



Youth screening depression: Validation of the Patient Health Questionnaire-9 (PHQ-9) in a representative sample of adolescents

Eduardo Fonseca-Pedrero^{a,*}, Adriana Díez-Gómez^a, Alicia Pérez-Albéniz^a, Susana Al-Halabí^b, Beatriz Lucas-Molina^c, Martín Debbané^{d,e}

^a Universidad de La Rioja, Spain

^b Universidad de Oviedo, Spain

^c Universidad de Valencia, Spain

^d Developmental Clinical Psychology Research Unit, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland

^e Research Department of Clinical, Educational and Health Psychology, University College London, United Kingdom

ARTICLE INFO

Keywords:

Depression
Adolescents
Psychometric
Validation
PHQ-9
Mental health

ABSTRACT

Background: Depression symptoms and mood disorders constitute one of the major public health challenges among youths. Thus, early prevention and intervention for depression should be a priority. The main goal of the present study was to validate the Patient Health Questionnaire-9 (PHQ-9) scores in a school-based sample of non-clinical adolescents.

Method: Stratified random sampling was conducted. Participants were 2235 students ($M = 14.49$, $SD = 1.76$, range = 12–18 years), 52.9 % were female, from 34 secondary schools in Spain. Several previously validated self-reported questionnaires of mental health and psychopathology were administered.

Results: The unidimensional factorial model of the PHQ-9 items showed adequate goodness of fit indices. Strong measurement invariance across gender was found. Omega for the PHQ-9 total score was 0.87. The PHQ-9 total score was positively associated with anxiety symptoms and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life.

Conclusions: The PHQ-9 is a brief, easy, and reliable tool for assessing self-reported depressive symptoms in both clinical and school settings. PHQ-9 may be used as a screening tool for universal early detection and monitoring of depression symptoms during adolescence.

1. Introduction

Emotional problems (e.g., depression, anxiety) are among the leading causes of associated disability and global burden of disease in young people. The *Global Burden of Diseases, Injuries, and Risk Factors Study GBD (2019)* (GBD 2019 Mental Disorders Collaborators, 2022) revealed that mental disorders remained among the top ten causes of burden of disease worldwide, with no evidence of an overall reduction in burden since 1990. The World Health Organization (WHO, 2022) informs that one billion people worldwide have a diagnosis of a mental disorder (more than one in eight adults and adolescents). Depression in young people is a rising concern, so it is our duty as a society to promote, protect and care for the mental health of the entire population, in particular of one of the most vulnerable groups, children and adolescents. Thus, missed opportunities for depression identification and

treatment can be quite costly from both personal and public health perspectives (Davis et al., 2022).

Depression, as a continuum phenotype, encompasses a range of mood-related concepts and a spectrum of difficulties, placing at its most extreme end clinical disorders that could be expressed as mood syndromes; while at the other end, depression can refer to a mood state in the context of normative affective experience (Hankin, 2015; Thapar et al., 2022). Across this quantitative variation, intermediate (subclinical or subthreshold depression) expressions can also be identified. Subthreshold depression is associated with low mood and additional depressive symptoms such as loss of interest and enjoyment, but without reaching the diagnostic threshold (Thapar et al., 2022). Previous works have found that youth depression is continuously, not categorically, distributed. This viewpoint of the dimensional alternative to traditional Nosologies (e.g., Diagnostic and Statistical Manual of Mental Disorders)

* Corresponding author at: University of La Rioja, C/ Luis de Ulloa, 2, Edificio VIVES; C.P: 26004, Logroño, La Rioja, Spain.

E-mail address: eduardo.fonseca@unirioja.es (E. Fonseca-Pedrero).

are related with psychopathology models as, the Hierarchical Taxonomy of Psychopathology (HiTOP) (Kotov et al., 2021). Conceptualizing depression (internalizing spectra) as a graded dimension has implications for theory and practice, as well as for methods and measurement (Eaton et al., 2023; Hankin et al., 2005).

The prevalence put forward by developmental epidemiological research for major depressive disorder and dysthymia is situated around 8 % (95 % CI: 0.02–0.13) and 4 % (95 % CI: 0.01–0.07) among adolescents, respectively. In addition, the global point prevalence rate of elevated self-reported depressive symptoms from 2001 to 2020 was 34 % (95 % CI: 0.30–.38) (Shorey et al., 2022). That is, 34 % of adolescents worldwide, aged 10–19 years, are at risk of developing clinical depression, which exceeds the reported estimates of individuals aged 18 to 25 years. The age of onset, severity, persistence, and comorbidity are factors to consider in the study of depression symptoms during this developmental stage. Previous studies have found that the average age of development of any mental disorders is 14.5 years (Solmi et al., 2022). The proportion of individuals with onset of mood disorders before the age of 14, 18, 25 were, respectively, 2.5 %, 11.5 %, 34.5 %, and the peak age was 20.5 years ($k = 79$, median = 31, IQR = 21–46) (Solmi et al., 2022). Additionally, adverse outcomes associated with clinical and subclinical depression during adolescence include the onset of other mental health disorders (e.g., anxiety, substance abuse, and conduct disorders). Elevated depressive symptoms are associated with many outcomes as risk behaviours, health problems, and adverse psychosocial outcomes in interpersonal, social, educational, and occupational functioning as well as suicidal behaviours (Thapar et al., 2022). For instance, in the educational sphere, depression was associated with poorer school grades (Riglin et al., 2014) and school absenteeism (Finning et al., 2019).

