental health. *Int J Cult Ment Health.* 2012;5(2):148–9. <https://doi.org/10.1080/17542863.2010.512732>
- Ramos-Vera C, Quispe Callo G, Basauri Delgado M, Vallejos Saldarriaga J, Saintila J. Factorial and network structure of the Reynolds Adolescent Depression Scale (RADS-2) in Peruvian adolescents. *Plos one* 2023;18(5):e0286081. <https://doi.org/10.1371/journal.pone.0286081>
- Villarreal-Zegarra D, Copez-Lonzoy A, Bernabé-Ortiz A, Melendez-Torres GJ, Bazo-Alvarez JC. Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics. *PloS one* 2019;14(9):e0221717. <https://doi.org/10.1371/journal.pone.0221717>
- Baños-Chaparro J, Ynquillay-Lima P, Delgado FL, Guillen FGF. Evidencias psicométricas del Patient Health Questionnaire-2 para la evaluación de síntomas depresivos en adultos peruanos. *Rev Ciencias Médicas.* 2021;25(5):1-9. [URL](https://doi.org/10.1007/s00127-023-02446-8)
- Caycho-Rodríguez T, Vilca LW, Carbajal-León C, White M, Vivanco-Vidal A, Saroli-Aranibar D, et al. Coronavirus Anxiety Scale: New psychometric evidence for the Spanish version based on CFA and IRT models in a Peruvian sample. *Death Stud.* 2022;46(5):1090-9. <https://doi.org/10.1080/07481187.2020.1865480>
- Caycho-Rodríguez T, Valencia PD, Vilca LW, Carbajal-León C, Vivanco-Vidal A, Saroli-Aranibar D, et al. Cross-cultural validation of the new version of the Coronavirus Anxiety

- Scale in twelve Latin American countries. *Curr Psychol*. 2022;1-18. <https://doi.org/10.1007/s12144-021-02563-0>
23. Zhong QY, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sánchez SE, Williams MA. Diagnostic validity of the generalized anxiety disorder-7 (GAD-7) among pregnant women. *PloS one*. 2015;10(4):e0125096. <https://doi.org/10.1371/journal.pone.0125096>
  24. Carlos-Colchado EM, Chavez-Aburto LA, Dextre-Paye KJ, Mallqui-García SA, De La Cruz-Valdiviano CB. Escalas de Depresión, Ansiedad y Estrés (DASS-21): evidencias psicométricas en universitarios. *Eureka*. 2023;20(1):26-42. [URL](https://doi.org/10.1016/S0033-3182(09)70864-3).
  25. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. 2009;50(6):613-21. [https://doi.org/10.1016/S0033-3182\(09\)70864-3](https://doi.org/10.1016/S0033-3182(09)70864-3)
  26. Kroenke K, Baye F, Lourens SG. Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *J Affect Disord*. 2019;246:437-43. <https://doi.org/10.1016/j.jad.2018.12.098>
  27. Staples LG, Dear BF, Gandy M, Fogliati V, Fogliati R, Karin E, ... Titov. Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: The PHQ-2, GAD-2, and K-6. *Gen Hosp Psychiatry*. 2019;56:13-8. <https://doi.org/10.1016/j.genhosppsy.2018.11.003>
  28. Lenz AS, Li C. Evidence for measurement invariance and psychometric reliability for scores on the PHQ-4 from a rural and predominately Hispanic community. *Meas Eval Couns Dev*. 2022;55(1):17-29. <https://doi.org/10.1080/07481756.2021.1906157>
  29. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: American psychiatric association; 2013
  30. Byers AL, Yaffe K, Covinsky KE, Friedman MB, Bruce ML. High occurrence of mood and anxiety disorders among older adults: The National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2010;67(5):489-96. <https://doi.org/10.1001/archgenpsychiatry.2010.35>
  31. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychol Bull*. 2014;140(3):816-45. <https://doi.org/10.1037/a0034733>
  32. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J Affect Disord*. 2010;122(1-2):86-95. <https://doi.org/10.1016/j.jad.2009.06.019>
  33. Kim HW, Shin C, Lee SH, Han C. Standardization of the Korean version of the Patient Health Questionnaire-4 (PHQ-4). *Clin Psychopharmacol Neurosci*. 2021;19(1):104-11. <https://doi.org/10.9758/cpn.2021.19.1.104>
