

# Cuestionario de Salud del Paciente-9 (PHQ-9): una revisión sistemática y un metaanálisis de la generalización de la confiabilidad

**Patient Health Questionnaire (PHQ-9): a systematic review and a meta-analysis of reliability generalization**

Mafalda Ortiz Morán<sup>1,a</sup> , Jeanette Vásquez Vega<sup>1,a</sup> ,  
Iris Gladys Correa Aranguren<sup>1,a</sup> , Javier Andrés Reyes Rodríguez<sup>1,b</sup> ,  
Emerson Walter Laura Barraza<sup>1,c</sup> , José Livia Segovia<sup>1,a</sup> 

## RESUMEN

La depresión es un problema relevante de salud pública, razón por la cual es necesario conocer los aspectos psicométricos de los instrumentos utilizados en su evaluación, a fin de posibilitar mejores decisiones. El Cuestionario de Salud del Paciente (PHQ-9) es una herramienta que satisface notablemente el nivel de exactitud en la medida de depresión, por lo que el objetivo del presente estudio fue evaluar su fiabilidad mediante un metaanálisis de generalización. Una búsqueda sistemática en Web of Science y Scopus, entre el 2020 y el 2022, permitió evaluar 32 trabajos que utilizaron el PHQ-9 y reportaban el alfa de Cronbach. Se realizó un metaanálisis de efectos aleatorios para estimar la fiabilidad de la prueba, con estimación de la heterogeneidad mediante los abordajes estadísticos Q de Cochran e I<sup>2</sup>. Posibles variables moderadoras fueron analizadas mediante metarregresión y análisis de subgrupos. La fiabilidad combinada del PHQ-9 es de 0,846 (IC 95 %: 0,824-0,862); el test de Egger y el test de correlación ordinal  $\tau$  de Kendall para sesgo de publicación permitieron documentar la ausencia de evidencia significativa de tal sesgo en los datos evaluados. Se determinó también una heterogeneidad estadísticamente significativa en la muestra total, con base en la prueba Q y el índice I<sup>2</sup> de variabilidad. El análisis de metarregresión señaló que el tipo de muestra (clínico-general) es un moderador estadísticamente significativo. El presente

### Citar como:

Ortiz M, Vásquez J, Correa IG, Reyes JA, Laura EW, Livia J. Cuestionario de Salud del Paciente-9 (PHQ-9): una revisión sistemática y un metaanálisis de la generalización de la confiabilidad. *Rev Neuropsiquiatr.* 2024; 87(3): 273-288. DOI: 10.20453/rnp.v87i3.5264

**Recibido:** 22-02-2024

**Aceptado:** 17-07-2024

**En línea:** 30-09-2024

### Correspondencia:

Iris Gladys Correa Aranguren  
Contacto: [icorrea@unfv.edu.pe](mailto:icorrea@unfv.edu.pe)



Artículo de acceso abierto, distribuido bajo los términos de la Licencia Creative Commons Atribución 4.0 Internacional.

© Los autores

© *Revista de Neuro-Psiquiatría*

<sup>1</sup> Universidad Nacional Federico Villarreal. Lima, Perú.

<sup>a</sup> Doctor en Psicología.

<sup>b</sup> Magíster en Psicología.

<sup>c</sup> Estudiante de Psicología.

estudio metaanalítico muestra que el PHQ-9 posee buena consistencia interna y su uso es apropiado con propósitos de investigación.

**Palabras clave:** depresión; PHQ-9; metaanálisis; generalización de la fiabilidad; alfa de Cronbach.

## ABSTRACT

Depression is a significant public health problem, which makes it necessary to understand the psychometric properties of the instruments used for its evaluation to facilitate better decision-making. The Patient Health Questionnaire (PHQ-9) is an instrument that achieves optimal accuracy levels in measuring depression. Therefore, the objective of this study was to conduct a meta-analysis to generalize its reliability. A systematic search conducted in Web of Science and Scopus between 2020 and 2022 retrieved 32 studies that applied the PHQ-9 and reported Cronbach's alpha. A random effects meta-analysis was performed to estimate the test reliability, while heterogeneity was assessed using Cochran's Q and I<sup>2</sup> statistics. Potential moderating variables were examined through meta-regression and subgroup analyses. The combined reliability of the PHQ-9 was 0.846 (95% CI: 0.824-0.862). The Egger test and Kendall's  $\tau$  rank correlation test indicated no significant evidence of publication bias in the evaluated data. Also, there was also statistically significant heterogeneity in the total sample, determined by the Q test and the I<sup>2</sup> variability index. The meta-regression analysis showed that the type of sample (clinical vs. general) is a statistically significant moderator. This meta-analytical study demonstrates that the PHQ-9 has good internal consistency and is suitable for research purposes.

**Keywords:** depression; PHQ-9; meta-analysis; reliability generalization; Cronbach's alpha.

## INTRODUCCIÓN

Para la Organización Mundial de la Salud (1), la depresión representa un problema de salud pública; según estimaciones a nivel internacional, más de 280 millones de personas la padecen, manifestándose en sufrimiento, irritabilidad, pena permanente, y escaso interés por las actividades familiares, escolares y laborales. Así, el 3,8 % de la población mundial padece depresión, siendo la prevalencia mayor en los adultos (5 %), así como en los adultos mayores (5,7 %); además, este malestar constituye un 4,3 % de la carga mundial de morbilidad y es una de las principales causas de discapacidad a nivel mundial, proyectándose que para el 2030 será la primera causa de esta condición (2).

El Instituto de Salud Mental Honorio Delgado-Hideyo Noguchi (3-6) realizó estudios epidemiológicos en el Perú, encontrando que el desorden más frecuente es la depresión, cuyas prevalencias de vida fueron 18,2 % en Lima, 16,2 % en la sierra, 21,4 % en la región de la montaña, y 14,1 % en la costa.