Depression symptoms and disorders in adolescents are frequently misdiagnosed and undertreated. Routine screening, from a preventive approach, has the potential to improve the early and reliable identification of depressive symptoms. According to NICE Guidelines [NG134] for depression in children and young people, healthcare professionals in primary care, schools and other relevant community settings should be trained to detect symptoms of depression, and to assess children and young people who may be at risk of depression (NICE Guidance, 2019). The detection of these individuals with subclinical depression, whether in health, social or educational settings, requires the availability of adequate tools to make informed and data-driven decisions. The PHQ-9 (Kroenke et al., 2001; Spitzer et al., 1999) has become a standard measure of depression research and clinical practice. The PHQ-9 is a self-report developed to assess the severity of depression according to DSM criteria. Its psychometric properties have been adequately examined (El-Den et al., 2018; Kroenke, 2021). Previous studies showed strong evidence that the PHQ-9 can be used as a unidimensional measure of depressive symptoms (Bianchi et al., 2022). In primary care settings, the brief nature and ease of scoring of this instrument make it an excellent choice for providers and researchers looking to implement depression screening (Richardson et al., 2010). The evidence suggests that assessing the factors separately will not provide any useful information for most patients (Boothroyd et al., 2019). An adolescent version of the PHQ-9 was also designed (Johnson et al., 2002). In addition, the standard PHQ-9 has been validated in previous work with adolescents (Anum et al., 2019; Borghero et al., 2018; Burdzovic & Brunborg, 2017; Fatiregun and Kumapayi, 2014; Leung et al., 2020; Rhew et al., 2016; Richardson et al., 2010; Sinclair-McBride et al., 2018). For instance, Burdzovic and Brunborg (2017), in a sample of Norwegian adolescents, found a single-factor structure for the PHQ-9.

Many instruments can be used to assess depression symptoms, but further adaptation of tests is needed to identify these experiences at an early age (Bernaras et al., 2019). This developmental period involves physical, psychological, and social changes, which may increase an individual's sensitivity and reactivity to stress exposure (Beck et al., 2021). To date, although the standard PHQ-9 has been validated in

previous studies, yet there is little information on the psychometric properties of the PHQ-9 scores in large and representative samples of the general population. For instance, few studies have tested the measurement invariance by gender or gathered new validity evidence using modern psychometric methods as item response theory (IRT) for this measurement tool.

In this context, the main goal of the present study was to validate the PHQ-9 scores in a school-based sample of adolescents. This study aimed to: a) analyze the prevalence of depressive symptoms; b) examine the internal structure of the PHQ-9 scores; c) test the measurement invariance of the PHQ-9 by gender; d) estimate the reliability of the PHQ-9 scores; and e) analyze the association between PHQ-9 scores and psychometric indicators of mental health and quality of life. In line with previous literature, it was hypothesized that the one-factor model of the PHQ-9 would have adequate goodness-of-fit indices. In addition, we further hypothesized that this hypothesized dimensional model would be equal across gender. We also expected that the reliability estimation of the PHQ-9 scores would be adequate. Finally, we expected that depressive symptoms would be related with emotional and behavioural difficulties.

2. Method

2.1. Participants

Stratified random sampling was conducted at the class level in the total student population of La Rioja (region in Northern Spain). The students belonged to different public and charter educational centers, compulsory secondary education and vocational training. Strata were formed depending on the public and charter nature of the educational institutions, and the educational level. A total of 34 schools and 98 classrooms participated in the study.

The initial sample consisted of 2640 students. Those participants that: a) showed a high score on the Oviedo Infrequency Response Scale (more than 2 points) ($n = 175$) and were over 18 years of age ($n = 247$) were removed. Thus, a total of 2235 students, 1045 men (46.8 %), 1183 (52.9 %) women, and 7 (0.3 %) non-binary identity participated in the study. The mean age was 14.49 years ($SD=1.76$), age range between 12 and 18 years. The age distribution was as follows: 12 years, $n = 280$; 13 years, $n = 387$; 14 years, $n = 396$; 15 years, $n = 408$; 16 years, $n = 371$; 17 years, $n = 240$; and 18 years, $n = 153$. The 90.8 % of the sample was identified as Spanish.

2.2. Instruments

Socio-demographics, mental health problems, and lifestyle. An *ad hoc* instrument was developed to assess age, school grade, sex, gender, and nationality. In addition, family history of mental disorder was assessed. Information about lifestyle was also collected: general health, hours of sleep, time to fall asleep, and frequency of free-time activities.

Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001; Spitzer et al., 1999). The PHQ-9 is composed of nine questions designed to assess depressive symptomatology according to DSM criteria. The items are answered according to the frequency of symptoms (0 = *not at all*, 1 = *some days*, 2 = *more than half of the days*, 3 = *almost every day*). A higher score is indicative of greater depressive symptomatology. The PHQ-9 has been validated into Spanish (González-Blanch et al., 2018).

Generalized Anxiety Disorder Assessment (GAD-7) (Spitzer et al., 2006). The GAD-7 is a seven-item instrument used to measure or assess the severity of generalized anxiety disorder. Each item asks the individual to rate the severity of his or her symptoms over the past two weeks. Response options include 0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, 3 = *almost every day*. The GAD-7 has been validated into Spanish (Muñoz-Navarro et al., 2017). In this study, the reliability of the total score was adequate (McDonald's Omega = 0.90)

Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). The

SDQ is a self-report questionnaire that is widely used for the assessment of different emotional and behavioural problems related to mental health in adolescents. The SDQ is made up of a total of 25 statements distributed across five subscales: Emotional symptoms, Conduct problems, Hyperactivity, Peer problems, and Prosocial behavior. The first four subscales yield a Total difficulties score. In this study we used a Likert-type response format with three options (0 = *not true*, 1 = *some-what true*, 2 = *certainly true*). The validated Spanish version of the SDQ was used in the present study (Ortuño-Sierra et al., 2022). The SDQ total difficulties score showed adequate reliability in this sample (McDonald's Omega = 0.75).

