



Evidencias psicométricas de la versión peruana del Patient Health Questionnaire (PHQ-9)

Psychometric evidence of the Peruvian version of the Patient Health Questionnaire, PHQ-9

Kennedy Quispilcaya Capcha¹ , Giuliana Salazar-Álvarez² ,
Elizabeth Dany Araujo-Robles³ 

Resumen

El objetivo del presente estudio fue analizar las propiedades psicométricas de la versión peruana del Patient Health Questionnaire (PHQ-9) en una muestra de 33,613 personas, con datos obtenidos de una base de datos de la ENDES 2019. La metodología es de nivel básico no experimental y con diseño instrumental. Los resultados indican buena calidad de los ítems, evidenciando capacidad para medir el mismo constructo y diferenciar a las personas con distintos niveles de depresión. Se identificó un modelo factorial de dos dimensiones (somático y cognitivo) con una adecuada bondad de ajuste (RMSEA = .097; CFI = .932; NFI = .932; TLI = .906). Finalmente, se obtuvo un alto nivel de confiabilidad ($\omega = .892$).

Palabras clave: depresión, PHQ-9, propiedades psicométricas, población peruana, tamizaje

Abstract

The objective of this study was to analyze the psychometric properties of the Peruvian version of the Patient Health Questionnaire (PHQ-9), with a sample of 33,613 people, obtained from a database of the ENDES 2019. The methodology is basic level, not experimental and instrumental design. The results indicate good quality of the items, showing the ability to measure the same construct and differentiate people with distinct levels of depression. A two-dimensional factor (somatic and cognitive) model, with adequate goodness of fit was identified (RMSEA = .097; CFI = .932; NFI = .932; TLI = .906). Finally, a high level of reliability was obtained ($\omega = .892$).

Keywords: depression, PHQ-9, psychometric properties, Peruvian population, screening

INTRODUCCIÓN

La depresión es uno de los trastornos psiquiátricos que se presenta con mayor frecuencia en las personas y es considerado como la principal causa de discapacidad (Villarreal-Zegarra et al., 2019). Se estima que el 4.4% de la población mundial sufren con esta enfermedad, considerando con mayor vulnerabilidad a las mujeres (3.6% de los hombres y 5.1% de las mujeres) (OMS, 2017), además, la sintomatología depresiva en América

¹ Universidad Internacional de Valencia, Valencia, España.

² Asociación Peruana Contextual Conductual de Psicología, Lima, Perú.

³ Universidad Peruana Cayetano Heredia, Lima, Perú.

del Norte, América Latina y el Caribe se calculó en 15% (OMS, 2017). Por ejemplo, en el Perú, durante la pandemia en el 2021, fueron atendidos 313,455 casos a través de diferentes servicios de salud mental, lo cual reflejó un aumento de 12% en comparación a la etapa de prepandemia (MINSA, 2022).

Teniendo en cuenta que la depresión está en camino a ser la principal causa de años perdidos por discapacidad para el 2030 y considerando su alta frecuencia, es necesario contar con instrumentos psicométricos actualizados en el contexto peruano para su evaluación y diagnóstico temprano (Vega, 2018).

Los instrumentos de versión corta más utilizados para evaluar la depresión son la CES-D, la Escala de Depresión de Hamilton, la EDZ, el Inventario de Depresión de Beck, el EPDS y el Cuestionario de Salud del Paciente, PHQ-9 (Prieto-Molinaria et al., 2020). El Patient Health Questionnaire 9 (PHQ-9) es un cuestionario que evalúa la depresión de acuerdo a los criterios del DSM-IV. Sin embargo, aún existen debates respecto a su estructura factorial.

En los últimos años se reportaron investigaciones con posturas contrarias sobre la estructura factorial (modelo unidimensional, bidimensional y bifactorial) del cuestionario PHQ-9. Un modelo unifactorial que pertenece al análisis de primer orden; es decir, trabaja con un conjunto de ítems que pertenecen a un solo factor latente o constructo. Por lo tanto, un instrumento psicométrico se considerará unidimensional cuando las respuestas están basadas en un único atributo (Ferrando & Anguiano, 2010).

Asimismo, otro modelo propuesto es el modelo bidimensional, el cual explica directamente la variación de los ítems con base en dos dimensiones; pero, al igual que el anterior, también es considerado un análisis de primer orden. No obstante, cada dimensión debe medirse como una combinación formativa; es decir con puntajes directos obtenidos de la suma por ambos factores, pero cada dimensión también puede ser medida por separado. Además, los modelos bidimensionales tienen menos puntajes de error y logran una medición más precisa. Por otra parte, el modelo bifactorial es un modelo de segundo orden, en el que las dimensiones se explican mejor por la presencia de un factor general.