La depresión es el síntoma psiquiátrico más común en la atención primaria (7). Se estima que la prevalencia es alta en este entorno, especialmente entre personas con enfermedades crónicas (8). Es causa primordial de discapacidad, ausencia laboral, productividad disminuida o pérdida y mayor uso de recursos de atención médica (9, 10). Asimismo, está relacionada con la baja calidad de vida (11) y el costo elevado de la atención médica (12). Por lo tanto, la depresión tiene importantes implicaciones para la salud pública (13).

Según Kroenke (14), la depresión no se detecta en más de la mitad de los pacientes de atención primaria, casi el 30-50 % de los pacientes con depresión pasan desapercibidos durante el diagnóstico por un médico de atención primaria (13), siendo la evaluación inexacta una de las barreras importantes para una atención eficaz (15, 16). A pesar de la alta carga, la depresión —a menudo— no se reconoce ni se trata lo suficiente (17, 18).

Los instrumentos de detección son muy útiles en encuestas comunitarias y pueden usarse para referir

pacientes al personal de salud mental. De ahí la necesidad de evaluar las diferentes escalas de depresión, sus ventajas, limitaciones y propiedades psicométricas.

En relación con los instrumentos usados que evalúan la depresión, tenemos los siguientes: Inventario de Depresión de Beck (BDI), que evalúa la presencia y la severidad de la depresión; Escala Heteroaplicada de Hamilton (HRSD), que reconoce síntomas relacionados a la depresión; Inventario de Salud Mental de 5 reactivos (MHI-5), escala autoaplicable que identifica la presencia de indicadores del trastorno; Montgomery Asberg Depression Rating Scale (MADRS), escala heteroaplicada que mide la severidad de los síntomas; así como el Cuestionario de Salud del Paciente (PHQ-9), herramienta autoaplicable que estima la presencia de síntomas (19).

Por otra parte, es necesario analizar la validez y la fiabilidad de los instrumentos (20), situación crítica que reflejan Quispe y Lévano (21), quienes señalan que el 53,3 % de los psicólogos no examina las condiciones locales relacionadas al lugar de origen de los instrumentos, lo cual puede afectar la validez; además, el 42,2 % realiza evaluaciones usando test inadecuados; el 46,7 % realiza interpretaciones que no corresponden; y otro 48,9 % de los encuestados señala que para utilizar las pruebas de medición estas deberían ser autorizadas previa evaluación de su competencia. Asimismo, Moscoso (22) establece que los instrumentos actuales sobre depresión presentan niveles diagnósticos simplistas y ambiguos, no diferencian las dimensiones objetivas de los trastornos depresivos; asimismo, señala que los instrumentos existentes solo tienen el propósito de evaluar la intensidad de los síntomas y la frecuencia de la ocurrencia, tomando como referencia el *DSM-V*. En este contexto, Livia y Ortiz (23), en un estudio realizado sobre los test psicológicos en el Perú, señalaron que no existe una política de control que regule el uso adecuado de las pruebas y que asegure la fiabilidad de los resultados, por lo cual se hace un mal uso de las pruebas psicométricas por la falta de estandarización de los instrumentos.

Cjuno et al. (24) señalan que el PHQ-9 es una de las pruebas más estudiadas en la población peruana, cuyas evidencias de validez y fiabilidad en nuestro país no disponen de certeza, observándose una ausencia de la generalización de la consistencia interna, razón por la cual se ha decidido llevar a cabo un análisis metaanalítico de la fiabilidad.

El PHQ-9 se considera una variante autoaplicada de la herramienta diagnóstica PRIME-MD (Primary

Care Evaluation of Mental Disorder) para trastornos mentales comunes. Spitzer et al. (25) determinaron la validez y la utilidad diagnóstica en la atención primaria en clínicas de Estados Unidos, comparable con el PRIME-MD original administrado por un médico, así como en una muestra de pacientes de obstetricia y ginecología de Nueva York (26). El PHQ-9, en el apartado de depresión, puntúa cada uno de los nueve criterios del *DSM-IV* de 0 (nada) a 3 (casi diariamente). Su puntuación puede oscilar entre 0 y 27. Mide, asimismo, la sintomatología depresiva y se ha empleado frecuentemente en investigaciones como herramienta de cribado, principalmente en investigación epidemiológica. En el estudio original de Kroenke et al. (27), en donde utilizan las entrevistas para profesionales de salud mental (MHP) como criterio estándar, para un puntaje PHQ-9  $\geq 10$  se obtuvo una sensibilidad y especificidad del 88 % en ambos para la depresión mayor. Los puntajes del PHQ-9 de 5, 10, 15 y 20 presentaron depresión leve, moderada, moderadamente grave y grave, respectivamente, cumpliendo un doble propósito: la capacidad del instrumento como medio de apoyo al diagnóstico y la capacidad de identificar la severidad de los síntomas (27). En un reciente estudio sobre los aspectos psicométricos del PHQ-9, se encontró una distribución de los ítems en un solo factor con una varianza explicada de 54,8 % y un  $\alpha = 0,896$  (28).

### Enfoque de generalización de la fiabilidad

El metaanálisis es una metodología que combina y resume los resultados de varios estudios individuales, lo cual sintetiza la información de varias investigaciones y genera un aumento del tamaño de la muestra (29). Por otro lado, el mayor tamaño de muestra repercute en una mayor potencia estadística, mejorando la posibilidad de detectar diferencias cada vez menores y permite hacer más precisos los estimados de los intervalos de confianza. Este tipo de investigación se ha considerado como antídoto contra la llamada crisis de replicación que aflige a las ciencias sociales (30), dada la necesidad de la comunidad científica de poder resumir la información dispersa en la literatura.

