

**Assessing the Reliability and Validity of the PHQ-9  
Depression Scale Among Mexican Medical Students, 2018**

School of Global Public Health, New York University, New York, NY, USA

Spring 2024, GPH-GU 2225 Psychometric Measurement and Analysis

Helen Liang, Ivy Zhao

## **Abstract**

*Objective:* Major Depressive Disorder (MDD) is a significant public health issue, particularly among medical students due to their high-stress environment. This study focuses on the reliability and validity of the PHQ-9 Depression Scale in assessing depression among Mexican medical students to provide early screening and intervention of depression or depressive symptoms.

*Methods:* A sample of 776 students from a private Mexican university completed the PHQ-9, GAD-7 Anxiety Scale, and Epworth Sleepiness Scale. Data collection occurred in 2014 with a response rate of 64.6%. Reliability was assessed through Cronbach's alpha, and validity was evaluated through existing literature, factor analysis, and pairwise Pearson's correlation.

*Results:* The study found that the PHQ-9 Depression Scale exhibits good internal consistency with a Cronbach's alpha of 0.8097. Factor analysis identified a single factor, "Negative Emotions/Thoughts", explaining 70.04% of the variance. Significant moderate to strong association was observed between the PHQ-9 and GAD-7 ( $r = 0.6796$ ) and a moderate association with the Epworth Sleepiness Scale ( $r = 0.3884$ ), supporting its construct validity.

*Conclusion:* The 9-item PHQ-9 Depression Scale is a reliable and valid tool for screening and early intervention of depression among Mexican medical students and can also be effectively used in diverse settings. However, cultural adaptability should be considered to enhance its applicability and accuracy across different populations. Furthermore, the study's generalizability may be limited to similar educational and

cultural settings. The self-report nature of the PHQ-9 may also introduce biases related to cultural differences in expressing depressive symptoms.

*Keywords:* depression, depression scale, PHQ-9, reliability, validity, Mexican medical students

## **Introduction**

### *Depression and Depression Scales*

Major Depressive Disorder (MDD) is a growing public health concern and one of the leading causes of disability. The symptoms of MDD can be characterized by a lowering of mood, reduction of energy, reduced capacity for enjoyment, disturbed appetite, unresponsive mood, and other physical symptoms <sup>[1]</sup>. Mental disorders are indiscriminate and can affect all types of individuals, but those who have endured trauma or major life events may be more vulnerable. Considering these challenges, it becomes imperative to prioritize mental health and ensure that public health systems are equipped with effective screening methods.

Valid and efficient psychometric tests are crucial to target those at risk and provide appropriate interventions. There are two common self-reported questionnaires for screenings and diagnosis of common mental health disorders: the Patient Health Questionnaire (PHQ-9) with a 9-item depression module and the Generalized Anxiety Disorder (GAD-7) with a 7-item anxiety scale. Both tests are self-administered questionnaires where the PHQ-9 is a tool to identify and measure the severity of depression, and the GAD-7 is a tool that measures the severity of anxiety. This research will primarily focus on the PHQ-9 Depression Scale, assessing its reliability and validity in the diagnosis of depression.

The PHQ-9 consists of 9 criteria, with each scored as “0” (not at all) to “3” (nearly every day), on which the diagnosis of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) about depression is based and allows for a two-step measure where more questions must be asked to establish depression diagnosis when scores are high <sup>[2]</sup>. Thus, it serves as a dual-purpose tool for both diagnosing depressive disorders and grading the severity of depressive symptoms.

### *Depression Among Mexican Medical Students*

MDD among Mexican medical students, particularly, is a public health concern that requires more attention. In Mexico, the prevalence of depression among medical undergraduates is typically around 20.2%, with 16.2% of them experiencing major depressive disorder <sup>[3]</sup>. Globally, medical students face more financial, health, and academic burdens than the average college student, making them a sensitive population prone to mental health disorders such as depression and anxiety <sup>[4]</sup>. In Mexico, the path to medical school generally includes four years of undergraduate education followed by four years in medical school. Even after graduating from medical school, medical school graduates enter a residency period that can last three to seven years. Furthermore, as of 2023, the average medical school graduate was burdened with more than \$240,000 in total student debt <sup>[5]</sup>. The high-stakes environment, combined with the pressure to pay off loans, fear of making mistakes, and the emotional toll of dealing with patients' suffering and death, can significantly impact their mental health <sup>[6]</sup>.