Respecto al discutido modelo factorial del PHQ-9, González-Blanch et al. (2018), Villarreal-Zegarra et al. (2019) y Zhong et al. (2014) reportaron

evidencias psicométricas indicando que los datos del instrumento se ajustan y se ven mejor representados mediante un modelo unidimensional. Sin embargo, De la Flor (2020), Merz et al. (2011), Patel et al. (2020) y Saldivia et al. (2019), respaldan la estructura de dos dimensiones, indicando que el PHQ-9 bidimensional es el más apropiado para evaluar la depresión. Por otro lado, hasta la fecha el único estudio que reportó un buen ajuste de los datos para un modelo bifactorial fue el de Huarcaya-Victoria et al. (2020).

Otras evidencias psicométricas del PHQ-9 se encuentran en los estudios de Zhong et al. (2014), quienes hallaron correlaciones ítem-test ($r = .3$) y obtuvieron una adecuada consistencia interna ($\text{PSI} = .972$). Por su parte, Villarreal-Zegarra et al. (2019), también identificaron un adecuado nivel de confiabilidad ($\omega = .87$). Conviene subrayar que no se encontraron estudios sobre el análisis de los ítems.

En el presente estudio se realizará el análisis factorial acompañado con los índices de discriminación, homogeneidad y comunalidades, así como los coeficientes de consistencia interna. La revisión de las evidencias psicométricas de este instrumento para tamizaje de depresión beneficiará en la atención primaria, a través de pruebas de tamizaje; mediante el empleo del PHQ-9 en el primer nivel de atención para reducir el tiempo en las evaluaciones. De esa manera, se puede evaluar a más personas en poco tiempo, esto considerando que las consultas en el primer nivel de atención son de corta duración y con alta demanda de pacientes.

MÉTODO

Participantes

Se recurrió a un análisis de datos secundarios, a partir de la Encuesta de Demografía y Salud del Perú (ENDES), una encuesta representativa que se realiza anualmente a nivel nacional. El diseño de la ENDES incluye una técnica de muestreo caracterizada por ser probabilística de tipo equilibrado, bietápica, estratificada e independiente. La muestra de estudio es representativa en función de la población peruana, conformada por hombres y mujeres, de entornos urbanos y rurales, de origen multiétnico y de diferentes niveles socioeconómicos. En el presente estudio se incluyeron datos de personas desde los 15 años de edad o más; sin embargo, se excluyó a los participantes con datos incompletos en las

variables de interés (sexo, edad y los nueve ítems del PHQ). Finalmente, se analizaron los datos de 33,613 participantes, 14,301 varones y 19,312 mujeres.

Instrumento

El PHQ-9 fue originalmente desarrollado por Spitzer et al. (1999). Puede ser aplicado individual o colectivamente a adultos de 15 años o más, con un tiempo de aplicación aproximado de 5 minutos. El PHQ-9 evalúa la depresión de acuerdo a los criterios del DSM-IV mediante 9 ítems con un formato de respuesta tipo Likert que incluye alternativas desde 0 (*nunca*) hasta 3 (*casi todos los días*) y que indica la severidad de la sintomatología depresiva, proporcionando una puntuación total cuantitativa de intervalo del 0 hasta 27 (Kroenke et al., 2001; Kroenke & Spitzer, 2002).

Procedimiento y análisis de datos

El presente estudio psicométrico se desarrolló con fuente de datos secundarias de acceso libre; por consiguiente, no se aplicó el instrumento directamente. Inicialmente, se realizó la búsqueda de la fuente de datos. Una vez obtenidos los datos, se evaluó la normalidad de las puntuaciones del PHQ-9 mediante la prueba de Kolmogorov-Smirnov. Asimismo, se analizó la calidad de los ítems, mediante los estadísticos elemento total corregido, communalidades y *t* de Student para los índices de homogeneidad, communalidades y discriminación, respectivamente.

A continuación, se realizaron las pruebas KMO y de esfericidad de Bartlett ($p < .05$) y, luego, se realizaron análisis factoriales confirmatorios (AFC) para evaluar los modelos unidimensionales, bidimensionales y bifactor mediante el método de extracción de máxima verosimilitud. Los estadísticos que se tomaron en cuenta para la evaluación de cada modelo fueron las medidas de ajuste absoluto (Chi cuadrado y RMSEA), las medidas de ajuste incremental (TLI, CFI y NFI) y las medidas de ajuste de la parsimonia (PRATIO, PCFI, PNFI y AIC). La confiabilidad se evaluó mediante el método de consistencia interna, para ello se calculó el estadístico Omega.