Por lo señalado, efectuar un metaanálisis de la fiabilidad del PHQ-9 se hace necesario para conocer la precisión de las puntuaciones del instrumento, el mismo que no se ha efectuado, para así determinar si se puede generalizar la fiabilidad. En una primera indagación en las bases de datos Web of Science (WoS) y Scopus, no se encontraron metaanálisis sobre la fiabilidad de este instrumento; sin embargo, sí se ubicaron algunos

estudios de revisión sistemática que hacen referencia fundamentalmente a la validez (31, 32).

El estudio de la fiabilidad del instrumento es necesario, considerando que se trata de una propiedad psicométrica que cambia de acuerdo con variables externas, como la composición de la muestra o el contexto en el que se aplica (33, 34), por lo que es recomendable disponer de estimaciones de la fiabilidad a partir de diferentes diseños para lograr una mejor comprensión del error que afecta a las puntuaciones, variando sus coeficientes de acuerdo a las poblaciones, además de la variabilidad de las muestras de personas (20). Asimismo, las estimaciones de las puntuaciones de las pruebas deben basarse en sus propios datos, en lugar de ser inducidas por resultados anteriores (35), a lo cual Vacha-Haase et al. (36) denominaron como «inducción a la fiabilidad»; ello refleja la concepción errónea de que la consistencia de las puntuaciones es una propiedad inmutable. De igual manera, se ha indicado que una forma de observar el comportamiento de la consistencia interna de los puntajes de una prueba es estudiar con el metaanálisis la generalización de la fiabilidad y dar cuenta de la heterogeneidad.

La realización de un metaanálisis de confiabilidad es una metodología que permite encontrar factores involucrados en dicha variabilidad. También ayudará a estudiar la adecuación y generalización de un instrumento, considerando que la generalización de la fiabilidad corresponde al análisis psicométrico de las propiedades de un cuestionario en las distintas situaciones en las que ha sido aplicado (37).

Si bien se han realizado investigaciones de las propiedades psicométricas del PHQ-9, no se ha encontrado la ejecución de un metaanálisis de la generalización de la fiabilidad de los puntajes sobre dicho instrumento; por tanto, los objetivos de la investigación fueron describir las características de los estudios, estimar la fiabilidad media de las puntuaciones de la escala, evaluar la heterogeneidad de las estimaciones de la fiabilidad y la búsqueda de variables moderadoras que permitan dar cuenta de tal variabilidad.

## MATERIALES Y MÉTODOS

### Búsqueda y selección de artículos

El estudio se inició con una revisión de la literatura, de acuerdo al método PRISMA (38), tomando en consideración los criterios de inclusión-exclusión y ecuación de búsqueda, todo en relación con el objetivo

del estudio, para luego realizar el metaanálisis, siguiendo las indicaciones de la lista de verificación REGEMA (REliability GEneralization Meta-Análisis) (39).

### Búsqueda bibliográfica

Las investigaciones seleccionadas debían cumplir cuatro criterios: i) ser una investigación empírica que haya aplicado el PHQ-9 en personas de las diferentes edades; ii) haberse informado entre el 2020 y el 2022; iii) reportar el estadístico de fiabilidad alfa de Cronbach obtenido con los propios datos de la muestra; y iv) estar en el idioma español o inglés.

La búsqueda bibliográfica se realizó con cuatro investigadores independientes, quienes analizaron la evaluación de los requisitos de elegibilidad e inclusión. Los artículos fueron seleccionados de las bases de datos WoS y Scopus, en los cuales se efectuó la búsqueda con tres descriptores de manera independiente: “Psychometrics” AND “Patient Health Questionnaire-9” (PHQ-9), “Validity” AND “Reliability” AND “patient Health Questionnaire-9” (PHQ-9), “Reliability” AND “patient Health Questionnaire- 9” (PHQ-9).

El proceso se inició revisando 528 registros de búsqueda (369 de WoS y 159 de Scopus), de los cuales 150 eran duplicados, por lo que se analizaron los resúmenes de los 378 restantes. Posteriormente, se excluyeron 232 artículos debido a que comprendían estudios no psicométricos. Se evaluaron los 146 estudios restantes para determinar su elegibilidad, de los cuales se excluyeron 114, por contener otras versiones del instrumento PHQ-9 (n = 50), por tener acceso solamente al *abstract* (n = 24) y por no contener estudios de fiabilidad (n = 40). Finalmente, el número de documentos seleccionados quedó conformado por 32 artículos que cumplieron los criterios requeridos para realizar el estudio. La búsqueda de artículos se realizó de acuerdo con el método PRISMA para las revisiones sistemáticas, las mismas que se detallan en la figura 1.

### Extracción y codificación de los datos

De los 32 estudios seleccionados, se obtuvieron los coeficientes de fiabilidad para las diversas muestras. A cada uno de los reportes se aplicó una lista de chequeo, la misma que incluyó las siguientes variables: autor, año de publicación, tamaño muestral, continente (Asia, África, Europa, Latinoamérica, América del Norte), idioma (inglés y otros), tipo de muestra (general,

problemas de salud), grupo generacional (adolescente, joven, adulto, adulto mayor, mixta) y coeficiente de fiabilidad alfa de Cronbach.

### Análisis estadísticos

Los estudios seleccionados reportaron al menos un coeficiente de confiabilidad alfa obtenido de los análisis con la muestra. Con el objetivo de normalizar su distribución, previo a la elaboración de los análisis estadísticos, los valores alfa de Cronbach se modificaron mediante la fórmula propuesta por Bonett (40), con la finalidad de realizar un ajuste en los datos para que los coeficientes alfa sigan una distribución más simétrica y equilibrar sus varianzas, lo cual facilita su interpretación y análisis, asegurando, a su vez, robustez y precisión. Realizado los análisis estadísticos, los valores se devolvieron a la métrica del coeficiente alfa para comprender su interpretación.