  34. Khubchandani J, Brey R, Kotecki J, Kleinfelder J, Anderson J. The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch Psychiatr Nurs*. 2016;30(4):457-62. <https://doi.org/10.1016/j.apnu.2016.01.014>
  35. Rodríguez-Muñoz MDLF, Ruiz-Segovia N, Soto-Balbuena C, Le HN, Olivares-Crespo ME, Izquierdo-Méndez N. The psychometric properties of the patient health questionnaire-4 for pregnant women. *Int J Environ Res Public Health*. 2020;17(20):7583. <https://doi.org/10.3390/ijerph17207583>
  36. Mills SD, Fox RS, Pan TM, Malcarne VL, Roesch SC, Sadler GR. Psychometric evaluation of the patient health questionnaire-4 in Hispanic Americans. *Hisp J Behav Sci*. 2015;37(4):560-71. <https://doi.org/10.1177/0739986315608126>
  37. Kocalevent RD, Finck C, Jimenez-Leal W, Sautier L, Hinz A. Standardization of the Colombian version of the PHQ-4 in the general population. *BMC Psychiatry*. 2014;14(1):1-8. <https://doi.org/10.1186/1471-244X-14-205>
  38. Renovanz M, Soebianto S, Tsakmaklis H, Keric N, Nadji-Ohl M, Beutel M, et al. 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). *Support Care Cancer*. 2019;27:4469-77. <https://doi.org/10.1007/s00520-019-04718-z>
  39. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract*. 2007;57(535): 144-51. [URL](https://doi.org/10.7326/0003-4819-146-5-200703060-00004).
  40. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med Clin*. 2007;146(5):317-25. <https://doi.org/10.7326/0003-4819-146-5-200703060-00004>
  41. Nipp RD, El-Jawahri A, D'Arpino SM, Chan A, Fuh CX, Johnson PC, et al. Symptoms of posttraumatic stress disorder among hospitalized patients with cancer. *Cancer*. 2018;124(16):3445-53. <https://doi.org/10.1002/cncr.31576>
  42. Whooley MA, Caska CM, Hendrickson BE, Rourke MA, Ho J, Ali S. Depression and inflammation in patients with coronary heart disease: findings from the Heart and Soul Study. *Biological Psychiatry*. 2007;62(4):314-20. <https://doi.org/10.1016/j.biopsych.2006.10.016>
  43. Zhang WR, Wang K, Yin L, Zhao WF, Xue Q, Peng M, et al. Mental health and psychosocial problems of medical health workers during the COVID-19 epidemic in China. *Psychother Psychosom*. 2020;89(4):242-50. <https://doi.org/10.1159/000507639>
  44. Bowen N, Guo S. Structural equation modeling. Oxford: Oxford University Press, Inc; 2012.
  45. Bean GJ, Bowen NK. Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *J Evid Based Soc Work*. 2021;18(6):597-618. <https://doi.org/10.1080/26408066.2021.1906813>
  46. Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. *Psychol Methods*. 2007;12(1):58-79. <https://doi.org/10.1037/1082-989X.12.1.58>
  47. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. *J Stat Softw*. 2012;48:1-29. <https://doi.org/10.18637/jss.v048.i06>
  48. Meijer RR, Sijtsma K, Smid NG. Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Appl Psychol Meas*. 1990;14(3):283-98. <https://doi.org/10.1177/014662169001400306>
  49. Zamora-Araya JA, Smith-Castro V, Montero-Rojas E, Moreira-Mora TE. Advantages of the Rasch Model for Analysis and Interpretation of Attitudes: The Case of the Benevolent Sexism Subscale. *Revista Evaluar*. 2018; 18(3):1-13. <https://doi.org/10.35670/1667-4545.v18.n3.22201>
  50. Encina López F, Buongermini Méndez R, Michel Kolm K, Torales Benítez J. Determinación de salud auto-percibida en una población rural del Paraguay implementando el cuestionario de Goldberg -GHQ12- en guaraní. *An. Fac. Cienc. Méd*. 2013;46(2):15-24. [URL](https://doi.org/10.1016/j.jad.2009.06.019).