RESULTADOS

Los 9 reactivos presentan un nivel significativo en el índice de discriminación ($p < .05$). Además, los datos de la Tabla 1 indican valores adecuados del índice de homogeneidad, mediante el índice del

elemento total corregido ($> .30$). Estos hallazgos indican que todos los reactivos miden la misma variable (depresión) de manera selectiva y cuentan con capacidad para discernir entre los niveles leve, moderado y severo de depresión. Sin embargo, las communalidades indican que los reactivos no van a explicar adecuadamente el factor ($> .50$), lo que supone un desajuste en el modelo factorial, el cual se abordará más adelante. Se aprecia que los ítems 5, 8 y 9 explican la varianza entre 36% y 47.7%, por lo que tendrían que eliminarse; sin embargo, cuentan con un nivel adecuado de homogeneidad.

Tabla 1
Índices para el análisis de la calidad de reactivos

	IHC	h^2
Ítem 1: Anhedonia	.656	.542
Ítem 2: Distimia	.715	.623
Ítem 3: Sueño	.646	.533
Ítem 4: Cansancio	.694	.597
Ítem 5: Apetito	.601	.477
Ítem 6: Desconcentración	.635	.527
Ítem 7: Psicomotricidad	.627	.519
Ítem 8: Ideación suicida	.500	.360
Ítem 9: Culpa	.558	.436

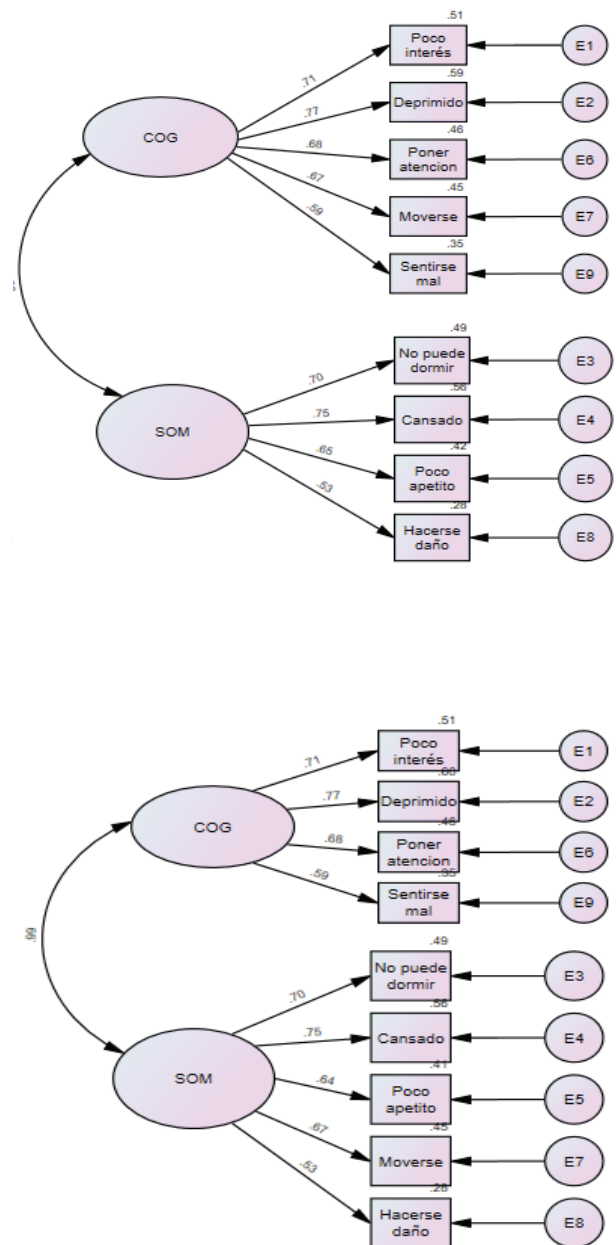
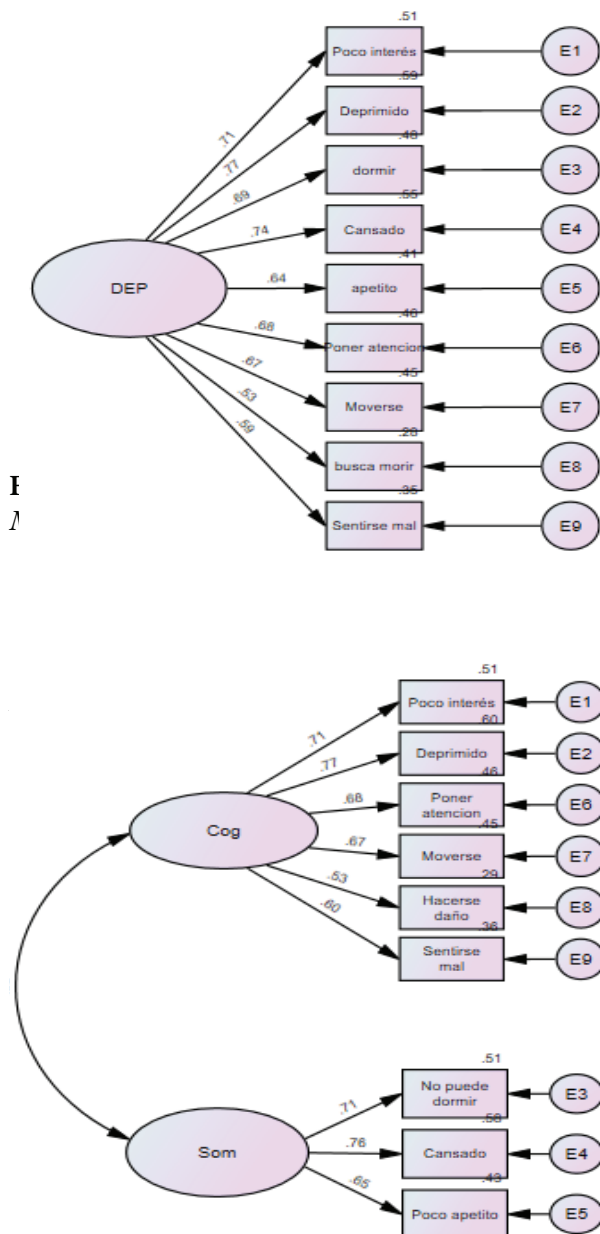
Nota. IHC = índice de homogeneidad corregido;
 h^2 = communalidades.