## **Methods**

### *Participants and Data Collection*

The participants in this research are medical students at a private Mexican university. The data were collected from a cohort of 776 Mexican medical students in 2014. Initially, 1200 medical students were randomly selected from a total enrollment of 4,520 students at the university, representing a response rate of 64.6%. Participants completed the self-administered questionnaires: PHQ-9 Depression Scale, GAD-7 Anxiety Scale, and the Epworth Sleepiness Scale, which measure the severity of depression, severity of anxiety, and daytime sleepiness<sup>[7]</sup>, respectively.

### *Statistical Analyses*

Reliability was assessed by examining the internal consistency of the PHQ-9 through Cronbach's alpha. It is a useful tool to measure reliability in the health science community, as it measures all items in a scale measure are interrelated under the same construct<sup>[8]</sup>. Cronbach's alpha is expressed as a number ranging from zero to one, where levels above 0.7 indicate acceptable internal consistency. However, for an established research tool such as the PHQ-9, a Cronbach's alpha of at least 0.8 is required for acceptability, and an alpha above 0.9 would suggest excellent internal consistency, thus confirming the tool as reliable.

Validity, including face, content, and construct validity, were evaluated through existing literature, factor analysis, and pairwise Pearson's correlation with p-value. While a predictor for a certain type of psychopathology cannot be proven accurate unless it can be shown that the predictor is connected to a reliable sign of that mental illness<sup>[9]</sup>, Construct validity will be implemented to measures whether a tool measures the construct it claims to measure, in this case, depression. Factor analysis separates variables into groups and variables under the same factor group should show high

correlations, thus confirming construct validity. Factor analysis scores above 0.4 is considered acceptable and stable. Items with scores lower than 0.3 will be suppressed as they may not tap into the construct. Pairwise Pearson's correlation will be used to assess the correlations between the PHQ-9 and the GAD-7/Epworth Sleepiness Scale. By examining the correlations between these constructs, construct validity can be further confirmed by ensuring that both tools are effectively measuring related mental disorders.

## **Results**

StataSE Version 18.0 is used for all data analyses performed in this study. Table 1 presents key characteristics of Mexican medical students at a private Mexican university (N = 776). Students in Year 1 (35.44%) comprised the largest group in the 4-year medical school program, suggesting a possible recent increase in admissions or retention of first-year students. The distribution across semesters varies, with the highest percentage of students in Semester 1 and 4 (23.84%) and the lowest in Semester 8 (7.09%). Notably, there are no students in Semester 3, indicating possibly an error or specific academic structuring at the university. The mean age (years) is 20.22 (SD = 1.79, range = 12-31). The presence of students with a wide age range could indicate non-traditional pathways into the program. However, the median age, 20 (IQR = 19-21), reflected a more concentrated age group in the typical college-attending age range. There is an almost equal gender distribution with a slight female majority (51.55%). Additional key characteristics of our study sample presented in Table 1 include height (meters) and weight (kg) <sup>[10]</sup>.

### *Reliability*

Table 2 presents descriptive statistics and reliability analysis of the PHQ-9 Depression Scale for our study sample. Key metrics include the mean response scores ranging from “0” (not at all) to “3” (nearly every day), standard deviations (SD), Cronbach’s alpha for the full scale, and Cronbach’s alpha for each individual item if it was removed from the scale. The means of the items vary, indicating different levels of reported symptoms. The highest mean scores are associated with Item 4: “Feeling tired or having little energy” and Item 3: “Trouble falling or staying asleep or sleep too much” with means (SD) of 1.31 (0.79) and 1.05 (0.94), respectively, suggesting that these are more common symptoms among the students. Lower scores were on more severe symptoms, such as “Thoughts that you would be better off dead or of hurting yourself in some way”, indicating these are less commonly reported but are critical in understanding the severity of depressive symptoms. The PHQ-9 Depression Scale has an overall good internal consistency reliability across all 9 items, indicated by a Cronbach’s alpha of 0.8097, suggesting that the scale items reliably measure the same construct, which is depression in this case. All 9 items are retained in the final scale since Cronbach’s alpha would not improve much by dropping any items.