Se consideró el modelo de efectos aleatorios para el análisis estadístico, de tal manera que los coeficientes alfa se ponderaron en función de la inversa de su varianza. La heterogeneidad obtenida por las estimaciones de la fiabilidad se calculó mediante el estadístico Q de Cochran y el índice I<sup>2</sup>. Se realizó un análisis de moderación usando un análisis de varianza (ANOVA). El Test of Moderators, en el contexto

del análisis de moderadores, evalúa si las variables moderadoras tienen un efecto significativo en el modelo. Para la variable de la escala, el resultado del análisis según el moderador categórico tipo fue QM. Los análisis se efectuaron en el paquete Metafor en un interfaz de R Studio.

Al tratarse de un estudio documental, no se realizaron intervenciones con personas ni muestras biológicas, ni fue necesario solicitar autorización de un comité de ética institucional.

### RESULTADOS

El proceso de revisión sistemática obtuvo un total de 528 artículos. Primero se retiraron 150 documentos que estaban duplicados. Luego se obtuvo un tamizaje a través de la revisión del título y el resumen, eliminándose 232 trabajos por no ser psicométricos o no tener datos de confiabilidad. Posteriormente, se efectuó un análisis completo de los 146 estudios y se eliminaron 114 por no cumplir con los criterios de inclusión, resultando un total de 32 investigaciones para el metaanálisis. En la figura 1, se distingue el diagrama PRISMA que describe los pasos en la búsqueda y definición de estudios, mientras que en la tabla 1 se observa el uso del PHQ-9 en los años 2020 (n = 13), 2021 (n = 9) y 2022 (n = 10).

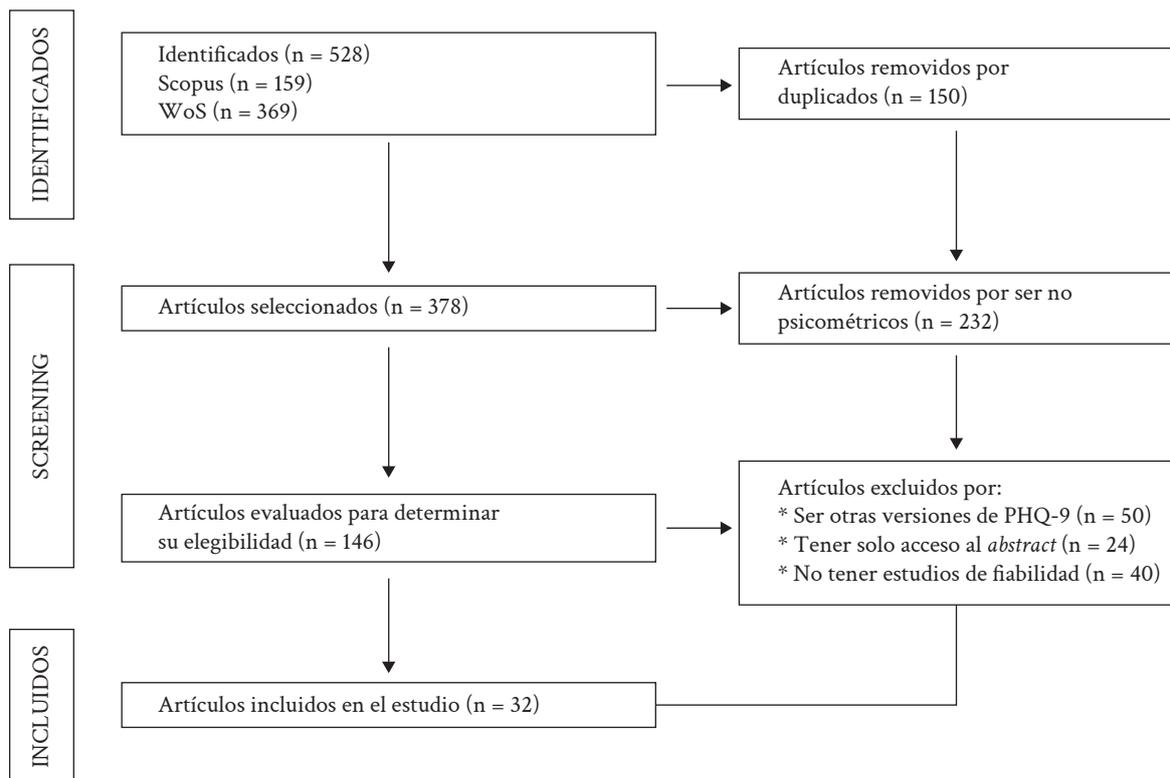


Figura 1. Diagrama del proceso de los artículos seleccionados.

Los estudios registrados se realizaron en diferentes países y continentes. El país que presentó más estudios fue China con 5 artículos. En cuanto a la distribución por continentes, se obtuvo lo siguiente: Asia (47 %), África (25 %), América del Sur (12,5 %), Europa (9,4 %) y América del Norte (6,3 %). Respecto a las etapas de vida, se realizaron estudios con adolescentes (6 %), jóvenes (3 %), personas adultas

(75 %), adultos mayores (3 %), adolescentes-adultos (3 %), adultos-adultos mayores (3 %) y jóvenes-adultos (6 %). De los artículos analizados, se observó que el tamaño de muestra fue de 39 747 sujetos. Se identificó que 19 estudios (59,4 %) se realizaron con muestras de personas que tenían problemas de salud, y 13 artículos en población general (40,6 %). Todos los trabajos incluyeron mujeres y varones (tabla 1).