Kidscreen-10 Index (Ravens-Sieberer et al., 2010). The Kidscreen-10 Index is a measurement instrument developed and validated to assess health-related quality of life in children and adolescents aged 8 to 18 years. It presents a total of 10 questions in a Likert 5-choice response format, where a higher score is indicative of higher quality of life. The Kidscreen-10 has been validated in Spain (Aymerich et al., 2005). The Kidscreen-10 scores showed good reliability in this sample (McDonald's Omega = 0.91).

The Oviedo Infrequency Scale-revisited (INF-OV-R) (Fonseca-Pedrero et al., 2009). The INF-OV-R was administered to the participants to detect those who responded in a random, pseudorandom or dishonest manner. The INF-OV-R instrument is a self-report composed of 10 items in a dichotomic scale format (Yes/No). Students with more than two incorrect responses on the INF-OV-R scale were eliminated from the sample.

2.3. Procedure

The research was approved by the Ethical Committee of Clinical Research of La Rioja (CEImLAR, PI 552). The psychometric measures were administered collectively, through personal computers, in groups of 10 to 30 students, during school hours and in a classroom specially prepared for this purpose. Administration took place under the supervision of the researchers trained in a standard protocol. No incentive was provided for their participation. Participants' parents were asked to sign an informed consent form so that their children could participate in the study. Participants were informed of the confidentiality of their responses and of the voluntary nature of the study. This work is part of a broader project called PSICE (Evidence-based Psychology in Educational Contexts) (Fonseca-Pedrero et al., 2023b) (ClinicalTrials.gov. Ref: NCT05322642).

2.4. Data analyses

First, we calculated the prevalence and descriptive statistics of the PHQ-9 items. The PHQ-9 total score was divided into the following categories of increasing severity: 0–4 (minimal), 5–9 (mild), 10–14 (moderate), 15–19 (moderately severe), and 20–27 points (severe).

Second, in order to analyze the internal structure of the PHQ-9, several confirmatory factor analyses (CFA) were performed. Attending to previous studies, a one-dimensional model was examined. Diagonally Weighted Least Squares estimator was used. The following goodness-of-fit indices were used: Chi-square (χ^2), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and 90 % Confidence interval, and Standardized Root Mean Square Residual (SRMR). Hu and Bentler (1999) suggested that RMSEA should be 0.06 or less for a good model fit and CFI and TLI should be 0.95 or more, though any value over 0.90 tends to be considered acceptable.

Third, in order to test measurement invariance across gender, successive multigroup CFAs were conducted. Basically, a hierarchical set of steps are followed when testing measurement invariance, typically starting with the determination of a well-fitting multigroup and baseline model and continuing with the establishment of successive equivalence constraints in the model parameters across groups. The baseline model is

called the configural model, which is the first and least restrictive model specified and is important because it represents the baseline model against which all subsequent specified invariance models are compared. The configural model is established by specifying and testing the CFA model for each group separately. Once the theoretical model has been validated in both groups, configural invariance is then examined, requiring that the same pattern of fixed and freely estimated model parameters is equivalent across groups; however, no equality constraints are imposed on the model parameters between groups. Configural invariance is tested by assessing the model fit. When configural invariance is met (i.e., the model fits the data), it suggests that at least the general factor structure is similar, but not necessarily equivalent, across groups. The next step is to impose equality constraints on the factor loadings across the groups to test metric or weak invariance. If the model fit with the constrained parameters is significantly and practically worse than the baseline or configural model, then weak invariance is not supported. When metric invariance is met, it suggests that the same unit of measurement is being used for the item across the groups and that the participants interpret and respond to the measure in a similar manner (Horn and McArdle, 1992). The final step is to impose constraints on the item intercepts and factor loadings to test strong or scalar invariance across groups. The confirmation of the invariance of the intercepts permits comparison of the latent means in both groups. The analyzed models are nested in that the imposed constraints are progressively added. Due to the limitations of the $\Delta \chi^2$ regarding its sensitivity to sample size, Cheung and Rensvold (2002) proposed a more practical criterion, the Δ CFI, to determine if nested models are practically equivalent. In this study, when Δ CFI is greater than 0.01 between two nested models, the more constrained model is rejected since the additional constraints have produced practically worse fit. However, if the change in CFI is less than or equal to 0.01, it is considered that all specified equal constraints are tenable; therefore, we can continue with the next step in the analysis of measurement invariance.

Fourth, reliability estimation of the PHQ-9 scores were estimated using McDonald's Omega. In addition, from the IRT framework with the 2-PL Model, the test information function was computed. Classical test theory methods cannot give us direct guidance on the latent trait of a measure to accurately assess depressive experiences at various points along the continuum (Hambleton et al., 1991). IRT methods provide estimates of the position on the latent trait (i.e., the theta level) where the tool provides the most information. Test information function graphically depict the regions of the latent trait continuum most precisely assessed. Greater information reflects greater measurement accuracy, or reliability. Test information function are estimated on the same latent trait scale (standardized $M = 0$; $SD = 1$), yielding information that is comparable across tests (Olino et al., 2012).

Fifth, the associations between PHQ-9 scores and other mental health indicators were calculated. SPSS 22.0, FACTOR 10.5.01, and JASP were used for data analyses.

3. Results

3.1. Descriptive statistics

Prevalence and descriptive statistics for the PHQ-9 items are shown in Table 1. The prevalence rates of depressive symptoms according to the recommended cut-off points were: 52.1 % (minimal), 20.2 % (mild), 15.2 % (moderate), 8.1 % (moderately severe), and 4.4 % (severe).

3.2. Confirmatory factor analysis of the PHQ-9 items

The standardized factor loadings for the total sample and by gender are shown in Table 2. Goodness-of-fit indices for the one-dimensional model were adequate (see Table 3).