### *Factor Analysis*

Table 3 displays results of factor analysis for the PHQ-9 Depression Scale for our study sample. Factor 1 (“Negative Emotions/Thoughts”), representing 70.04% of the total variance, was selected according to the scree plot (Figure 1) and Kaiser-Guttman rule (eigenvalue = 3.1424 > 1.0000). None of the 9 items were excluded from the final scale since all factor loadings were above 0.3. The internal consistency reliability remains the same with a Cronbach’s alpha of 0.8097.

## *Validity*

Face validity was checked in this study as an informal way of content validity, with content validity checked by verifying that the scale items of the PHQ-9 Depression Scale align with existing studies. Construct validity was initially assessed using factor analysis with one factor (“Negative Emotions/Thoughts”) selected containing all 9 items, then confirmed through pairwise Pearson’s correlation between the PHQ-9 Depression Scale and other potential related factors such as the GAD-7 Anxiety Scale (see Table 4a) and the Epworth Sleepiness Scale (see Table 4b). A moderate to strong association with the GAD-7 ( $r = 0.6796$ ) and a moderate association with the Epworth Sleepiness Scale ( $r = 0.3884$ ), both with  $p$ -values  $< 0.0001$ , established consistency with existing studies on the associations between depression and anxiety disorders and daytime sleepiness, respectively.

## **Discussion**

### *Findings*

The PHQ-9 Depression Scale demonstrated good internal consistency reliability and validity among Mexican medical students. This scale is a 1-factor structure containing all 9 items with 70.04% of total variance that can be explained by the model. The PHQ-9 has good face and content validity because this 9-item scale can largely reflect the content domain of depression and has been validated by literature. Construct validity of the scale is also verified, highlighting that the scale has only one factor containing all 9 items and that there are significant associations between depression and anxiety/daytime sleepiness. Both methods established good construct validity.



In conclusion, the PHQ-9 is a highly standardized instrument and is used in diverse settings such as clinical and research environments, making it easy to administer and score for screening and early intervention in mental health disorders among medical students. It should be used in conjunction with clinical interviews and other diagnostic assessments to provide a full picture of depression diagnosis among medical students.

### *Limitations*

One of the main limitations of this study is that it may not be generalizable to other samples or populations since it only focused on a specific study sample. Another limitation is that as a self-report tool, the validity of the PHQ-9 can vary among different cultural and ethnic groups due to varying expressions of depression or depressive symptoms, causing recall and self-report biases, and thus resulting in overreporting or underreporting. Integrating the PHQ-9 for different cultural contexts such as administering the questionnaires in different languages and/or modifying the questions/items so that they are more applicable to medical students in other regions can reduce biases, ensure accuracy and inclusivity in depression diagnosis.

## References

1. World Health Organization. <https://icd.who.int/browse10/2016/en#/F30-F39>. Accessed December 12, 2018.
2. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001 Sep;16(9):606-13. doi: 10.1046/j.1525-1497.2001.016009606.x. PMID: 11556941; PMCID: PMC1495268.
3. Romo N., Bobadilla R., Tafoya S., Guízar D., Gutiérrez J., Carriedo P., Heinze G. Major depressive disorder in Mexican medical students and associated factors: a focus on current and past abuse experiences. *J. Affect. Disord*. 2019;245(15):834–840.
4. Mirza AA, Baig M, Beyari GM, Halawani MA, Mirza AA. Depression and Anxiety Among Medical Students: A Brief Overview. *Adv Med Educ Pract*. 2021 Apr 21;12:393-398. doi: 10.2147/AMEP.S302897. PMID: 33911913; PMCID: PMC8071692.
5. Mohareb AM, Brown TS. Medical Student Debt and the US Infectious Diseases Workforce. *Clin Infect Dis*. 2023 Apr 3;76(7):1322-1327. doi: 10.1093/cid/ciac862. PMID: 36318609; PMCID: PMC10396319.
6. Bannatyne AJ, Jones C, Craig BM, Jones D, Forrest K. A systematic review of mental health interventions to reduce self-stigma in medical students and doctors. *Front Med (Lausanne)*. 2023 Jun 15;10:1204274. doi: 10.3389/fmed.2023.1204274. PMID: 37396888; PMCID: PMC10311217.

