



Psychometric Properties of the PHQ-A among a Sample of Children and Adolescents in Puerto Rico

Stephanie López-Torres¹ · Coralee Pérez-Pedrogo^{1,2} · Israel Sánchez-Cardona¹ · Marizaida Sánchez-Cesáreo^{1,2}

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

This study aimed to examine the psychometric properties of the Spanish translation of the Patient Health Questionnaire-9 modified for adolescents (PHQ-A) on a sample of children and adolescents. This is especially pertinent considering that depression often goes unnoticed among children and adolescents, and research about screening instruments is limited. The sample consisted of secondary data of 567 students residing in Puerto Rico. We conducted a confirmatory factor analysis to test the factorial structure of the PHQ-A. A multi-group confirmatory factor analysis (MG-CFA) was performed to examine invariance across gender, and correlations between depression, anxiety and trauma scores were evaluated as evidence of convergent validity. Results indicated a better fit when removing the ninth item of the scale which evaluates suicidal ideation, making reference to the PHQ-8. The MG-CFA also suggests that symptoms of depression are similar between youth despite gender. PHQ-8 scores were positively correlated with anxiety and trauma scores, which evidence convergent validity. In conclusion, the PHQ-8 modified for adolescents (PHQ-8-A) could prove to be a useful instrument for the assessment and early treatment of depression symptoms considering groups who are at-risk such as minorities and youth.

Keywords Adolescents · Children · Depression · Instrumental study · PHQ-A

Introduction

The Patient Health Questionnaire-9 modified for adolescents (PHQ-A) is suggested to be an accessible, reliable and commonly used instrument for the assessment of depression symptoms among children and adolescents (Johnson et al. 2002). Based on these findings, the following research aims to evaluate the psychometric properties of the Spanish version of the PHQ-A among a sample of children and adolescents,

from grades fifth through twelfth that reside in Puerto Rico. We conducted a confirmatory factor analysis to test the factorial structure of the PHQ-A. Additionally, a multi-group confirmatory factor analysis was performed to examine invariance across gender. Correlations between PHQ scores, the Trauma Screening Questionnaire (TSQ) and the Generalized Anxiety Disorder 7 item – adapted for adolescents (GAD-7-A) were evaluated as evidence of convergent validity.

Depression among Hispanic Youth

An estimated 9% of adolescents in the United States between the ages of 12 to 17 reported at least one episode of major depression with severe impairment in the year 2016 (Center for Behavioral Health Statistics and Quality 2017). Since depression is on the rise, early interventions strategies have gained importance especially when depression is frequently associated to anxiety disorders, drug use, high risk sexual behaviors, and risk for suicide (up to 16% of adolescents in US have considered commit suicide) (Centers for Disease Control and Prevention 2017; Hallfors et al. 2004; Jennings et al. 2018). Several risk factors contribute to the development of depression disorders among youth, including parent mental health problems, lower socioeconomic status, and the

✉ Stephanie López-Torres
slopez046@sunmail.albizu.edu

Coralee Pérez-Pedrogo
coperez@albizu.edu

Israel Sánchez-Cardona
isanchez2@albizu.edu

Marizaida Sánchez-Cesáreo
mcsanchez@albizu.edu

¹ Instituto Tercera Misión, Carlos Albizu University 00902-3711, Puerto Rico

² Division of Community Services, University of Puerto Rico, Medical Sciences Campus San Juan Puerto Rico

interaction between sex hormones and other developmental factors (Klasen et al. 2015; Thapar et al. 2012). In addition, children and adolescents who are Latino and Hispanic suffer from higher rates of depression than their peers, are less likely to receive mental health treatment, and are at a higher risk of engaging in delinquency (Jennings et al. 2018; Ramirez et al. 2017). When evaluating depression among Puerto Rican children and adolescents, most population-based studies tend to assess those living in the United States, and only a few consider Puerto Ricans that live on the main island (Langellier et al. 2012). Children and adolescents that live on the island also present more health disadvantages when compared with those that live in the United States, with some authors suggesting poverty prevalence as a main risk factor (Langellier et al. 2012). In the epidemiological study titled *Consulta Juvenil IX*, 3986 students from seventh to twelfth grade in 155 private and public schools of Puerto Rico were surveyed to evaluate risk and protective factors associated with substance abuse. Results indicated that between the years 2012 to 2013, 19.3% of participants presented symptoms of major depression. Significant differences were found by gender, where females had a higher prevalence of major depression (23.2% vs. 14.5%) and suicidal ideation (24.9% vs. 12.7%) than males (Colón, 2013). It seems evident that depression is a serious public health issue and properly validated instruments are required for its evaluation and subsequent referral for early treatment, especially considering racial and regional disparities.