**Tabla 1.** Variables de los estudios incluidos para el metaanálisis de generalización de la fiabilidad del PHQ-9.

N.º	Autores	Año	Muestra	Tipo	Grupo generacional	Alfa	Idioma	Continente-país
1	Aslan et al. (41)	2020	577	General	Adulto mayor	0,78	Español	Latinoamérica-Chile
2	Cumbe et al. (42)	2020	502	Problemas de salud	Adultos. Pacientes de atención primaria en sala de espera	0,84	Portugués	África-Mozambique
3	Dajpratham et al. (43)	2020	115	Problemas de salud	Adultos. Pacientes con accidente cerebrovascular	0,78	Tailandés	Asia-Tailandia
4	Degefa et al. (44)	2020	163	Problemas de salud	Adultos. Pacientes oncológicos ambulatorios	0,78	Amárico	África-Etiopía
5	Kigozi (45)	2020	208	Problemas de salud	Adultos. Pacientes con TBC que asisten a centros de atención primaria	0,84	Zulú	África-Sudáfrica
6	Leung et al. (46)	2020	10 933	General	Adolescentes	0,86	Mandarín	Asia-China
7	Molebatsi et al. (47)	2020	257	Problemas de salud	Adultos. Pacientes de cuidado primario en sala de espera	0,80	Inglés	África-Botsuana
8	Killian et al. (48)	2020	499	Problemas de salud	Adultos. Pacientes con depresión y distimia	0,67	Inglés	América del Norte-EE. UU.
9	Sebera et al. (49)	2020	434	Problemas de salud	Adulto. Pacientes con epilepsia	0,87	Inglés	África-Ruanda
10	Shin et al. (50)	2020	10 759	General	Adultos	0,79	Coreano	Asia-Corea
11	Sun et al. (51)	2020	109	Problemas de salud	Adultos. Pacientes con depresión mayor	0,89	Mandarín	Asia-China
12	Ye et al. (52)	2020	148	Problemas de salud	Adultos. Pacientes con psoriasis	0,94	Mandarín	Asia-China
13	Mwangi et al. (53)	2020	787	General	Adultos	0,84	Suajili	África-Kenia
14	Jiraniramai et al. (54)	2021	3204	General	Adultos trabajadores de la salud	0,80	Tailandés	Asia-Tailandia
15	Kianimoghadam et al. (55)	2021	150	Problemas de salud	Adultos. Pacientes con VIH	0,88	Persa	Asia-Irán
16	Pilunthanakul et al. (56)	2021	101	Problemas de salud	Adolescentes autistas	0,81	Inglés	Asia-Singapur
17	Quiñonez-Freire et al. (57)	2021	366	Problemas de salud	Adultos. Pacientes hospitalarios	0,85	Español	Latinoamérica-Ecuador

N.º	Autores	Año	Muestra	Tipo	Grupo generacional	Alfa	Idioma	Continente-país
18	Wisting et al. (58)	2021	793	General	Adultas	0,92	Noruego	Europa-Noruega
19	Ma et al. (59)	2021	1741	Problemas de salud	Adultos. Pacientes con un trastorno depresivo mayor	0,89	Mandarín	Asia-China
20	Farrahi et al. (60)	2021	463	General	Adultos. Estudiantes universitarios	0,86	Persa	Asia-Persia
21	Lopez et al. (61)	2021	368	Problemas de salud	Adultos con síndrome de Down	0,87	Inglés	Europa-Reino Unido
22	Miller et al. (62)	2021	2969	General	Adolescentes-adultos	0,73	Inglés	África-Uganda
23	Sun et al. (63)	2022	277	Problemas de salud	Adultos del servicio de neurología	0,84	Mandarín	Asia-China
24	Yona et al. (64)	2021	103	Problemas de salud	Adultos, adultos mayores. Pacientes	0,79	Hebreo	Asia-Israel
25	Matrángolo et al. (65)	2022	1100	General	Adultos	0,86	Español	Latinoamérica-Argentina
26	Mohamadian et al. (66)	2022	246	General	Jóvenes. Estudiantes de Ciencias Médicas	0,82	Persa	Asia-Irán
27	Errazuriz et al. (67)	2022	897	General	Jóvenes, adultos. Inmigrantes de habla hispana	0,90	Español	Latinoamérica-Chile
28	Rahman et al. (68)	2022	677	General	Jóvenes, adultos. Universitarios	0,82	Bengalí	Asia-Bangladés
29	Molebatsi et al. (69)	2022	257	Problemas de salud	Adultos. Asistentes de atención primaria	0,80	Inglés	África-Botsuana
30	Beswick et al. (70)	2022	103	Problemas de salud	Adultos. Pacientes con esclerosis múltiple	0,89	Inglés	Europa-Escocia
31	Ghazisaeedi et al. (71)	2022	400	General	Adultos universitarios	0,88	Persa	Asia-Irán
32	Beaman et al. (72)	2022	41	Problemas de salud	Adultos. Pacientes de la clínica	0,86	Inglés	América del Norte-EE. UU.

### Estimación de la fiabilidad media

El análisis de los 32 artículos identificados señaló valores alfa entre ,67 y ,94. El 97 % de los estudios alcanzó valores alfa de Cronbach por encima de ,70 (n = 31) y solo un trabajo reportó un alfa de ,67. El 19 % (n = 6) mostró valores entre ,70 y ,79; el 69 % (n = 22), entre ,80 y ,89; y solo el 3 % (n = 1) se ubicó por encima de ,90. Los análisis con las puntuaciones normalizadas mediante la transformación de Bonett para la muestra de 32 coeficientes alfa analizada mediante un modelo de efectos aleatorios alcanzó un valor de ,846 (IC 95 %: ,824-,862). Por otra parte, el metaanálisis efectuado con los alfa de Cronbach sin transformar dio un resultado global de ,88 (IC 95 %: ,87-,90).