51. Vecchio EA, Dickson M, Zhang Y. Indigenous mental health and climate change: A systematic literature review. *J Clim Chang Health*. 2022;6:100121. <https://doi.org/10.1016/j.joclim.2022.100121>
52. Carranza Esteban RF, Mamani-Benito O, Cjuno J, Tito-Betancur M, Lingán-Huamán SK, Arias-Chávez D. Translation and validation of the WHO-5 General well-being index into native language Quechua of the Peruvian South. *Heliyon*. 2023;9(6):e16843. <https://doi.org/10.1016/j.heliyon.2023.e16843>
53. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res*. 2005;58(2):163-71. <https://doi.org/10.1016/j.jpsychores.2004.09.006>
54. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41(11):1284-92. <https://doi.org/10.1097/01.mlr.0000093487.78664.3c>
55. López Guerra V, Aguirre Mejía AJ, Guerrero Alcedo JM. Propiedades psicométricas y estructura factorial del cuestionario de salud del paciente PHQ-4 en estudiantes universitarios ecuatorianos. *Rev Cubana Enferm*. 2022;38(3):e4885. [URL](https://doi.org/10.1016/j.rce.2022.03.006).
56. Hernández A, Hidalgo MD, Hambleton RK, Gómez Benito J. International test commission guidelines for test adaptation: A criterion checklist. *Psicothema*. 2020;32(3):390-8. <https://doi.org/10.7334/psicothema2019.306>
57. Aiken LR. Three coefficients for analyzing the reliability and validity of ratings. *Educ Psychol Meas*. 1985;45(1):131-42. <https://doi.org/10.1177/0013164485451012>
58. Brown TA. *Confirmatory Factor Analysis for Applied Research*, Second Edition. Guilford Publications; 2015.
59. Kline RB. *Principles and Practice of Structural Equation Modeling* (4th ed.). The Guilford Press; 2015.
60. Schumacker RE, Lomax RG. *A Beginner's Guide to Structural Equation Modeling: Fourth Edition* (4th ed.). Routledge; 2015.
61. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 2015;16(3):297-334. <https://doi.org/10.1007/BF02310555>
62. McDonald RP. *Test Theory: A Unified Treatment*. Taylor & Francis; 1999.
63. Raykov T, Hancock GR. Examining change in maximal reliability for 63 multiple-component measuring instruments. *Br J Math Stat Psychol*. 2005;58(1):65-82. <https://doi.org/10.1348/000711005X38753>
64. Hambleton RK, Swaminathan H, Rogers J. *Fundamentals of item response theory*. Sage; 1991.
65. Samejima F. Graded response model. In Van der Linden WJ, Hambleton RK, eds. *Handbook of modern Item Response Theory*. New York, NY: Springer; 1997, pp. 85-100.
66. Cai L, Monroe S. *A New Statistic for Evaluating Item Response Theory Models for Ordinal Data*. California: University of California; 2014.
67. Maydeu-Olivares A, Joe H. Assessing Approximate Fit in Categorical Data Analysis. *Multivariate Behav Res*. 2014;49(4):305-28. <https://doi.org/10.1080/00273171.2014.911075>
68. Maydeu-Olivares A. Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement*. 2013;11(3):71-101. <https://doi.org/10.1080/15366367.2013.831680>
69. Lubbe D, Schuster C. A Graded Response Model Framework for Questionnaires with Uniform Response Formats. *Appl Psychol Meas*. 2019;43(4):290-302. <https://doi.org/10.1177/0146621618789394>
70. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw*. 2012;48(2):1-36. <https://doi.org/10.18637/jss.v048.i02>
71. Finney SJ, DiStefano C. Nonnormal and categorical data in structural equation modeling. In: *Structural equation modeling: A second course*, 2nd ed. Charlotte, NC, US: IAP Information Age Publishing; 2013. p. 439-92.
72. Malgaroli M, Calderon A, Bonanno GA. Networks of major depressive disorder: A systematic review. *Clin Psychol Rev*. 2021; 85:102000. <https://doi.org/10.1016/j.cpr.2021.102000>
73. Giuliani M, Gorini A, Barbieri S, Veglia F, Tremoli E. Examination of the best cut-off points of PHQ-2 and GAD-2 for detecting depression and anxiety in Italian cardiovascular inpatients. *Psychol Health*. 2021;36(9):1088-1101. <https://doi.org/10.1080/08870446.2020.1830093>
74. Klein MC, Ciotoli C, Chung H. Primary care screening of depression and treatment engagement in a university health center: a retrospective analysis. *J Am Coll Health*. 2011;59(4):289-95. <https://doi.org/10.1080/07448481.2010.503724>
75. Mulvaney-Day N, Marshall T, Downey Piscopo K, Korsen N, Lynch S, Karnell LH, et al. Screening for behavioral health conditions in primary care settings: a systematic review of the literature. *J Gen Intern Med*. 2018;33:335-46. <https://doi.org/10.1007/s11606-017-4181-0>