Table 1

Prevalence (%) of response and descriptive statistics for the items of the Patient Health Questionnaire-9 (PHQ-9) in the whole sample.

Item	Not at all	Several days	More than half of the days	Almost every day	M	SD	Skewness	Kurtosis
1 Little interest or pleasure in doing things	37.3	44.4	10.5	7.8	0.89	0.88	0.90	0.21
2 Feeling down, depressed, or hopeless	36.2	42.9	11.9	9	0.94	0.91	0.83	-0.06
3 Trouble falling or staying asleep, or sleeping too much	53.1	25.4	9.4	12.1	0.80	1.03	1.05	-0.18
4 Feeling tired or having little energy	59.1	21.3	10.3	9.4	0.70	0.99	1.21	0.20
5 Poor appetite or overeating	27.7	44.7	14.4	13.2	1.13	0.97	0.62	-0.53
6 Feeling bad about yourself – or that you are a failure	56.5	26.4	8.5	8.5	0.69	0.95	1.26	0.51
7 Trouble concentrating on things	37.9	34.8	15.2	12	1.01	1.01	0.68	-0.64
8 Moving or speaking so slowly that other people could have noticed	63.8	22.4	9.3	4.6	0.55	0.84	1.47	1.26
9 Thoughts that you would be better off dead or of hurting yourself	82.5	11.2	3.5	2.7	0.26	0.65	2.75	7.26

Table 2

Standardized factor loadings of the Patient Health Questionnaire-9 (PHQ-9) for the total sample and by gender.

Items	Total sample	Male	Female
1	0.843	0.802	0.832
2	0.737	0.667	0.796
3	0.710	0.593	0.714
4	0.732	0.609	0.729
5	0.722	0.625	0.735
6	0.835	0.802	0.803
7	0.607	0.590	0.627
8	0.561	0.498	0.554
9	0.799	0.778	0.773

Note. All standardized factor loadings estimated were statistically significant ($p < .01$).

3.3. Measurement invariance of the PHQ-9 scores across gender

Since the one-factor model showed a good fit, the measurement invariance of the unidimensional model of the PHQ-9 was tested as a function of gender. Goodness-of-fit indices for males and females were adequate (see Table 3). The configural, metric invariance and scalar measurement invariance models showed an adequate fit to the data. The ΔCFI between the constrained and unconstrained models was under 0.01, thus measurement invariance across gender for this unidimensional model was supported.

3.4. Reliability estimation of the PHQ-9 scores

The internal consistency of the PHQ-9 total frequency score, estimated with McDonald’s Omega, was 0.87 (95 % CI: 0.86–0.88). Item discrimination indices were higher than 0.30. According to the IRT framework, the test information function provides an optimal estimation at the medium-high latent trait (values between 0 and 2) (see Fig. 1). The tool reduces its accuracy around the lowest level of the latent trait.

Table 3

Goodness-of-fit indices for the hypothetical models tested and measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) by gender.

Model	χ^2	df	CFI	TLI	RMSEA (90 % IC)	SRMR	ΔCFI*
One-factor	203.802	27	0.993	0.991	0.054 (0.054–0.061)	0.043	
Measurement invariance							
Male (n = 1045)	99.404	27	0.988	0.983	0.051 (0.040–0.062)	0.054	
Female (n = 1183)	131.934	27	0.993	0.991	0.057 (0.048–0.067)	0.046	
Configural invariance	230.832	54	0.992	0.989	0.054 (0.047–0.067)	0.050	
Metric Invariance	275.388	62	0.990	0.988	0.056 (0.049–0.062)	0.054	<0.01
Scalar Invariance	396.285	79	0.985	0.987	0.060 (0.054–0.066)	0.052	<0.01

Note. χ^2 = Chi square; df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; IC = Interval Confidence; SRMR = Standardized Root Mean Square Residual; ΔCFI = Change in Comparative Fit Index.

Good model fit is indicated by a RMSEA \leq 0.06, CFI and TLI \geq 0.90 of 0.95, SRMR \geq 0.08.

*CFI $\Delta <$ 0.01 indicates measurement invariance across gender.

3.5. Evidence based on the relations of the PHQ-9 scores to other variables

We also studied the correlation between the PHQ-9 total score and mental health psychometric indicators. As shown in Table 4, the PHQ-9 total score was positively and statistically significant correlated with anxiety symptoms, and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life.

4. Discussion

Depression symptoms and disorder is a common problem among adolescents (Shorey et al., 2022). Routine screening for depression is recommended, yet standardization in screening and management is lacking (Cucci et al., 2022). Improving early recognition and assessment procedures, using reliable tools, can help us to prevent, as well as to promote evidence-based psychological treatments for mild and

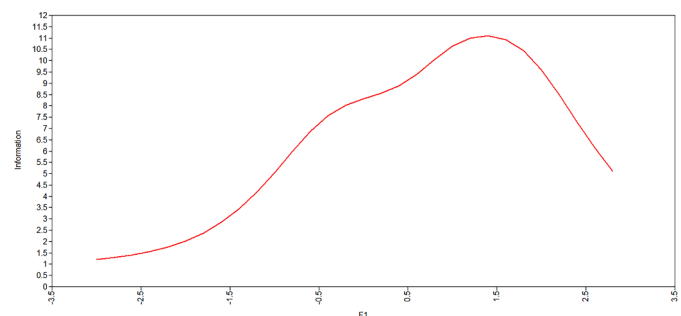


Fig. 1. Test information function of the Patient Health Questionnaire-9 (PHQ-9).

Note. F1=latent trait. Test information function graphically depict the regions of the latent trait continuum most precisely assessed. Greater information reflects greater measurement accuracy, or reliability.

Table 4

Pearson's correlations between the Patient Health Questionnaire-9 (PHQ-9) and the measures of mental health and quality of life.