### Evaluación de la heterogeneidad general

Los resultados en relación con los potenciales sesgos de publicación fueron calculados con el test de Egger:  $t(30) = -0,466$ ;  $p = 0,642$ ;  $b = 0,469$  (IC 95 %: 0,424-0,515). Los hallazgos sugieren que no se encontró evidencia de sesgo de publicación significativo en los datos evaluados, ya que el valor de p es alto (0,466), lo que señala que no hubo una diferencia significativa. Además, se usó el test de Rank de correlación ordinal:  $\tau$  Kendall =  $-0,127$ ;  $p = 0,307$ , para sesgo de publicación, mostrando una correlación baja y estadísticamente no significativa. Este resultado indica que no se halló evidencia significativa de sesgo de publicación en los datos evaluados.

Para valorar la variabilidad del alfa de Cronbach en las diferentes muestras, se calculó la heterogeneidad. Se obtuvo una heterogeneidad estadísticamente significativa sobre la muestra total Q (gl = 31) 1477,92; p < 0,0001. Por otra parte, el índice de variabilidad fue

de 98,19 % (>75 %, considerado alto). En la figura 2, se muestra la media de cada alfa de Cronbach para todos los estudios de la escala, con sus respectivos intervalos de confianza.

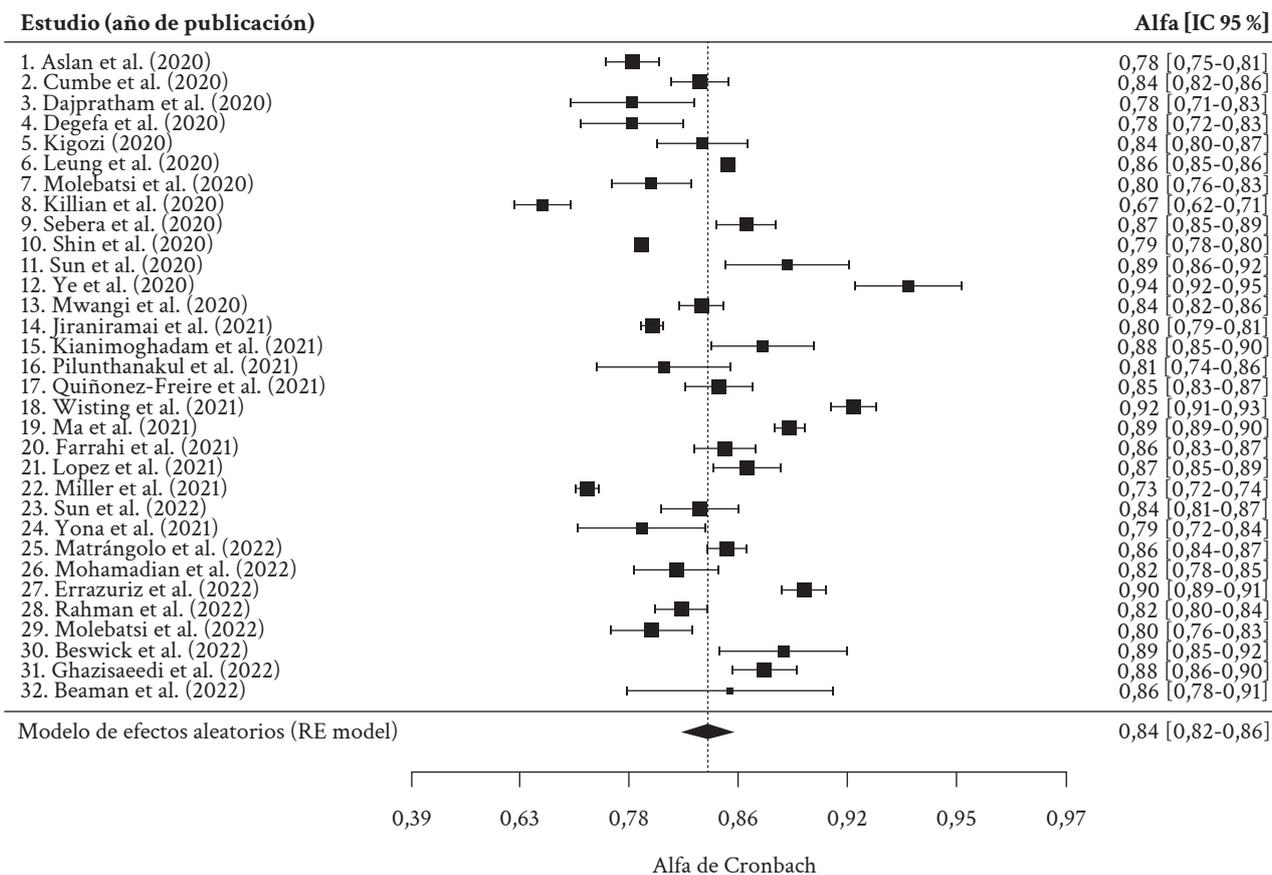


Figura 2. Forest plot del metaanálisis para la escala.

De los 32 estudios primarios, 11 presentaron valores medios e intervalos de confianza por debajo del promedio combinado en el metaanálisis, con una mediana de 0,79; 10 artículos compartieron valores con esta media, con una mediana de 0,84; y los valores

de los otros 11 estudios estuvieron por encima de esa media global combinada, con una mediana de 0,89. Las estadísticas descriptivas para cada factor se hallan en la tabla 2.

Tabla 2. Descriptivos del alfa de Cronbach para cada artículo según su posición respecto a la media global.