7. The official website of the Epworth Sleepiness Scale (ESS & Ess-Chad). Epworth Sleepiness Scale. (n.d.). <https://epworthsleepinessscale.com/>
8. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011 Jun 27;2:53-55. doi: 10.5116/ijme.4dfb.8dfd. PMID: 28029643; PMCID: PMC4205511.
9. Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Annu Rev Clin Psychol*. 2009;5:1-25. doi: 10.1146/annurev.clinpsy.032408.153639. PMID: 19086835; PMCID: PMC2739261.
10. López T., Juan C., García, David. (2018). Mental health data of 776 Mexican medical students (PHQ-9, GAD-7 and Epworth Sleepiness Scale Scores). figshare. Dataset. <https://doi.org/10.6084/m9.figshare.1305817.v3>

Table 1. Key Characteristics of Mexican Medical Students at a Private Mexican University (N=776), 2018

	N (%)
Total	776 (100.00)
School Year in Course	
Year 1	275 (35.44)
Year 2	185 (23.84)
Year 3	140 (18.04)
Year 4	176 (22.68)
Semester in Course	
Semester 1	185 (23.84)
Semester 2	90 (11.60)
Semester 3	185 (23.84)
Semester 4	0 (0.00)
Semester 5	84 (10.82)
Semester 6	56 (7.22)
Semester 7	121 (15.59)
Semester 8	55 (7.09)
Age ( <i>years</i> )	
Mean (SD, range)	20.22 (1.79, 12-31)
Median (IQR)	20 (19-21)
Gender	
Male	375 (48.32)
Female	400 (51.55)
Missing	1 (0.13)
Height ( <i>meters</i> )	
Mean (SD, range)	1.69 (0.09, 1.42-1.93)
Median (IQR)	1.69 (1.62-1.75)
Weight ( <i>kg</i> )	
Mean (SD, range)	67.66 (13.72, 35-143)
Median (IQR)	67 (58-75)

SD = standard deviation

IQR = interquartile range

↓

Table 2. Descriptive Statistics and Reliability Analysis of the PHQ-9 Depression Scale Among Mexican Medical Students at a Private Mexican University (N=776), 2018

Scale Item (Cronbach's $\alpha$ = 0.8097, N = 776)	Mean <sup>a</sup>	SD	$\alpha$ if Item Deleted
Little interest or pleasure in doing things	0.94	0.71	0.7954
Feeling down, depressed, or hopeless	0.75	0.73	0.7816
Trouble falling or staying asleep, or sleeping too much	1.05	0.94	0.7936
Feeling tired or having little energy	1.31	0.79	0.7852
Poor appetite or overeating	1.04	0.91	0.7876
Feeling bad about yourself - or that you are a failure or have let yourself or your family down	0.78	0.84	0.7801
Trouble concentrating on things, such as reading the newspaper or watching television	0.69	0.8	0.7871
Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been	0.47	0.74	0.7944
Thoughts that you would be better off dead, or of hurting yourself in some way	0.11	0.4	0.8088

<sup>a</sup> Measured on a 9-item scale with response categories ranging from 0 = not at all to 3 = nearly every day  
SD = standard deviation

↓

Table 3. Factor Analysis of the PHQ-9 Depression Scale Among Mexican Medical Students at a Private Mexican University (N=776), 2018

Statement	Factor 1 Negative Emotions/Thoughts
Little interest or pleasure in doing things	0.5222
Feeling down, depressed, or hopeless	0.6914
Trouble falling or staying asleep, or sleeping too much	0.5682
Feeling tired or having little energy	0.6161
Poor appetite or overeating	0.6052
Feeling bad about yourself - or that you are a failure or have let yourself or your family down	0.7045
Trouble concentrating on things, such as reading the newspaper or watching television	0.5989
Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	0.5407
Thoughts that you would be better off dead, or of hurting yourself in some way	0.4191
Eigenvalue	3.1424
Cronbach's $\alpha$	0.8097