	PHQ-9	GAD-7	SDQ PREM	SDQ PRCD	SDQ PRPP	SDQ HIP	SDQ PROS
GAD-7	0.766**						
SDQ PREM	0.709**	0.751**					
SDQ PRCD	0.397**	0.389**	0.291**				
SDQ PP	0.413**	0.383**	0.396**	0.257**			
SDQ HIP	0.474**	0.453**	0.362**	0.446**	0.184**		
SDQ PROS	-0.117**	-0.044*	0.034	-0.294**	-0.214**	-0.141**	
KS-10	-0.552**	-0.471**	-0.485**	-0.248**	-0.316**	-0.300**	0.177**

Note. ** $p < .01$;

PHQ-9= Patient Health Questionnaire-9; GAD-7 = Generalized Anxiety Disorder Assessment-7; SDQ = Strengths and Difficulties Questionnaire; SDQ PREM = SDQ emotional problems; SDQ PRCD = SDQ conduct problems; SDQ PRPP = SDQ peer problems; SDQ HIP = SDQ hyperactivity; SDQ PROS = SDQ prosocial behavior; KS-10= KIDSCREEN-10 Index.

moderate to severe depression. Thus, the main objective of the present study was to validate the PHQ-9 scores in a community-based sample of adolescents. The prevention, using reliable screening methods, are fundamental in the management of this phenomena (Patra and Kumar, 2023), considered as one of the major barriers for the family, educational, health and societal systems.

During the last two weeks, the 12.5 % of the adolescent sample reported, symptoms of depression of moderately severe (8.1 %) and severe (4.4 %). The results found in the present study seem consistent with previous international reports examining self-reported depression symptoms in adolescent school samples using the PHQ-9 (Burdzovic and Brunborg, 2017; Tsai et al., 2014). For instance, 5.8 % prevalence of clinically-elevated symptoms among Norwegian adolescents or 5.1 %, among Chinese high-school students (Tsai et al., 2014). Nonetheless, since we used internationally defined cut-off scores but not tested in Spain, these results should be considered as preliminary. Furthermore, the PHQ-9 has a frequency-based item response system, which does not necessarily imply depressive severity.

The unidimensional factorial model of the PHQ-9 showed adequate goodness of fit indices. Similar results have been found in previous research (Burdzovic and Brunborg, 2017; Anum et al., 2019; Leung et al., 2020). For instance, Leung et al. (2020) found, using CFA, that the one-factor model with three pairs of item correlations fitted the PHQ-9 data well, and that measurement invariances by age and gender were supported. In another work, Burdzovic and Brunborg (2017) found a single-factor structure for the PHQ-9 and adequate reliability estimation for both genders. Also, these results are in line with those found in the adult population (e.g., Bianchi et al., 2022). At this regard, previous studies supported robust evidence for the unidimensional structure of PHQ-9 to assess depressive symptoms (Bianchi et al., 2022), and add useful information for most patients (Boothroyd et al., 2019). From a conceptual point of view and given the overlap with other emotional problems and disorders (e.g., anxiety), we should consider, during adolescence, the PHQ-9 total score as a proxy for the general dimension of emotional dysregulation or internalizing factor (e.g., Piqueras et al., 2021). This view is fully congruent with current transdiagnostic and psychopathology models (e.g., HiTOP) consider that psychological phenomena as representing unbroken spectra ranging from very low to very high levels (Eaton et al., 2023). From this broader framework, subjective distress should be largely determined by the presence of emotional -or internalizing symptoms- such as anxiety and depression (Piqueras et al., 2021).

Multi-group CFAs showed that the one-factor model of the PHQ-9 had strong measurement invariance across gender. Previous studies have found mixed results. For instance, Burdzovic & Brundorg (2017) found no evidence of metric or scalar equality across genders. Anum et al. (2019) conducted a multi-group CFA and supported a one-factor structure of the PHQ-9 that was invariant across gender. These results showed that all PHQ-9 items were equivalent across gender (none showed differential items functioning). It should be stressed that if measurement invariance does not hold, the validity of such scores

should be questioned. Comparability between different groups only makes sense if it can be guaranteed that participants interpret and understand the latent construct in a similar manner (Horn and McArdle, 1992).

The reliability, estimated with McDonald's Omega, for the PHQ-9 total score was 0.87. This result is convergent with those found in previous studies conducted in adolescent populations (Burdzovic & Brunborg, 2017; Anum et al., 2019; Leung et al., 2020). In addition, we computed, from IRT, the test information function (TIF). The TIF provides an optimal estimation at the medium-high latent trait, that is, the PHQ-9 provided information at middle and higher levels of depression severity continuum. It is important in the context of ability or true score estimation because the TIF serves as an estimate of the latent trait accuracies of depression (i.e., the levels of the latent trait are measured with less standard error of measurement).

The PHQ-9 was positively associated with anxiety symptoms and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life. Similar results have found in prior research in both adolescent and adult samples. For example, Anum et al. (2019) found that the PHQ-9 correlated significantly with measures of anxiety, depression, mental wellbeing, and suicidal behavior. Furthermore, youth depression has been associated with a wide variety of risk and protective factors (e.g. Beck et al., 2021; Thapar et al., 2022). The risk factors identified include, among others, female sex, older age, poorer performance at school, lower interpersonal trust, social stress, atypicality, anxiety, feelings of incompetence, somatization, exposure to adverse events such as illness or death of a family member, physical or sexual abuse, bullying, poor academic achievement, poor sleep, more negative body image, more problematic use of social media or computer games, as well as poorer family functioning and inconsistent parental discipline (Beck et al., 2021).