Posición relativa a la media global	Número de artículos (% del total)	Media	Mediana	Desviación estándar	Mínimo	Máximo
Por debajo	11(34 %)	0,776	0,79	0,041	0,67	0,81
Compartido	10(13 %)	0,845	0,84	0,015	0,82	0,86
Por encima	11(53 %)	0,893	0,89	0,022	0,87	0,94

### Análisis del moderador

Se realizó un análisis de moderación usando un ANOVA. El Test of Moderators, en el contexto del análisis de moderadores, evalúa si las variables moderadoras (*coefficients* 1:2) tienen un efecto significativo en el modelo. Para la variable de la escala, el resultado del análisis, según el moderador categórico tipo de muestra (clínico-general), fue  $QM(df = 2) = 873,05$ ;  $p < 0,0001$ , lo que indica una evidencia estadísticamente significativa en el papel moderador de la variable tipo de muestra. Es decir, la estimación de la fiabilidad es distinta según el tipo de muestra (general o clínico) al que se le aplicó la escala. En tal sentido, la estimación de fiabilidad para el tipo general fue 0,83 y para el tipo clínico fue 0,84.

Para el moderador categórico idioma, se obtuvo  $QM(df = 2) = 914,06$ ;  $p < 0,0001$ , lo que indica una evidencia estadísticamente significativa en el papel moderador de la variable idioma. Por tanto, la estimación de la fiabilidad difiere según el idioma en que se haya aplicado la escala. En efecto, la fiabilidad para la escala en inglés fue 0,823 y para otros idiomas fue 0,852.

Para el moderador categórico continente, se obtuvo  $QM(df = 2) = 1121,82$ ;  $p < 0,0001$ , lo que indica una evidencia estadísticamente significativa en el papel moderador de la variable continente. Por tanto, la estimación de la fiabilidad difiere según el continente en que se haya aplicado la escala. Así, la fiabilidad para la escala aplicada en África es 0,816; en Asia, 0,851; en Europa, 0,896; en Latinoamérica, 0,853; y en América del Norte, 0,764.

Para el moderador categórico grupo generacional, se obtuvo  $QM(df = 2) = 878,18$ ;  $p < 0,0001$ , lo que indica una evidencia estadísticamente significativa en el papel moderador de la variable grupo generacional. Por tanto, la estimación de la fiabilidad difiere según el grupo generacional en que se haya aplicado la escala. Así, la fiabilidad para la escala aplicada en adolescentes fue 0,831 y para adultos fue 0,847.

En cuanto al moderador categórico año de publicación, se obtuvo  $QM(df = 3) = 872,81$ ;  $p < 0,0001$ , lo que señala una evidencia estadísticamente significativa en el papel moderador de la variable del año de publicación. Por

tanto, la estimación de la fiabilidad difiere según el año de publicación en que se haya aplicado la escala. Así, la fiabilidad para la escala aplica en 2020 fue 0,826; para el 2021 fue 0,855; y para el año 2022 fue 0,847.

Para la variable numérica tamaño de la muestra de los estudios, se obtuvo  $QM(df = 1) = 0,476$ ;  $p < 0,490$ , lo que indica que no hay evidencia estadísticamente significativa en el papel moderador del tamaño de la muestra. En otras palabras, cambiar el tamaño de la muestra no tiene un impacto considerable en los resultados de la escala aplicada.

### Estimación robusta

Para la escala, se realizó un análisis de sensibilidad a fin de establecer qué tan sensible es la estimación combinada de la confiabilidad respecto a cada uno de los estudios. Primero se realizó un *leave-one-out* para detectar valores influyentes, se retiró un estudio a la vez. Se encontró una estimación promedio para la confiabilidad de 0,836, con desviación típica de 0,002, un mínimo de 0,834 y un máximo de 0,843, que al comparar con la estimación de la fiabilidad combinada de 0,846 no mostró distorsión respecto a los resultados globales. Esto indica que no hay estudios influyentes en el metaanálisis.

Por otro lado, se consideraron como estudios influyentes aquellos con un alfa de Cronbach fuera del intervalo de confianza (95 %) obtenido de la estimación combinada del metaanálisis. Luego de identificados los estudios potenciales, se retiraron y se repitió el análisis original. En la figura 3, se reporta la media combinada del coeficiente alfa de Cronbach de 0,845 (IC 95 % : 0,837-0,854); una heterogeneidad  $Q(gl = 10) = 28,31$ ,  $p < 0,002$ , estadísticamente significativa; y el índice de variabilidad fue de 60,69 % (entre 50 y 75 %, considerado moderado. Por otro lado, al restar la estimación combinada (0,846) y la estimación sin datos atípicos (0,845), podemos concluir que la atenuación fue muy pequeña (-1 %). Si bien la heterogeneidad disminuyó sustancialmente al quitar algunos estudios primarios, la estimación global no sufrió cambios importantes, lo que sugiere que los estudios excluidos no fueron influyentes en el metaanálisis.