Los valores de la medida de adecuación muestral (KMO) y de la prueba de esfericidad de Barlett (0.914 y $p < .001$, respectivamente), indicaron la viabilidad del análisis factorial (Montoya, 2007). La prueba estadística exacta chi-cuadrada χ^2 indicaría que los datos no se ajustan a ninguno de los 5 modelos evaluados ($p < .001$). En la Tabla 2 se puede observar, además, los restantes estadísticos de bondad de ajuste considerados. El valor del RMSEA para los modelos 1, 2, 3 y 4 es 0.098, 0.097, 1.0 y 1.0, respectivamente; lo que significa que las medidas de ajuste absolutos presentan un nivel mediocre; no obstante, en el modelo 5 se obtuvo un valor del RMSEA de 0.077, el cual indicaría un mejor ajuste absoluto, sin embargo, no es completamente adecuado. Por su parte, los indicadores de ajuste incremental para el modelo 1 evidencian un nivel adecuado (CFI=.929; NFI=.928; TLI=.905), así como para el modelo 2 (CFI=.932; NFI=.932; TLI=.906), el modelo 3 (CFI=.929; NFI=.929; TLI=.902) y el modelo 4 (CFI=.929;

NFI=.928; TLI=.901), mientras que el modelo 5 presenta niveles altos (CFI=.971; NFI=.971; TLI=.942). Asimismo, frente a los modelos 1, 2, 3 y 4, el modelo 5 tiene medidas de ajuste de la parsimonia menor, lo que estaría indicando un mejor ajuste de los datos. Sin embargo, sus cargas factoriales son negativas y muy bajas. Las Figuras 1 a 5 muestran gráficamente los distintos modelos puestos a prueba.