One line of research in the field of depression is based on the idea of early detection, prevention and intervention in individuals who report subclinical depression with the aim of mitigating or reducing the impact that the disorder may have on the personal, family, academic, health and social spheres (Bernaras et al., 2019). Compared to other methods (e.g., clinical interviews), the use of these tools constitutes a rapid, efficient and non-invasive method of assessment. Moreover, the delay in diagnosis and treatment, the inadequate supply of mental health services, and the adverse consequences of depressive symptoms and disorders reinforce the importance of screening and treatment of this phenomena (as an internalizing disorder, depression is much less likely to be detected) at critical stages of human development such as adolescence. Due to the rising trend of depressive symptoms and the increasing number of at-risk teenagers, clinicians, researchers and practitioners should be more vigilant and proactive in their outreach activities to raise awareness and promote access and availability of services for this vulnerable group (Shorey et al., 2022). The United States Preventive Services Task Force recommends screening for depression in adolescents aged 12 to 18 years in the primary care setting (Grade B recommendation) (Forman-Hoffman et al., 2016).

Depression screening and prevention programs for young people need to go beyond the clinic walls (Guo and Jhe, 2021). Educational settings are at the forefront of mental health promotion and prevention during childhood and adolescence. Universal school-based screening would improve the detection of adolescent depression and reduce the limited capacity of the healthcare system to provide adequate mental health care to those who clearly need treatment. Schools are the “natural” place for actions to promote mental well-being and, specifically, for the prevention of mental health problems. Most adolescents spend long periods of time in classrooms, with schools being one of the main agents involved in socialization, as well as in formation and promotion of optimal development (Fonseca-Pedrero et al., 2023a). In this regard, the WHO Guidelines on School Health Services (WHO, 2021) highlight that schools are essential environments for the acquisition of socio-emotional skills (e.g., self-regulation and resilience). In the last decade, a range of psychological interventions to promote mental health and prevent mental health problems in schools have been tested with varying degrees of success (González-Roz et al., 2023).

Taken together, the results showed that the PHQ-9 is a brief, easy, and reliable tool for assessing self-reported depressive symptoms in school settings. This research provides further support for the validity of the PHQ-9 scores in a large and representative sample of non-clinical adolescents. In addition, the results indicated that the PHQ-9 could be a brief and useful tool to assess depression in general population samples. Future studies should continue to analyze the protective and risk factors for depression, add new psychometric procedures (e.g., network models) and digital based technologies (Elosua et al., 2023) to prevent mental health disorders in young people.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by a national project awarded by the Ministry of Science and Innovation of the Government of Spain and the Agency and the European Regional Development Fund (Project “PID2021-127301OB-I00” funded by MCIN / AEI /10.13039/501100011033 FEDER, UE).

References

- Anum, A., Adjorlolo, S., Kugbey, N., 2019. Depressive symptomatology in adolescents in Ghana: examination of psychometric properties of the Patient Health Questionnaire-9. *J. Affect. Disord.* 256, 213–218. <https://doi.org/10.1016/j.jad.2019.06.007>.
- Aymerich, M., Berra, S., Guillaumon, I., Herdman, M., Alonso, J., Ravens-Sieberer, U., Rajmil, L., 2005. Development of the Spanish version of the KIDSCREEN, a health-related quality of life instrument for children and adolescent. *Gas. Sanit.* 19, 93–102.
- Beck, A., LeBlanc, J.C., Morissette, K., Hamel, C., Skidmore, B., Colquhoun, H., Lang, E., Moore, A., Riva, J.J., Thombs, B.D., Patten, S., Bragg, H., Colman, I., Goldfield, G.S., Nicholls, S.G., Pajcar, K., Potter, B.K., Meeder, R., Vasa, P., Hutton, B., Shea, B.J., Graham, E., Little, J., Moher, D., Stevens, A., 2021. Screening for depression in children and adolescents: a protocol for a systematic review update. *Syst. Rev.* 10 <https://doi.org/10.1186/s13643-020-01568-3>.
- Bernaras, E., Jaureguizar, J., Garaigordobil, M., 2019. Child and adolescent depression: a review of theories, evaluation instruments, prevention programs, and treatments. *Front. Psychol.* 10, 543. <https://doi.org/10.3389/fpsyg.2019.00543>.
- Bianchi, R., Verkuilen, J., Toker, S., Schonfeld, I.S., Gerber, M., Brähler, E., Kroenke, K., 2022. Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. *Psychol. Assess.* 34, 595–603. <https://doi.org/10.1037/PAS0001124>.
- Boothroyd, L., Dagnan, D., Muncer, S., 2019. PHQ-9: one factor or two? *Psychiatry Res.* 271, 532–534. <https://doi.org/10.1016/j.psychres.2018.12.048>.
- Borghero, F., Martínez, V., Zitzko, P., Vöhringer, P.A., Cavada, G., Rojas, G., 2018. [Screening depressive episodes in adolescents. Validation of the Patient Health Questionnaire-9 (PHQ-9)]. *Rev. Med. Chil.* 146, 479–486. <https://doi.org/10.4067/S0034-98872018000400479>.
- Burdzovic, A.J., Brunborg, G.S., 2017. Depressive symptomatology among Norwegian adolescent boys and girls: the patient health Questionnaire-9 (PHQ-9) psychometric