Patient Health Questionnaire-9 Modified for Adolescents, Spanish Version

The PHQ-9 is a brief and self-administered instrument that has been frequently used and validated to measure depression symptoms and their severity among adults in both clinical and research settings (Kroenke et al. 2001; Kroenke and Spitzer 2002). As a self-administered instrument it is compared in a superior way to instruments administered by clinicians, and has been translated to several languages including Spanish, Arabic, Thai, Portuguese, Italian, Dutch, German, Malay, and Konkani (Gilbody et al. 2007).

The PHQ-A, a modified version which examines the depression symptomatology among youth was developed, with its sensitivity and specificity being similar to adult samples (Johnson et al. 2002; Richardson et al. 2010). The PHQ-A is based on the original scale as it has the same number of items. It is based on the diagnostic criteria of the DSM-IV and includes a question regarding functional impairment. Although the Spanish version of the instrument is available and might present favorable properties, psychometric studies among Hispanic adolescents are limited. This is unfortunate considering that the majority of depressed youth are commonly

undiagnosed and untreated for depression and there are limited data on screening tools (Williams et al. 2009; Young et al. 2010).

We consider that the PHQ-A is a brief, convenient and favorable instrument for primary care, community and clinical health settings. Therefore, more research is needed to evaluate its psychometric properties especially among at-risk samples such as Hispanic youth. For this study we aim to evaluate the psychometric properties of the Spanish version of the PHQ-A among a sample of children and adolescents through an analysis of reliability, validity, and factor analysis. We expect that the PHQ-A will present optimal properties in both male and female participants, and will perform equally among middle and high school samples.

Method

Participants and Procedures

The sample consisted of 567 students (298 Female; 266 Male) from fifth to twelfth grade that attended public schools located in the southern part of Puerto Rico (see Table 1). Of the 567 participants, 46.5% attended middle school and 50.2% attended high school at the time of screening. The data did not include the age of participants and the mean age could not be obtained.

Secondary data were acquired from the service program titled: Traumatic Events in Children: An Evidence-Based

Table 1 Sample Socio-demographics

Variable	<i>f</i>	%
Gender		
Female	298	52.6
Male	266	46.9
No response	3	0.5
Grade		
5th grade	25	4.4
6th grade	58	10.2
7th grade	96	16.9
8th grade	84	14.8
9th grade	76	13.4
10th grade	88	15.5
11th grade	68	12.0
12th grade	53	9.3
No response	19	3.4
Intention to leave school		
No	468	85.2
Yes	75	13.2
No response	24	4.2

Step-Care Model as Alternative for Its Management and Mitigation. We obtained permission to use the data from the principal investigators, which provided data sets with de-identified information of the participants. This secondary data analysis study was also approved by the Institutional Review Board at Carlos Albizu University, San Juan Campus. Raw data of the PHQ-A as well as data of anxiety and trauma measures were analyzed to examine the construct and convergent validity of the PHQ-A. Independent sample *t*-tests were performed to test for mean differences considering gender and school level.

We obtained the informed consent and assent from all individual participants for whom identifying information is not included in this article. The students that participated in the study were recruited after obtaining a consent form from their primary caregiver and after filling out an assent form themselves. Staff from the service program individually provided information about the program to caregivers and students, and explained the content of the consent form and their rights. The screening instruments that measured depression, anxiety, and trauma would then be administered to students that assented. The consent forms were handed out after parents attended a report card meeting, while the assent form and screening instruments were administered on another day while the students were in school. Students that could not participate due to not having the signed consent or not assenting were then handed an “activity book” while participants finished the screening. This process was previously discussed and planned with the director and staff of each of the schools. Staff from the service program were previously trained for the management of adverse experiences, and psychologists were available on-site to manage possible crises.

Instruments

Patient Health Questionnaire-9 Modified for Adolescents (PHQ-A), Spanish Version This instrument has nine items for the evaluation of depression symptoms and their severity based on the criteria of the Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition (DSM-IV). It is adapted to be used among adolescents and can be administered in multiple settings such as primary care, clinical and research. Item scores range from zero to three, depending on the frequency of symptoms in the last two weeks, where zero means absence of symptoms or “never” having experienced symptoms, a score of one references a low frequency (“several days”, 1–6 days), a score of two references a regular frequency (“more than half the days”, 7–12 days), while a score of three means high frequency of symptoms (“nearly every day”, 12–14 days). The participant is likely to present clinical depression when he/she scores two to three in five of the nine items, including scores on the first two items that explore symptoms of anhedonia and depressed mood. Similar to the

PHQ-9, the lowest total score is zero and the highest score is 27, where a score between 10 and 14 is a moderate degree of depression, 15 to 19 is moderate to severe, and 20–27 is severe (Johnson et al. 2002; Kroenke and Spitzer 2002). The Spanish version used for this study was previously translated and adapted using the cultural equivalence model (Chávez and Canino 2005).