Figura 1

Modelo 1 del PHQ-9: unidimensional



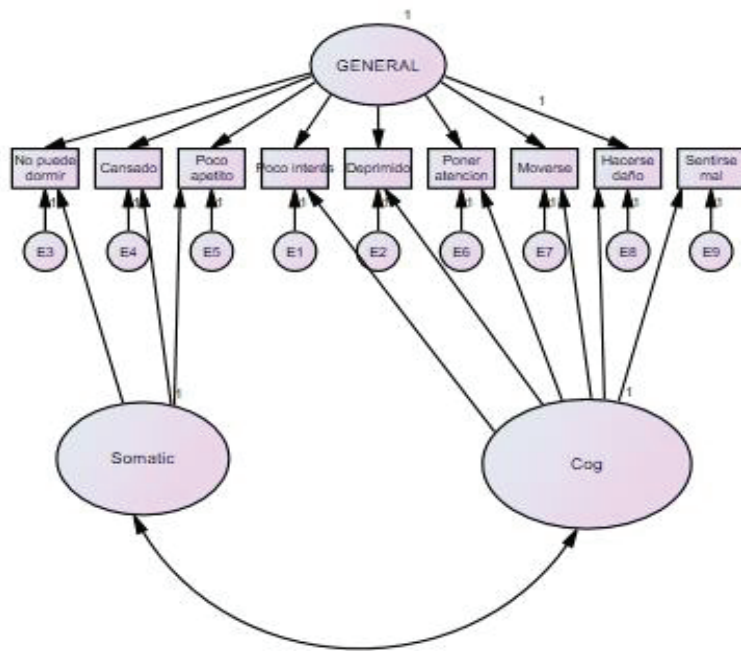


Tabla 2
Índices de bondad de ajustes de los diferentes modelos evaluados

Modelos	RMSEA	CFI	TLI	NFI	PRATIO	PCFI	PNFI	AIC
1	0.098	0.929	0.905	0.928	0.750	0.696	0.696	8796.712
2	0.097	0.932	0.906	0.932	0.722	0.673	0.673	8367.711
3	1.00	0.929	0.902	0.929	0.722	0.672	0.671	8736.132
4	1.00	0.929	0.901	0.928	0.722	0.671	0.671	8762.992
5	0.077	0.971	0.942	0.971	0.528	0.512	0.512	3719.817

Para la presente investigación se considera como más adecuado el modelo 2 (bifactorial), debido a que obtuvo, ligeramente, mejores puntajes en los índices de bondad de ajuste; además teóricamente responde a la literatura y facilita el monitoreo de las personas con síntomas depresivos. Cabe resaltar que los puntajes directos para la evaluación se obtendrán de la suma de ambas dimensiones (somática y cognitiva).

Finalmente, se evaluó la confiabilidad de consistencia interna y confiabilidad según el método de dos mitades con un total de 9 ítems, con resultados adecuados mediante los estadísticos Omega ($\omega = .892$), Spearman-Brown (.824) y Guttman (.753). Por lo que se infiere buena fiabilidad del modelo y buena correlación entre variables.

DISCUSIÓN

A partir de los hallazgos encontrados, se evidenció buena calidad de los ítems de la versión peruana del PHQ-9. Esta valoración responde a criterios establecidos en la literatura (Domínguez, 2013; Salazar & Del Castillo, 2018). De este modo, el índice de homogeneidad cumplió los criterios estadísticos de Kline (1999), quien indica que los puntajes mayores a .30 son adecuados para evaluar la homogeneidad. Este resultado refleja que los puntajes se adecuan perfectamente y que todos los ítems contribuyen para medir el mismo constructo (depresión).

Por otra parte, los ítems que explican la proporción de la varianza, son representados por las comunales. Los resultados obtenidos son aceptables y superan el mínimo requerido 0.5 (Hair et al., 2013). No obstante, como se mencionó, los ítems 5, 8 y 9 explican la varianza entre 36% y 47.7% por lo que tendrían que eliminarse. Si bien es cierto que cuentan con un gran porcentaje de varianza de error y no estarían bien representados en el modelo, lo cierto es que dichos ítems cuentan con adecuada homogeneidad. Además, las comunales en la psicología generalmente son bajas, ya que las medidas como los tests y sobre todo los ítems, contienen intrínsecamente muchos errores de medida (Ferrando & Anguiano, 2010). Por consiguiente, se conservan los 9 ítems.

Dentro de este orden de ideas, la capacidad discriminativa fue significativa ($p < .05$) en todos los ítems. En otras palabras, todos los ítems pueden diferenciar a las personas con niveles leves,

moderados y severos de depresión (Molina et al., 2020). Este resultado resalta la calidad de los ítems, ya que una característica deseable para un reactivo y sobre todo para aumentar su calidad es una alta capacidad de discriminación (Hurtado, 2018).

Aunque se recomienda analizar el índice de dificultad de los ítems, este no es recomendable para pruebas clínicas, sino para instrumentos que miden el rendimiento, sobre todo en el campo educativo (Pérez-Sala et al., 2019). Como se pudo observar, existe un buen nivel de calidad de los ítems del PHQ-9 para el diagnóstico de depresión.

También se ha verificado la validez relacionada al constructo, mediante el AFC. Dentro de este marco, para procesar los datos se consideraron 5 modelos diseñados y re-diseñados por Villarreal-Zegarra et al. (2019) y por Huarcaya-Victoria et al. (2020). Es preciso señalar que no se modificó ningún modelo y que solo se tomó las estructuras para procesar los datos y elegir el modelo más apropiado para el instrumento.

El Chi cuadrado indica desajuste en todos los modelos sometidos a prueba; no obstante, este resultado es sensible al tamaño de muestra que se considera para un estudio (Rose et al., 2017), además este estadístico no aporta la información necesaria para valorar el modelo, ya que no emplea un modelo alternativo para comparar, sino que recurre a la matriz reproducida por el método de estimación (Medrano & Muñoz-Navarro, 2017). Como consecuencia y al ser visible la cantidad de muestra que se empleó en este estudio, no se considera como determinante los resultados del estadístico Chi cuadrado.