- properties and correlates. *Front. Psychol.* 8, 887. <https://doi.org/10.3389/fpsyg.2017.00887>.
- Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Cucci, K.E., Thompson, J., Derouin, A., 2022. Bridging the gap in adolescent depression: an action plan intervention. *J. Pediatr. Health Care* 36, 264–269. <https://doi.org/10.1016/j.pedhc.2021.10.008>.
- Davis, M., Jones, J.D., So, A., Benton, T.D., Boyd, R.C., Melhem, N., Ryan, N.D., Brent, D. A., Young, J.F., 2022. Adolescent depression screening in primary care: who is screened and who is at risk? *J. Affect. Disord.* 299, 318–325. <https://doi.org/10.1016/j.jad.2021.12.022>.
- El-Den, S., Chen, T.F., Gan, Y.L., Wong, E., O'Reilly, C.L., 2018. The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. *J. Affect. Disord.* 225, 503–522. <https://doi.org/10.1016/j.jad.2017.08.060>.
- Eaton, N.R., Bringmann, L.F., Elmer, T., Fried, E.I., Forbes, M.K., Greene, A.L., Krueger, R.F., Kotov, R., McGorry, P.D., Mei, C., Waszczuk, M.A., 2023. A review of approaches and models in psychopathology conceptualization research. *Nat. Rev. Psychol.* 1–15. <https://doi.org/10.1038/s44159-023-00218-4>.
- Elosua, P., Aguado, D., Fonseca-Pedrero, E., Abad, F.J., Santamaría, P., 2023. New trends in digital technology-based psychological and educational assessment. *Psicothema* 35, 50–57. <https://doi.org/10.7334/PSICOTHEMA2022.241>.
- Fatiregun, A.A., Kumapayi, T.E., 2014. Prevalence and correlates of depressive symptoms among in-school adolescents in a rural district in southwest Nigeria. *J. Adolesc.* 37, 197–203. <https://doi.org/10.1016/j.jad.2013.12.003>.
- Finning, K., Ukoumunne, O.C., Ford, T., Danielsson-Waters, E., Shaw, L., Romero De Jager, I., Stentiford, L., Moore, D.A., 2019. The association between child and adolescent depression and poor attendance at school: a systematic review and meta-analysis. *J. Affect. Disord.* 245, 928–938. <https://doi.org/10.1016/j.jad.2018.11.055>.
- Fonseca-Pedrero, E., Lemos-Giráldez, S., Paino, M., Villazón-García, U., Muñiz, J., 2009. Validation of the schizotypal personality questionnaire brief form in adolescents. *Schizophr. Res.* 111, 53–60. <https://doi.org/10.1016/j.schres.2009.03.006>.
- Fonseca-Pedrero, E., Díez-Gómez, A., Pérez-Albéniz, A., Lucas-Molina, B., Al-Halabí, S., Calvo, P., 2023a. Psychology professionals in educational contexts: an unavoidable necessity. *Papeles del Psicólogo/ Psychologist Papers* 44, 112–124. <https://doi.org/10.23923/pap.psicol.3018>.
- Fonseca-Pedrero, E., Pérez-Albéniz, A., Al-Halabí, S., Lucas-Molina, B., Ortuño-Sierra, J., Díez-Gómez, A., Pérez-Sáenz, J., Inchausti, F., García, A.V.V., García, A.G., Solana, R.A., Ródenas-Perea, G., Clemente, M.P.D.V., López, A.C., Debbané, M., 2023b. PSICE Project Protocol: evaluation of the unified protocol for transdiagnostic treatment for adolescents with emotional symptoms in school settings. *Clínica y Salud* 34, 15–22. <https://doi.org/10.5093/CLYSA2023A3>.
- Forman-Hoffman, V., McClure, E., McKeeman, J., Wood, C.T., Middleton, J.C., Skinner, A.C., Perrin, E.M., Viswanathan, M., 2016. Screening for major depressive disorder in children and adolescents: a systematic review for the U.S. preventive services task force. *Ann. Intern. Med.* 164, 342–349. <https://doi.org/10.7326/M15-2259>.
- GBD 2019 Mental Disorders Collaborators, 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiat.* 9, 137–150. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3).
- González-Blanch, C., Medrano, L.A., Muñoz-Navarro, R., Ruiz-Rodríguez, P., Moriana, J. A., Limonero, J.T., Schmitz, F., Cano-Vindel, A., 2018. Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE* 13, e0193356. <https://doi.org/10.1371/JOURNAL.PONE.0193356>.
- González-Roz, A., Martínez-Loredo, V., Maalouf, W., Fernández-Hermida, J.R., Al-Halabí, S., 2023. Protocol for a trial assessing the efficacy of a universal school-based prevention program for addictive behaviors. *Psicothema* 35, 41–49. <https://doi.org/10.7334/PSICOTHEMA2022.251>.
- Goodman, R., 1997. The strengths and difficulties questionnaire: a research note. *J. Child Psychol. Psychiatry* 38, 581–586.
- Guo, S., Jhe, G.B., 2021. Universal depression screening in schools—promises and challenges in addressing adolescent mental health need. *JAMA Netw. Open* 4, e2132858. <https://doi.org/10.1001/jamanetworkopen.2021.32858>.
- Hambleton, R., Swaminathan, H., Rogers, J., 1991. *Fundamentals of Item Response Theory*. Sage.
- Hankin, B.L., 2015. Depression from childhood through adolescence: risk mechanisms across multiple systems and levels of analysis. *Curr. Opin. Psychol.* <https://doi.org/10.1016/j.copsyc.2015.01.003>.
- Hankin, B.L., Fraley, R.C., Lahey, B.B., Waldman, I.D., 2005. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *J. Abnorm. Psychol.* 114, 96–110. <https://doi.org/10.1037/0021-843X.114.1.96>.
- Horn, J.L., McArdle, J.J., 1992. A practical and theoretical guide to measurement equivalence in aging research. *Exp. Aging Res.* 18, 117–144.
- Hu, L.-T., Bentler, P.M., 1999. Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55.
- Johnson, J.G., Harris, E.S., Spitzer, R.L., Williams, J.B.W., 2002. The Patient Health Questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J. Adolesc. Heal.* 30, 196–204. [https://doi.org/10.1016/S1054-139X\(01\)00333-0](https://doi.org/10.1016/S1054-139X(01)00333-0).
- Kotov, R., Krueger, R.F., Watson, D., Cicero, D.C., Conway, C.C., DeYoung, C.G., Eaton, N.R., Forbes, M.K., Hallquist, M.N., Latzman, R.D., Mullins-Sweatt, S.N.,