↓

Table 4a. Pairwise Pearson's Correlation Between the PHQ-9 Depression Scale and the GAD-7 Anxiety Scale Among Mexican Medical Students at a Private Mexican University (N=776), 2018

	PHQ-9 Depression Scale									PHQ-9 Total
	Little interest or pleasure in doing things	Feeling down, depressed, or hopeless	Trouble falling or staying asleep, or sleeping too much	Feeling tired or having little energy	Poor appetite or overeating	Feeling bad about yourself - or that you are a failure or have let yourself or your family down	Trouble concentrating on things, such as reading the newspaper or watching television	Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	Thoughts that you would be better off dead or of hurting yourself in some way	
<b>GAD-7 Anxiety Scale</b>										
Feeling nervous, anxious, or on edge	0.2623	0.4653	0.3186	0.4067	0.3090	0.4093	0.3347	0.3425	0.1919	
Not being able to stop or control worrying	0.2795	0.4858	0.3510	0.4307	0.3488	0.4803	0.3662	0.3630	0.2276	
Worrying too much about different things	0.2763	0.4340	0.3432	0.3923	0.2896	0.4183	0.3319	0.2932	0.2118	
Trouble relaxing	0.2642	0.4508	0.3413	0.4144	0.3080	0.4090	0.3638	0.3403	0.1864	
Being so restless that it is hard to sit still	0.2129	0.2822	0.2971	0.2958	0.3194	0.2812	0.3345	0.4562	0.1721	
Becoming easily annoyed or irritable	0.3239	0.3593	0.2775	0.3513	0.3171	0.4075	0.2935	0.3252	0.2009	
Feeling afraid, as if something awful might happen	0.2463	0.4765	0.2559	0.281	0.272	0.4019	0.3268	0.2875	0.2660	
<b>GAD-7 Total</b>										0.6796 (p < 0.0001)

↓

Table 4b. Pairwise Pearson's Correlation Between the PHQ-9 Depression Scale and the Epworth Sleepiness Scale Among Mexican Medical Students at a Private Mexican University (N=776), 2018

	PHQ-9 Depression Scale									PHQ-9 Total
	Little interest or pleasure in doing things	Feeling down, depressed, or hopeless	Trouble falling or staying asleep, or sleeping too much	Feeling tired or having little energy	Poor appetite or overeating	Feeling bad about yourself - or that you are a failure or have let yourself or your family down	Trouble concentrating on things, such as reading the newspaper or watching television	Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	Thoughts that you would be better off dead or of hurting yourself in some way	
Epworth Sleepiness Scale										
Sitting and reading	0.1898	0.2298	0.2144	0.2396	0.2053	0.2571	0.2334	0.1786	0.1779	
Watching TV	0.1136	0.1569	0.1191	0.1808	0.1012	0.1483	0.1581	0.1522	0.1443	
Sitting inactive in a public place (e.g., a theater or a meeting)	0.1235	0.1494	0.1212	0.1079	0.0832	0.1754	0.1686	0.1599	0.2200	
As a passenger in a car for an hour without a break	0.0739	0.1367	0.1093	0.1804	0.1322	0.1038	0.0521	0.1260	0.1056	
Lying down to rest in the afternoon when circumstances permit	0.1187	0.0738	0.1681	0.2322	0.1445	0.0780	0.1195	0.1235	-0.0136	
Sitting and talking to someone	0.1020	0.1297	0.1309	0.1020	0.0701	0.1538	0.1731	0.1443	0.1646	
Sitting quietly after a lunch without alcohol	0.1242	0.1620	0.1648	0.1928	0.1681	0.1469	0.2114	0.1498	0.1123	
In a car, while stopped for a few minutes in traffic	0.1324	0.2023	0.1320	0.2124	0.1934	0.1885	0.1688	0.1827	0.1770	
Epworth Sleepiness Scale Total										0.3884 (p < 0.0001)

↓



Figure 1. Scree Plot of Eigenvalues from Factor Analysis of the PHQ-9 Depression Scale Among Mexican Medical Students at a Private Mexican University (N=776), 2018