Desde una perspectiva más específica, en el modelo 1 (unidimensional) se obtuvo un índice de error cuadrático mediocre, pero los índices de ajuste incremental obtuvieron valores apropiados. Investigadores como Villarreal-Zegarra et al. (2019) y González-Blanch, et al. (2018) apoyan la idea de un modelo unidimensional del PHQ-9, debido a que existe una alta correlación entre las dimensiones somática y cognitiva, esto con la finalidad de garantizar una simplicidad en los modelos.

Por otra parte, tenemos los modelos bidimensionales. En este estudio el modelo 2 es considerado el más apropiado, además de ser elegido como el modelo que representa al instrumento. Obtuvo un índice de error cuadrático mediocre (Tabla 2), sin embargo,

obtuvo valores apropiados en los índices de ajuste incremental y mejores medidas de ajuste de la parsimonia (Tabla 2) (Escobedo et al., 2016), esto en comparación con las otras estructuras factoriales procesadas. Los resultados del modelo 2 (Tabla 2) concuerdan con los estudios de Patel et al. (2019) quienes emplearon el AFCo con una muestra de 31,366 personas y respaldaron un modelo de dos dimensiones. Además, autores como Merz et al. (2011) y Saldivia et al. (2019) confirmaron una estructura de dos dimensiones del PHQ-9 para las poblaciones latinoamericanas.

Siguiendo con los modelos bidimensionales, se halló el RMSEA de los modelos 3 y 4, el cual es considerado como un nivel muy pobre, no obstante, obtuvieron puntajes adecuados en los índices de ajuste incremental (Tabla 2). Como se puede observar, en ambos modelos el índice de error cuadrático es muy pobre, excepto en los ajustes incrementales. Por lo tanto, los datos no se adecuan a dichos modelos y, en consecuencia, se descarta su utilidad para el instrumento.

Por tal motivo, el modelo 2 (bidimensional) es elegido como el modelo que explica los datos del instrumento en la presente investigación, para ello, se consideró 3 aspectos. Primero, obtuvo mejores niveles en la bondad de ajuste en comparación con los otros modelos; segundo, teóricamente responde a la literatura y está enfocado en la población clínica y tercero, facilitará el monitoreo de las personas con síntomas depresivos.

La interpretación de estos resultados fue complicada puesto que los modelos unidimensional y bidimensional mantienen resultados similares y ambos son respaldados por diferentes investigaciones. Además, en este estudio no se encuentran diferencias relevantes entre los valores de los índices de bondad de ajuste. Sin embargo, la diferencia podría estar relacionada con las características de la población; según Villarreal-Zegarra et al. (2019), los modelos bidimensionales están basados generalmente en poblaciones clínicas, mientras que los modelos unidimensionales, que reportan superposición de sus dimensiones, están basados en la población general.

Asimismo, los factores culturales sobre la percepción de los síntomas de la depresión podrían estar influenciando en los resultados (Escovar et al., 2018). Hay que destacar que, teóricamente, la depresión es

un constructo multidimensional, donde el ánimo deprimido y los sentimientos de inutilidad son los reactivos más importantes (Huarcaya-Victoria et al. 2020).

Finalmente, dentro de la validez relacionada al constructo, se procesó el modelo 5 (bifactorial), el cual obtuvo mejores puntajes en la bondad de ajuste (Tabla 2) incluso mejores medidas de ajuste de la parsimonia. Sin embargo, un hallazgo inesperado fue que las cargas factoriales se encontraron por debajo del mínimo requerido 0.5 (De la Fuente, 2011) y algunos ítems cargaron en negativo, excepto en el factor somático, donde sí se obtuvo cargas factoriales aceptables. En otras palabras, los resultados no son consistentes y existen muchos errores de medida en los ítems.

Es importante señalar que los modelos bifactoriales pueden llevar a obtener resultados falsos positivos si solo se consideran los índices de ajustes, es más ya existen evidencias que los índices tradicionales suelen favorecer a los modelos bifactor (Gignac, 2016; Morgan et al., 2015). Sería fructífero analizar otros índices, pero debido a limitaciones instrumentales y metodológicas, esta investigación no puede proporcionar una revisión exhaustiva de los índices que permiten valorar la robustez del modelo bifactor.

También se evaluó el estadístico de Omega obteniendo un valor de .892, este resultado según Campo-Arias y Oviedo (2008) se considera óptimo. Asimismo, este resultado concuerda con el del estudio de Villarreal-Zegarra et al. (2019), quienes trabajaron con una muestra representativa de 30,446 personas peruanas y como resultado obtuvieron un coeficiente Omega de .87 de la prueba de PHQ-9. Los mismos autores recomiendan el coeficiente Omega para investigaciones de corte psicométrico, porque permite tener medidas más precisas de la confiabilidad. Además, a diferencia del coeficiente Alfa, el estadístico Omega no depende de la cantidad de ítems y permite tener cálculos más estables, ya que solo trabaja con las cargas factoriales (Ventura-León & Caycho-Rodríguez, 2017).