Trauma Screening Questionnaire (TSQ), Spanish Version The TSQ is a brief 10-item self-report questionnaire for the screening of Posttraumatic Stress Disorder (PTSD). It is based on diagnostic criteria from the DSM-IV and items from the PTSD Symptom Scale - Self Report (PSS-SR; Foa et al. 1993), having five re-experiencing items and five arousal items. It was originally developed by Brewin et al. (2002) with the purpose of providing a brief screening instrument for likely PTSD. Authors consider a screen positive when at least 6 items are endorsed. This cutoff point has presented high levels of sensitivity (.86) and specificity (.93) (Brewin et al. 2002). The Spanish version of this instrument also presented favorable reliability estimates among children and adolescents that reside in Puerto Rico, presenting a Cronbach alpha of .80 (Pérez-Morales et al. 2019).

Generalized Anxiety Disorder 7-Item – Adolescent Version (GAD-7-a), Spanish Version The GAD-7 is a self-administered 7-item scale designed to assess symptoms of anxiety. It was developed by Spitzer, Kroenke, Williams, & Löwe in 2006 as a brief measure for anxiety symptoms. Each item is scored on a 0 to 3 point scale (“not at all” to “nearly every day”). Similar to the adult version, cut points of 5, 10, and 15 correspond to mild, moderate, and severe symptoms of anxiety. The instrument has presented good reliability and validity, with a cutoff point of cut point of 10 or greater was identified to optimized sensitivity (89%) and specificity (82%). The internal consistency of the scale was also excellent, resulting in a Cronbach α of .92 (Spitzer et al. 2006). It has presented acceptable specificity and sensitivity among a sample of adolescents, and the Spanish translation has presented optimal properties when administered to Hispanic samples (Mills et al. 2014; Mossman et al. 2017). For the following study, the Spanish version of the scale was used after being revised and adapted for Puerto Rican adolescents (GAD-7-A, (Castro-Díaz et al. 2018).

Data Analysis

Analyses were conducted with SPSS v. 23 and AMOS v. 23. First, descriptive analyses were performed to assess missing values. Analysis showed missing data was less than 5% in each variable (range from 0% to 4.8%). To address incomplete data, missing points were estimated with values derived from regression analysis (Byrne 2001; Kline 2011;

Table 2 Baseline Model PHQ-A (n = 567)

Model	χ^2	df	RMSEA	SRMR	CFI	TLI	AIC	$\Delta\chi^2$	Δdf	ΔCFI
M1 PHQ-A	175.01	27	.098	.050	.920	.89	229.01			
M2 PHQ-A revised (error covariance item 6 & 9)	125.52	26	.082	.042	.947	.926	181.518			
Diff. M1 & M2								48.49**	1	.027
M3 PHQ-8-A	100.98	20	.085	.040	.947	.926	148.977			
Diff. M2 y M3								24.54**	6	.000

n = 567; χ^2 = Chi-square; df = degree of freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized root mean square residual; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; AIC = Akaike Information Criterion; $\Delta\chi^2$ = difference in chi square; ΔCFI = difference in CFI; $\Delta RMSEA$ = Difference in RMSEA. * $p < 0.05$; ** $p < 0.01$

Tabachnick and Fidell 2013). Sample size for further analysis was $n = 567$.

We examined the PHQ-A factor structure with a Confirmatory Factor Analysis (CFA) and multi-group confirmatory factor analysis (MG-CFA) for invariance test using AMOS v. 23 (Arbuckle, 2014). First, a one-factor model was tested with PHQ-A items loading to a single factor. We re-specified the model based on modification indices, factor loadings, and goodness of fit indices. In line with available literature (Alpizar et al. 2017; Schantz et al. 2017), a second one-factor model was tested with only 8 items.

Once we established the best factorial structure, we conducted a multi-group confirmatory factor analysis (MG-CFA) in order to test invariance across gender. We proceeded to implement a logically organized strategy to test for multi-group invariance following the recommendation from Byrne (2008, 2016a). First, a baseline model for both female and male samples was fitted independently to test the best fitting model in both groups separately. Second, a configural model (no constraints) was fitted, in which no equality constraints are imposed. Second, a metric invariant model (*measurement equivalence*) was estimated, in which factor loadings are constrained equal across groups. Third, a *scalar invariant model* was assessed where the factor loadings and intercepts are constrained to be equal. Following recommendations, we do not examine equality of error variances since it represents and overly restricted test of the data (Byrne 2016b). All tests for equivalence were compared against this configural model.

Model fit was assessed with multiple absolute and relative indices of goodness-of-fit: chi-square (χ^2), the Root Mean Square Error of Approximation (RMSEA), the Standardized

Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI) and Tucker Lewis Index (TLI). Additionally, the Akaike Information Criterion (AIC) was evaluated; the lower the value, the better the fit. Values greater than .90 are indicative of good fit for CFI and TLI, values of RMSEA below .08 and .05 indicate a reasonable and good fit, respectively, while SRMR values lower than .08 are indicative of a good fit (Hu and Bentler 1999).