- Ruggero, C.J., Simms, L.J., Waldman, I.D., Waszczuk, M.A., Wright, A.G.C., 2021. The hierarchical taxonomy of psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. *Annu. Rev. Clin. Psychol.* 17, 83–108. <https://doi.org/10.1146/annurev-clinpsy-081219-093304>.
- Kroenke, K., 2021. PHQ-9: global uptake of a depression scale. *World Psychiatr.* 20, 135–136. <https://doi.org/10.1002/WPS.20821>.
- Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. <https://doi.org/10.1046/J.1525-1497.2001.016009606.X>.
- Leung, D.Y.P., Mak, Y.W., Leung, S.F., Chiang, V.C.L., Loke, A.Y., 2020. Measurement invariances of the PHQ-9 across gender and age groups in Chinese adolescents. *Asia. Pac. Psychiatry* 12. <https://doi.org/10.1111/APPY.12381>.
- Muñoz-Navarro, R., Cano-Vindel, A., Moriana, J.A., Medrano, L.A., Ruiz-Rodríguez, P., Agüero-Gento, L., Rodríguez-Enríquez, M., Pizà, M.R., Ramírez-Manent, J.I., 2017. Screening for generalized anxiety disorder in Spanish primary care centers with the GAD-7. *Psychiatry Res.* 256, 312–317. <https://doi.org/10.1016/J.PSYCHRES.2017.06.023>.
- NICE Guidance, 2019. *Depression in Children and Young people: Identification and Management*. NICE guideline [NG134].
- Olino, T.M., Yu, L., Klein, D.N., Rohde, P., Seeley, J.R., Pilkonis, P.A., Lewinsohn, P.M., 2012. Measuring depression using item response theory: an examination of three measures of depressive symptomatology. *Int. J. Method. Psychiatr. Res.* 21, 76–85. <https://doi.org/10.1002/mpr.1348>.
- Ortuño-Sierra, J., Sebastián-Enesco, C., Pérez-Albéniz, A., Lucas-Molina, B., Fonseca-Pedrero, E., 2022. Spanish normative data of the Strengths and Difficulties Questionnaire in a community-based sample of adolescents. *Int. J. Clin. Health Psychol.* 22. <https://doi.org/10.1016/J.IJCHP.2022.100328>.
- Patra, K.P., Kumar, R., 2023. Screening for depression and suicide in children. *StatPearls*.
- Piqueras, J.A., García-Olcina, M., Rivera-Riquelme, M., Martínez-González, A.E., Cuijpers, P., 2021. DetectaWeb-distress scale: a global and multidimensional web-based screener for emotional disorder symptoms in children and adolescents. *Front. psychol.* 12, 627604. <https://doi.org/10.3389/fpsyg.2021.627604>.
- Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., Power, M., Duer, W., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J., 2010. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Qual. Life Res.* 19, 1487–1500. <https://doi.org/10.1007/S11136-010-9706-5>.
- Rhew, I.C., Monahan, K.C., Oesterle, S., Hawkins, J.D., 2016. The communities that care brief depression scale: psychometric properties and criterion validity. *J. Commun. Psychol.* 44, 391–398. <https://doi.org/10.1002/jcop.21766>.
- Richardson, L.P., McCauley, E., Grossman, D.C., McCarty, C.A., Richards, J., Russo, J.E., Rockhill, C., Katon, W., 2010. Evaluation of the Patient Health Questionnaire (PHQ-9) for detecting major depression among adolescents. *Pediatrics* 126, 1117. <https://doi.org/10.1542/PEDS.2010-0852>.
- Riglin, L., Petrides, K.V., Frederickson, N., Rice, F., 2014. The relationship between emotional problems and subsequent school attainment: a meta-analysis. *J. Adolesc.* 37, 335–346. <https://doi.org/10.1016/J.ADOLESCENCE.2014.02.010>.
- Shorey, S., Ng, E.D., Wong, C.H.J., 2022. Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. *Br. J. Clin. Psychol.* 61, 287–305. <https://doi.org/10.1111/BJC.12333>.
- Sinclair-McBride, K., Morelli, N., Gusman, M., 2018. PHQ-9 administration in outpatient adolescent psychiatry services. *Psychiatr. Serv.* 69, 837–838. <https://doi.org/10.1176/appi.ps.201800145>.
- Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J.B., Jones, P., Kim, J.H., Kim, J.Y., Carvalho, A.F., Seeman, M.V., Correll, C.U., Fusar-Poli, P., 2022. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* 27, 281–295. <https://doi.org/10.1038/S41380-021-01161-7>.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. *JAMA* 282, 1737–1744. <https://doi.org/10.1001/JAMA.282.18.1737>.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166, 1092–1097. <https://doi.org/10.1001/ARCHINT.166.10.1092>.
- Thapar, A., Eyre, O., Patel, V., Brent, D., 2022. Depression in young people. *Lancet* 400, 617–631. [https://doi.org/10.1016/S0140-6736\(22\)01012-1](https://doi.org/10.1016/S0140-6736(22)01012-1).
- Tsai, F.J., Huang, Y.H., Liu, H.C., Huang, K.Y., Huang, Y.H., Liu, S.I., 2014. Patient health questionnaire for school-based depression screening among Chinese adolescents. *Pediatrics* 133, e402–e409. <https://doi.org/10.1542/peds.2013-0204>.
- WHO, 2021. *WHO Guideline On School Health Services*. WHO.
- World Health Organization, 2022. *World mental health report: transforming mental health for all*. WHO.