La confiabilidad se ve respaldada también con otros valores, para ello se evaluó los coeficientes de Spearman-Brown y Guttman, obteniéndose .824 y .753, respectivamente, estos valores son considerados como muy buenos. No obstante, la literatura científica desarrollada hasta el momento no muestra

evidencias de la utilización de estos estadísticos en el análisis psicométrico del PHQ-9

En síntesis, el PHQ-9 evidencia buena calidad de los ítems y el modelo bidimensional obtenido muestra un buen ajuste a los datos. La primera dimensión mide los síntomas somáticos y la segunda, los síntomas cognitivos. Evidencia también una alta confiabilidad. Se trata, por tanto, de un instrumento válido y confiable para la evaluación de depresión en la población peruana y se recomienda su uso para tales fines, ya sea para investigaciones epidemiológicas o como instrumento de screening.

Correspondencia:

Kennedy Quispilcaya Capcha

Email: kennedy.luigui97@gmail.com

REFERENCIAS

- Campo-Arias, A., & Oviedo, C. (2008). Propiedades psicométricas de una escala: la consistencia interna. *Revista de Salud Pública*, 10(5), 831-839. <https://doi.org/10.1590/S0124-00642008000500015>
- De la Flor, M. (2020). *Validación de la versión peruana del cuestionario sobre la salud del paciente (PHQ-9) para tamizaje de depresión en pacientes adultos con cáncer de un hospital público especializado de Lima Metropolitana* [Tesis de Licenciatura en Psicología]. Universidad Peruana Cayetano Heredia. https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/9023/Validacion_DelaFlorValdivia_Martin.pdf?sequence=1&isAllowed=y
- Domínguez, S. (2013). ¿Ítems politómicos o dicotómicos? Un estudio empírico con una escala unidimensional. *Revista Argentina de Ciencias del Comportamiento*, 5(3), 30-37. <https://doi.org/10.32348/1852.4206.v5.n3.5562>
- Escobedo, M., Hernández, J., Estebané, V., & Martínez, G. (2016). Modelos de ecuaciones estructurales: características, fases, construcción, aplicación y resultados. *Ciencia y Trabajo*, 18(55), 16-22. <http://dx.doi.org/10.4067/S0718-24492016000100004>
- Escovar, E., Craske, M., Roy-Byrne, P., Stein, M., Sullivan, G., Sherbourne, C., Bystritsky, A., & Chavira, D. (2018). Cultural influences on mental health symptoms in a primary care sample of Latinx patients. *Journal of Anxiety Disorders*, 55, 39-47. <https://doi.org/10.1016/j.janxdis.2018.03.005>
- Ferrando, P., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33. <https://www.redalyc.org/articulo.oa?id=77812441003>
- Gignac, G. (2016). The higher-order model imposes a proportionality constraint: that is why the bifactor model tends to fit better. *Intelligence*, 55, 57-68. <https://doi.org/10.1016/j.intell.2016.01.006>
- González-Blanch, C., Medrano, L., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J., Limonero, J., Schmitz, F., Cano-Vindel, A., & PsicAP Research Group (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(2), e0193356. <https://doi.org/10.1371/journal.pone.0193356>
- Hair, J., Black, W., Babin, B. J., & Anderson, R. E., (2013). *Multivariate data analysis* (7th. Ed.). Pearson.
- Huarcaya-Victoria, J., De Lama-Morán, R., Quirós, M., Bazán, J., López, K., & Lora, D. (2020). Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú. *Revista de Neuro-Psiquiatría*, 83(2), 72-78. <http://dx.doi.org/10.20453/rnp.v83i2.3749>
- Hurtado, L. (2018). Relación entre los índices de dificultad y discriminación. *Revista Digital de Investigación en Docencia Universitaria*, 12(1), 273-300. <https://dx.doi.org/10.19083/ridu.12.614>
- Kline, P. (1999). *Handbook of psychological testing*. Routledge.
- Kroenke, K., & Spitzer, R. L. (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, 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Medrano, L., & Muñoz-Navarro, R. (2017). Aproximación conceptual y práctica a los modelos de ecuaciones estructurales. *Revista Digital de Investigación en Docencia Universitaria*, 11(1), 219-239. <https://dx.doi.org/10.19083/ridu.11.486>
- Merz, E., Malcarne, V., Roesch, S., Riley, N., & Sadler, G. (2011). 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>