Since these models are nested, models were compared to each other using change values in χ^2 statistics and the related degrees of freedom. If $\Delta\chi^2$ value of two nested models is significant, it suggests that the constraints specified in the more restrictive model do not hold (Byrne 2008). Moreover, due to χ^2 sensitivity to sample size, we examined difference on ΔCFI to determine invariance. Values not exceeding 0.01 indicate invariance of the model (Cheung and Rensvold 2002).

Validity and Reliability

GAD and TSQ were used to examine convergent validity of the PHQ-A. As in previous studies (Comellas et al. 2015; Flory 2015), correlations were conducted to evaluate whether PHQ scores were positively associated with anxiety and trauma symptoms. We expect a strong and positive correlation among these constructs with scores being higher than .50 (Cohen et al. 2003). In addition, independent sample t test were performed to test gender differences in the PHQ scores. Finally, we assessed reliability of the PHQ measures using Cronbach alpha coefficient and composite reliability (Raykov and Shrout 2002). Reliability coefficients greater

Table 3 Baseline Model for Females and Males

Model	χ^2	df	RMSEA	SRMR	CFI	TLI	AIC	$\Delta\chi^2$	Δdf	ΔCFI
PHQ 8-A Female	67.95	20	.089	.043	.945	.923	115.95			
PHQ-8-A Male	89.75	20	.114	.058	.899	.859	137.75			
PHQ-8-A Male error covariance item 2 and 6	58.65	19	.088	.048	.94	.916	108.65	31.10**	1	.041

n = 567; χ^2 = Chi-square; df = degree of freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized root mean square residual; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; AIC = Akaike Information Criterion; $\Delta\chi^2$ = difference in chi square; ΔCFI = difference in CFI; $\Delta RMSEA$ = Difference in RMSEA. * $p < 0.05$; ** $p < 0.01$

Table 4 Goodness-of-Fit Statistics for the PHQ-8-A Test of Invariance across Gender

Model	χ^2	df	RMSEA	SRMR	CFI	TLI	AIC	$\Delta\chi^2$	Δdf	ΔCFI
M1 Configural model	114.71	39	.059	.050	.948	.926	212.713			
M2 Metric Invariance	121.06	46	.054	.050	.949	.938	234.723	6.35 n.s.	7	.001
M3 Scalar Invariance	141.88	54	.054	.050	.940	.938	209.88	20.87**	8	.008

$n = 567$; χ^2 = Chi-square; df = degree of freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized root mean square residual; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; AIC = Akaike Information Criterion; $\Delta\chi^2$ = difference in chi square; ΔCFI = difference in CFI; $\Delta RMSEA$ = Difference in RMSEA. * $p < 0.05$; ** $p < 0.01$

than 0.70 are generally considered as acceptable (Hair et al. 2006; Nunnally and Bernstein 1994).

Results

Confirmatory Factor Analysis of PHQ

Factor structure of the PHQ-A was tested with Confirmatory Factor Analysis. The one-factor structure of PHQ-A fit the data adequately (See Table 2). Examination of the modification indices showed that co-variation of measurement residual of item 6 and item 9, will improve model fit. The significant chi-square difference test showed an improvement of this model. Previous research suggest that item 9 which evaluates self-harm and suicidal ideation does not discriminate well between participants who are likely to be depressed from those who are not (Kroenke and Spitzer 2002), and also presents a lower factor loading (Merz et al. 2014). As recommended for large-scale community samples (Kroenke et al. 2009), we decided to test an 8-items PHQ one-factor model. This model presented a good fit to the data (see Table 2) and all items loaded significantly ($ps < .05$) and ranged from .58 to .77. We retained this PHQ-8 modified for adolescents (PHQ-8-A) as the best model.

Multi-Group Confirmatory Factor Analysis

We further analyze the PHQ-8-A invariance across gender. In order to establish a well-fitting baseline model for each group separately, CFA were conducted for female ($n = 298$) and male ($n = 266$) samples. Results of this CFA showed that the PHQ-

Table 5 Mean, Standard Deviation, Correlation and Reliability Estimates

	Mean	SD	CR	α	1	2
PHQ-8-A	5.02	5.03	.86	.85		
GAD-7-A	4.76	4.53	.86	.85	.763**	
TSQ	3.05	2.73	.81	.81	.618**	.643**

** $p < .01$; Cronbach's alpha on the diagonal

8-A model has a good fit to the data for the female sample; however, for the male sample model fit was improved if a covariance between measurement error of item 2 and 6 was specified (see Table 3). This covariance significantly improved the PHQ-8-A model fit for male sample ($\Delta\chi^2_{(1)} = 31.10, p < .01$). Factor loadings ranged between .59 and .77 in the female sub-sample, and .56 and .66 in the male sample; all statistically significant ($ps < .05$).

This best fitting model was tested across group of female and male. Since the baseline model for male includes an additional covariance among measurement error, this parameter was not constraint equal across groups. All other parameters were constrained equal to test for measurement and scalar invariance. Table 4 shows the results for the MG-CFA. Results showed metric invariance across gender. Value of $\Delta\chi^2$ was not significant and ΔCFI was less than .01, which may confirm metric invariance across gender. Additionally, although $\Delta\chi^2$ was statistically significant, ΔCFI was less than .01 indicating scalar invariance across gender. Thus, PHQ-8-A shows invariant factor structure across female and males.

Convergent Validity and Reliability of the PHQ-A

Table 5 shows descriptive statistics, correlations, Cronbach's alpha, and composite reliabilities. As expected, PHQ-8-A scores were strongly and positively correlated with anxiety ($r = .763, p < .01$) and trauma ($r = .618, p < .01$) scores. These correlations were also positive and strong in male and female samples (see Table 6). This provides evidence of the convergent validity of the scale. Mean differences were found in the PHQ-8-A among female and male, $t_{(557.71)} = -3.69, p < 0.01, d = .31$. Females reported significantly higher mean in the PHQ-8-A ($M = 5.69, SD = 5.36$) compared to males ($M = 4.17, SD = 4.38$). In terms of grade level, significant score differences were also observed $t_{(535.54)} = -4.90, p < 0.01, d = -.42$. High school students presented a higher mean of depression symptoms ($M = 6.04, SD = 5.41$) when compared to middle school students ($M = 3.97, SD = 4.38$).

Cronbach's alpha of the PHQ-8-A was above the recommended value of .70 ($\alpha_{females} = .87; \alpha_{male} = .80$). As well, composite reliability was above .70 ($CR_{female} = .86$;

Table 6 Mean, Standard Deviation, Correlation and Reliability Estimates by Gender

	<i>Mean Female</i>	<i>SD Female</i>	<i>Mean Male</i>	<i>SD Male</i>	α Female	<i>CR Female</i>	α Male	<i>CR Male</i>	1	2	3
PHQ-8-A	5.56	5.36	4.18	4.38	.87	.86	.80	.86		.702**	.558**
GAD-7-A	5.50	4.92	3.88	3.81	.86	.82	.83	.87	.794**		.534**
TSQ	3.42	2.91	2.60	2.42	.83	.79	.76	.83	.631**	.684**	

** $p < .01$; Correlation for female sample are shown below the diagonal. Correlations for the male sample are shown above the diagonal

$CR_{\text{male}} = .84$; $CR_{\text{middle school}} = .84$, $CR_{\text{high school}} = .86$) supporting reliability of the instrument (see Tables 5, 6, and 7).

Discussion

The present study aims to examine the psychometric properties of the PHQ-A among a sample of children and adolescents residing in Puerto Rico. On the basis of the results presented, the PHQ-A proved a better fit when the ninth item was removed. This version of the instrument makes reference to the Patient Health Questionnaire-8, which has the exact items of the original Patient Health Questionnaire-9 but omits the ninth item. This item evaluates self-harm and suicidal ideation. Although suicidal ideation is relevant to depression symptoms, some authors suggest that the ninth item of the PHQ-9 can present a lower factor loading because it does not discriminate well between participants who are likely to be depressed from those who are not (Kroenke & Spitzer., 2002). As a result, not endorsing the ninth item is common since it is not a sensitive indicator for depression, being retained strictly for clinical settings where follow-up is available and practical (Duffy et al. 2008; Huang et al. 2005). It is also relevant to mention that almost all of the positive responses to the ninth item tend to represent passive thoughts of death rather than suicidal ideation (Dhingra et al. 2011).

The main rationale for removing the ninth item was because of sensitivity reasons and to have a reliable screening tool for the general population (Kroenke et al. 2009). Research suggests that the ninth item of the PHQ-9 is normally associated to passive thoughts of dying rather than suicidal ideation (Dhingra et al. 2011). The PHQ-8 is also favored for clinical research where depression is a secondary outcome and for large-scale epidemiological studies. This especially considering that the PHQ-8 retains high levels of sensitivity (.88–.91)

and specificity (.88–.99) among the general population, with the same diagnostic cut-point (≥ 10) of the PHQ-9 (Dhingra et al. 2011; Kroenke and Spitzer 2002; Razykov, Ziegelstein, Whooley, & Thombs, 2012). The PHQ-8 has also been validated among the general population (Kroenke et al. 2009), adolescents (Peh et al. 2017), and adults who are Latino/a (Alpizar et al. 2017). To the best of our knowledge, a PHQ-8 modified for adolescents has not been considered prior to this study.

After retaining the PHQ-8-A as the best model, we tested for invariance across both groups of female and male participants. Results of the MG-CFA showed metric and scalar invariance across gender, suggesting that the participants across groups attribute the same meaning to the latent construct and presented similar factor loadings and level of underlying items (intercepts). We can conclude that the PHQ-8-A presented an equivalent factor structure in both male and female participants. Thus, the PHQ-8-A can be used in both gender groups effectively. Results of invariance also suggested that among male participants a covariance between measurement error of item 2 and 6 improved the model fit. Future exploration of this covariance among male participants should be further evaluated, as it makes reference to self-conscious moral emotions such as shame, guilt, embarrassment and pride (Else-Quest et al. 2012).