- Ministerio de Salud (MINSA). (2022). Más de 300 mil casos de depresión fueron atendidos durante el 2021. <https://www.gob.pe/institucion/minsa/noticias/575899-minsa-mas-de-300-mil-casos-de-depresion-fueron-atendidos-durante-el-2021/>
- Molina, M., Ochoa, C., & Ortega, E. (2020). Comparación de dos medias. Prueba de la T de Student. *Evidencias en Pediatría*, 16, 51. https://evidenciasenpediatria.es/files/41-13858-RUTA/51_Fundamentos_ComparacionMedias.pdf
- Montoya, O. (2007). Aplicación del análisis factorial a la investigación de mercados. caso de estudio. *Scientia et Technica*, Año XIII, N. 35, 281-286. <https://moodle2.utp.edu.co/index.php/revistaciencia/article/view/5443>
- Morgan, B., de Bruin, G., & de Bruin, K. (2014). Operationalizing burnout in the Maslach Burnout Inventory – Student Survey: personal efficacy versus personal inefficacy. *South African Journal of Psychology*, 44(2), 216-227. <https://doi.org/10.1177/0081246314528834>
- Organización Panamericana de la Salud y Organización Mundial de la Salud (2017). *Depresión y otros trastornos mentales comunes: estimaciones sanitarias mundiales*. <https://iris.paho.org/bitstream/handle/10665.2/34006/PAHONMH17005-spa.pdf?sequence=1&isAllowed=y>
- Patel, J., Oh, Y., Rand, K., Wu, W., Cyders, M., Kroenke, K., & Stewart, J. (2019). Measurement invariance of the patient health questionnaire-9 (PHQ-9) depression screener in U.S. adults across sex, race/ethnicity, and education level: NHANES 2005-2016. *Depression and Anxiety*, 36(9), 813-823. <https://doi.org/10.1002/da.22940>
- Pérez-Sala, C., Sirlopú, D., Cobo, R., & Awad, A. (2019). Análisis bifactorial de la escala de participación escolar en una muestra de estudiantes chilenos. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica. RIDEP*, 3(52), 27-39. <https://doi.org/10.21865/RIDEP52.3.03>
- Prieto-Molinaria, D., Aguirre, G., de Pierola, I., Luna, G., Merea, L., Lazarte, C., Uribe-Bravo, K., & Zegarra, Á. (2020). Depresión y ansiedad durante el aislamiento obligatorio por el COVID-19 en Lima Metropolitana. *Liberabit*, 26(2), e425. <https://dx.doi.org/10.24265/liberabit.2020.v26n2.09>
- Rose, A., Markman, B., & Sawilowsky, S. (2017). Limitations in the systematic analysis of structural equation model fit indices. *Journal of Modern Applied Statistical Methods*, 16(1), 69-85. <https://doi.org/10.22237/jmasm/1493597040>
- Salazar, C., & Del Castillo, S. (2018). *Fundamentos básicos de estadística* (1ra. Ed.). <http://librodigital.sangregorio.edu.ec/librosusgp/B0009.pdf>
- Saldivia, S., Aslan, J., Cova, F., Vicente, B., Inostroza, C., & Rincón, P. (2019). Psychometric characteristics of the Patient Health Questionnaire (PHQ-9). *Revista Médica de Chile*, 147(1), 53-60. <http://dx.doi.org/10.4067/S0034-98872019000100053>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & The Patient Health Questionnaire Primary Care Study Group. (1999). Validation and utility of a self-report version of PRIME-MD The PHQ Primary Care Study. *JAMA*, 282(18), 1737-1744. <https://doi.org/10.1001/jama.282.18.1737>
- Vega-Dienstmaier, J. (2018). Construcción de versiones cortas de la escala de psicopatología depresiva para tamizaje de depresión mayor y sus características psicométricas. *Revista de Neuropsiquiatría*, 81(3), 154-164. <https://doi.org/10.20453/rnp.v81i3.3383>

- Ventura J., & Caycho, T. (2017). El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627. <http://revistalatinoamericanaumanizales.cinde.org.co/docs/Abril2021/Cartas%20al%20Editor.%20El%20coeficiente%20Omega.pdf>
- Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabe-Ortiz, A., Melendez-Torres G., & Bazo-Alvarez, J. (2019). Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics. *PLoS ONE* 14(9), e0221717. <https://doi.org/10.1371/journal.pone.0221717>
- Zhong, Q., Gelaye, B., Fann, J., Sanchez, S., & Williams, M. (2014). Cross-cultural Validity of the Spanish version of PHQ-9 among pregnant Peruvian women: A Rasch item response theory analysis. *Journal of Affective Disorders*, 10(158), 148–153. <https://doi.org/10.1016/j.jad.2014.02.012>