In terms of limitations, it is important to note that the sample used for this study was not representative of the general population of youth living in Puerto Rico. The sample of this study consisted of students that attended public schools in the southern and rural area of Puerto Rico. This study did not include information regarding area of residency, and did not have information about students from schools located in urban areas. It is recommended to further research the psychometric properties of the PHQ-8 and the PHQ-8-A among various

Table 7 Mean, Standard Deviation, Correlation and Reliability Estimates by Grade Level

	<i>Mean Middle School</i>	<i>SD Middle School</i>	<i>Mean High School</i>	<i>SD High School</i>	α Middle School	<i>CR Middle School</i>	α High School	<i>CR High School</i>	1	2	3
PHQ-8-A	3.97	4.38	6.04	5.41	.83	.84	.86	.86		.722**	.594**
GAD-7-A	3.87	4.02	5.56	4.73	.83	.85	.87	.88	.768**		.658**
TSQ	2.76	2.59	3.31	2.82	.79	.80	.87	.87	.621**	.620**	

** $p < .01$; Correlation for middle school sample are shown below the diagonal. Correlations for the high school sample are shown above the diagonal.

populations to further expand the validity and reliability of the instrument. Studies that evaluate cut-off scores of the PHQ-8 in specific groups such as children, adolescents and persons who are Latino/a are scarce in comparison to studies on the sensitivity and specificity of the PHQ-9. Although the suggested cut-off score of the PHQ-8 is similar to the PHQ-9, cut-off scores can range depending on the sample's characteristics (Manea et al. 2012). Cut-off scores of the PHQ-9 among adolescents suggest a cut-off scores ranging from ≥ 8 to ≥ 11 (Allgaier et al. 2012; Richardson et al. 2010), while among samples of adults who are Latino/a this number ranged from ≥ 9 to ≥ 10 (Arrieta et al. 2017; Ramirez-Solá 2017). This is especially pertinent to address since studies that evaluate a cut-off score for Latino youth are limited. Research should be conducted to identify the sensitivity and specificity of the PHQ-8 and the PHQ-8-A to establish an optimal cut-off score for depression screening among youth residing in Puerto Rico and other populations that are Latino and Hispanic. We also suggest further research to evaluate multi-group invariance considering grade level.

The present research provides information regarding the psychometric properties of an accessible and reliable instrument, which can facilitate the assessment of depression symptoms and the development of proper intervention strategies among youth in Puerto Rico. To the best of our knowledge, little attention has been given to research evaluating the validity of the PHQ-8 among children and adolescents, and the PHQ-8 modified for adolescents has not been considered before. Most other validation studies consider the original PHQ-9 and the modified version for adolescents. Based on the obtained results, the PHQ-8 modified for adolescents – or PHQ-8-A proves to be a useful instrument for the assessment of depression symptoms among youth in large scale studies. Thus, examining the validity, specificity and sensibility of a measure that can be used in community setting as well as clinical facilities is highly relevant for public health.

Funding Information This work was supported by the Victim Compensation Grant Program, Department of Justice, grant number VA-GX-0039.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Allgaier, A.-K., Pietsch, K., Frühe, B., Sigl-Glöckner, J., & Schulte-Körne, G. (2012). Screening for depression in adolescents: Validity of the patient health questionnaire in pediatric care. *Depression and Anxiety, 29*(10), 906–913. <https://doi.org/10.1002/da.21971>.
- Alpizar, D., Plunkett, S. W., Whaling, K., Alpizar, D., Plunkett, S. W., & Whaling, K. (2017). Reliability and Validity of the 8-Item Patient Health Latino Emerging Adults for Measuring Depressive Symptoms of Latino Emerging Adults.
- Arrieta, J., Aguerrebere, M., Raviola, G., Flores, H., Elliott, P., Espinosa, A., et al. (2017). Validity and utility of the patient health questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: A cross-sectional study. *Journal of Clinical Psychology, 73*(9), 1076–1090. <https://doi.org/10.1002/jclp.22390>.
- Brewin, C. R., Rose, S., Andrews, B., Green, J., Tata, P., McEvedy, C., et al. (2002). Brief screening instrument for post-traumatic stress disorder. *British Journal of Psychiatry, 181*, 158–162. <https://doi.org/10.1192/bjp.181.2.158>.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah: Lawrence Erlbaum Associates.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872–882 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18940097>.
- Byrne, B. (2016a). Testing instrument equivalence across cultural groups: Basic concepts, testing strategies, and common complexities. In N. Zane, G. Bernak, & F. T. L. Leong (Eds.), *Evidence-based psychological practice with ethnic minorities: Culturally informed research and clinical strategies* (pp. 125–143). Washington: American Psychological Association.
- Byrne, B. M. (2016b). *Structural Equation Modeling : A Testing for Multigroup Invariance Using AMOS Graphics : A Road Less Traveled TEACHER ' S CORNER Testing for Multigroup Invariance Using AMOS Graphics : A Road Less Traveled*, (April 2013), 37–41. <https://doi.org/10.1207/s15328007sem1102>
- Castro-Díaz, B., López-Torres, S., Perez-Pedrogo, C., & Sánchez-Cardona, I. (2018). Psychometric properties of the Generalized Anxiety Disorder 7-item scale among Latino youth: An effective screening tool for school and community settings. Manuscript submitted to *Terapia Psicológica*.
- Center for Behavioral Health Statistics and Quality. (2017). *Results from the 2016 National Survey on Drug Use and Health: Detailed Tables*. Rockville, Maryland. Retrieved from <https://www.samhsa.gov/data/sites/default/files/NSDUH-DET-Tabs-2016/NSDUH-DET-Tabs-2016.pdf>
- Centers for Disease Control and Prevention. (2017). Suicide Among Youth. Retrieved from <https://www.cdc.gov/healthcommunication/toolstemplates/entertainmenttips/suicidemyouth.html>
- Chávez, L., & Canino, G. (2005). Toolkit on translating and adapting instruments. *Toolkit on Translating and Adapting Instruments*, 58. Retrieved from http://www.hsri.org/files/uploads/publications/PN54_Translating_and_Adapting.pdf
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Lawrence Erlbaum Associates Publishers.
- Comellas, R. M., Makhshvili, N., Chikovani, I., Patel, V., McKee, M., Bisson, J., & Roberts, B. (2015). Patterns of somatic distress among conflict-affected persons in the republic of Georgia. *Journal of Psychosomatic Research, 78*(5), 466–471. <https://doi.org/10.1016/j.jpsychores.2015.01.015>.
- Colón, H.M, Moscoso Álvarez M.R, Reyes Pulliza J.C., Rodríguez Figueroa L. (2013). El uso de sustancias en los escolares puertorriqueños: Consulta Juvenil IX, 2012-2013. Administración de Servicios de Salud Mental y Contra la Adicción, Bayamón, Puerto Rico.

- Dhingra, S., Kroenke, K., Zack, M., Strine, T., & Balluz, L. (2011). PHQ-8 days: A measurement option for DSM-5 major depressive disorder (MDD) severity. *Population Health Metrics*, 9(11). <https://doi.org/10.1186/1478-7954-9-11>.
- Duffy, F. F., Chung, H., Trivedi, M., Rae, D. S., Regier, D. A., & Katzelnick, D. J. (2008). Systematic use of patient-rated depression severity monitoring: Is it helpful and feasible in clinical psychiatry? *Psychiatric Services*, 59(10), 1148–1154. <https://doi.org/10.1176/appi.ps.59.10.1148>.
- Else-Quest, N. M., Higgins, A., Allison, C., & Morton, L. C. (2012). Gender differences in self-conscious emotional experience: A meta-analysis. *Psychological Bulletin*, 138(5), 947–981. <https://doi.org/10.1037/a0027930>.
- Flory, J. D. (2015). Comorbidity between post-traumatic stress disorder and major depressive disorder: Alternative explanations and treatment considerations. *Dialogues in Clinical Neuroscience*, 17(August), 141–150.
- Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, 6(4), 459–473. <https://doi.org/10.1007/BF00974317>.
- Gilbody, S., Richards, D., Brealey, S., & Hewitt, C. (2007). Screening for depression in medical settings with the patient health questionnaire (PHQ): A diagnostic meta-analysis. *Journal of General Internal Medicine*, 22(11), 1596–1602. <https://doi.org/10.1007/s11606-007-0333-y>.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate Data Analysis* (3rd ed.). Upper Saddle River: Pearson Prentice Hall.
- Hallfors, D. D., Waller, M. W., Ford, C. A., Halpern, C. T., Brodish, P. H., & Iritani, B. (2004). Adolescent depression and suicide risk: Association with sex and drug behavior. *American Journal of Preventive Medicine*, 27(3), 224–231. <https://doi.org/10.1016/j.amepre.2004.06.001>.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Huang, F. Y., Chung, H., Kroenke, K., Delucchi, K. L., & Spitzer, R. L. (2005). Using the patient health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients Frederick. *General Internal Medicine*, 21(6), 547–552. <https://doi.org/10.1111/j.1525-1497.2006.00409.x>.
- Jennings, W. G., Maldonado-Molina, M., Fenimore, D. M., Piquero, A. R., Bird, H., & Canino, G. (2018). The linkage between mental health, delinquency, and trajectories of delinquency: Results from the Boricua youth study. *Journal of Criminal Justice*, 22(2), 84–98. <https://doi.org/10.1016/j.jcrimjus.2018.08.003>.
- 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. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, 30(3), 196–204. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11869927>.
- Klasen, F., Otto, C., Kriston, L., Patalay, P., Schlack, R., & Ravens-Sieberer, U. (2015). Risk and protective factors for the development of depressive symptoms in children and adolescents: Results of the longitudinal BELLA study. *European Child & Adolescent Psychiatry*, 24(6), 695–703. <https://doi.org/10.1007/s00787-014-0637-5>.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.
- Kroenke, K., & Spitzer, R. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32(9), 509–515. <https://doi.org/10.3928/0048-5713-20020901-06>.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, 114(1–3), 163–173.
- Langellier, B. A., Martin, M. A., Canino, G., Garza, J. R., & Ortega, A. N. (2012). The health status of youth in Puerto Rico. *Clinical Pediatrics*, 51(6), 569–573. <https://doi.org/10.1177/0009922812443123>.
- Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the patient health questionnaire (PHQ-9): A meta-analysis. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, 184(3), E191–E196. <https://doi.org/10.1503/cmaj.110829>.
- Merz, E. L., Malcarne, V. L., Roesch, S. C., Riley, N., & Sadler, G. R. (2014). A multigroup confirmatory factor analysis of the patient health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity & Ethnic Minority Psychology*, 17(3), 309–316. <https://doi.org/10.1037/a0023883>.
- Mills, S. D., Fox, R. S., Malcarne, V. L., Roesch, S. C., Champagne, B. R., & Sadler, G. R. (2014). The psychometric properties of the generalized anxiety disorder-7 scale in Hispanic Americans with English or Spanish language preference. *Cultural Diversity & Ethnic Minority Psychology*, 20(3), 463–468. <https://doi.org/10.1037/a0036523>.
- Mossman, S. A., Luft, M. J., Schroeder, H. K., Vamey, S. T., Fleck, D. E., Barzman, D. H., et al. (2017). The generalized anxiety disorder 7-item scale in adolescents with generalized anxiety disorder: Signal detection and validation. *Annals of Clinical Psychiatry : Official Journal of the American Academy of Clinical Psychiatrists*, 29(4), 227–234A Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/29069107>.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, Inc..
- Peh, C. X., Shahwan, S., Fauziana, R., Mahesh, M. V., Sambasivam, R., Zhang, Y. J., et al. (2017). Emotion dysregulation as a mechanism linking child maltreatment exposure and self-harm behaviors in adolescents. *Child Abuse and Neglect*, 67, 383–390. <https://doi.org/10.1016/j.chiabu.2017.03.013>.
- Pérez-Morales, Á., Pérez-Pedrogo, C., López-Torres, S., Martínez-Taboas, A., & Sánchez-Cardona, I. (2019). Validación de una traducción al castellano del Trauma Screening Questionnaire con niños y adolescentes entre 8–17 años de edad residentes en Puerto Rico. Manuscript submitted to *Salud y Sociedad*.
- Ramirez, A. G., Gallion, K. J., Aguilar, R., & Surette Dembeck, Eurette Dembeck, E. (2017). *Mental Health and Latino Kids: A Research Review*. Retrieved from <https://salud-america.org/wp-content/uploads/2017/09/FINAL-mental-health-research-review-9-12-17.pdf>.
- Ramirez-Solá, M. (2017). Distress management in cancer patients in Puerto Rico. *The Journal of Community and Supportive Oncology*, 15(2), 68–73.
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 195–212. https://doi.org/10.1207/S15328007SEM0902_3.
- Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., et al. (2010). Evaluation of the patient health Questionnaire-9 item for detecting major depression among adolescents. *Pediatrics*, 126(6), 1117–1123. <https://doi.org/10.1542/peds.2010-0852>.

- Schantz, K., Reighard, C., Aikens, J. E., Aruquipa, A., Pinto, B., Valverde, H., & Piette, J. D. (2017). Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients. *The International Journal of Psychiatry in Medicine*, 52(4–6), 315–327. <https://doi.org/10.1177/0091217417738934>.
- Spitzer, R., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety DisorderThe GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.
- Thapar, A., Stephan, C., Pine, D., & Thapar, A. (2012). Depression in adolescence. *Lancet*, 379(9820), 1056–1067. [https://doi.org/10.1016/S0140-6736\(11\)60871-4](https://doi.org/10.1016/S0140-6736(11)60871-4).**Depression.**
- Williams, S. B., O'Connor, E. A., Eder, M., & Whitlock, E. P. (2009). *Screening for Child and Adolescent Depression in Primary Care Settings: A Systematic Evidence Review for the US Preventive Services Task Force*. *Pediatrics* (Vol. 123). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2008-2415>
- Young, J. F., Miller, M. R., & Khan, N. (2010). Screening and managing depression in adolescents. *Adolescent Health, Medicine and Therapeutics*, 1, 87–95. <https://doi.org/10.2147/AHMT.S7539>.

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