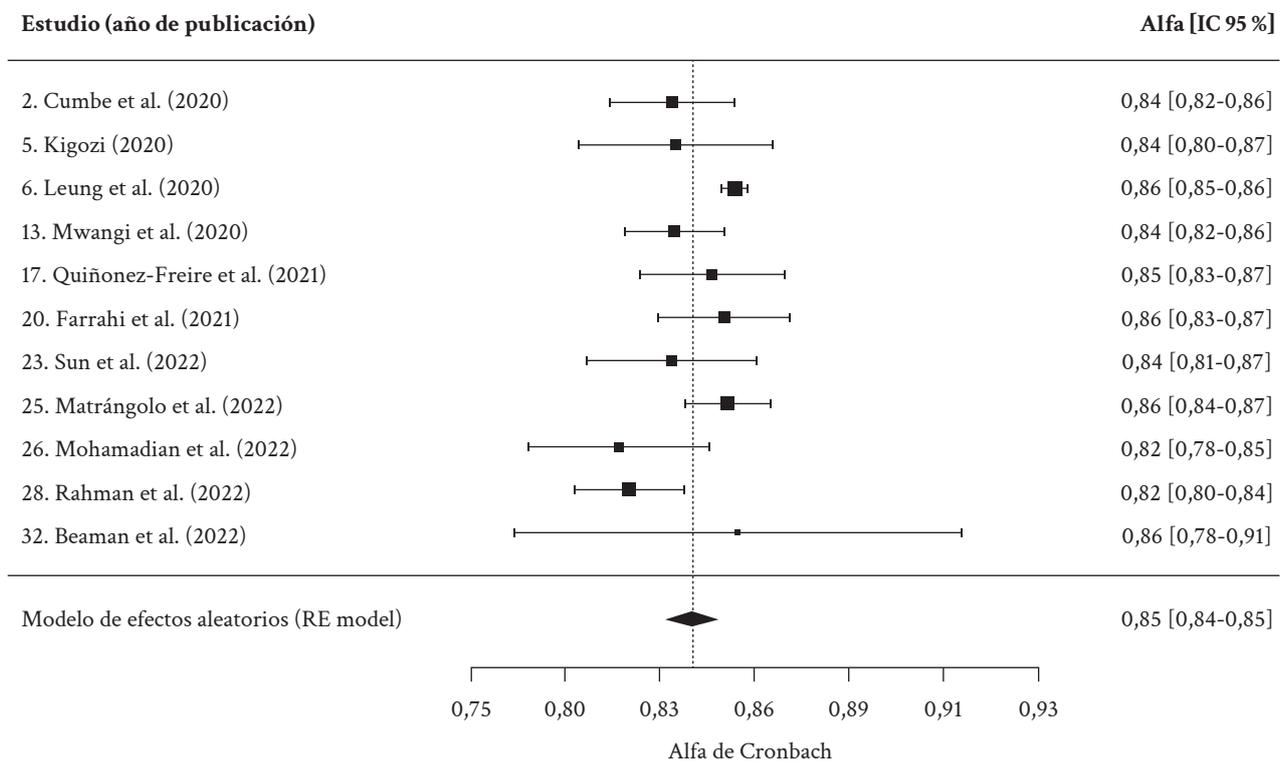


Figura 3. Forest plot del metaanálisis con estudios excluidos.

## DISCUSIÓN

El objetivo del estudio fue evaluar la fiabilidad de las puntuaciones del test PHQ-9 mediante un metaanálisis de generalización de la fiabilidad, el mismo que busca identificar la media de la varianza de error de medición entre estudios, utilizando un instrumento en particular, así como las fuentes de variabilidad de dichas variaciones entre los estudios (73), integrándose cuantitativamente alguna estimación de la fiabilidad obtenida al aplicar un determinado test psicológico (74).

La importancia de la fiabilidad de las puntuaciones de un test radica en que permite estimar su consistencia, en tanto no se trata de una propiedad estática, ni inherente al test, sino que varía según su estructura y las características de las muestras y el contexto de su aplicación (36, 75). Como señalan Prieto y Delgado (20), se debe evitar el error de considerar que un solo estudio de estimación de la fiabilidad refleja la verdadera y única confiabilidad del test, por lo que es necesario en la elaboración del test informar los métodos de cuantificación, las características de las muestras y las circunstancias en las que se han recopilado los datos, cuya responsabilidad también alcanza a los usuarios, no debiendo inducir la fiabilidad (36, 76).

Considerando que entre los problemas de salud mental más frecuentes y con mayor impacto en el bienestar y calidad de vida de las personas están los síntomas y trastornos depresivos (77), se hace necesaria su detección oportuna, además de disponer de instrumentos breves pero útiles para realizar tamizajes y hacer un seguimiento de las intervenciones. Asimismo, teniendo en cuenta que, dentro de los instrumentos de tamizaje, el PHQ-9 satisface el nivel óptimo de exactitud para depresión (78), surgió la necesidad de evaluar las propiedades psicométricas de este instrumento, pero de manera particular la fiabilidad de las puntuaciones, ya que sobre la validez hay varias investigaciones, como, por ejemplo, los referidos a la sensibilidad y especificidad (79, 80) o a su estructura unidimensional (81).

Dado que la confiabilidad no es una propiedad intrínseca del test, sino de las puntuaciones obtenidas del mismo en una aplicación específica, es imperativo realizar estudios metaanalíticos para generalizar la confiabilidad. Esto nos permitirá estimar la confiabilidad promedio y entender cómo varía en las diversas aplicaciones de una misma escala. A pesar de que el PHQ-9 es ampliamente utilizado en contextos aplicados y en estudios epidemiológicos, son escasos

los estudios que proporcionan algún tipo de coeficiente de confiabilidad con datos específicos de la muestra.

El coeficiente de fiabilidad promedio para el PHQ-9 fue de ,84, valor que se ubica por encima del límite de lo reportado en diversas recomendaciones (82, 83) que van a partir de ,60, aunque la mayoría coincide con lo establecido por Nunnally y Bernstein (84), quienes fijaron un valor mínimo de ,70 como confiabilidad aceptable para las investigaciones exploratorias. La escala de depresión muestra una confiabilidad media que logra el umbral de ,80, por lo que resulta satisfactorio para su uso con propósitos de investigación en general, debiendo evaluarse su uso en el contexto clínico, donde se consideran valores a partir de ,90 (85). Por los resultados, se puede concluir que la escala de depresión exhibe, en promedio, una fiabilidad moderada.

Los resultados examinados son sólidos y confiables, ya que los análisis de sensibilidad realizados muestran escasas diferencias con el resultado global obtenido en el metaanálisis. Además, no se observa sesgo de publicación, ya que la prueba de regresión de Egger presenta un resultado que no es estadísticamente significativo, complementado con el test de Rank de correlación ordinal y  $\tau$  Kendall para sesgo de publicación.

La confiabilidad observada mostró una considerable heterogeneidad entre los estudios, ya que los resultados fueron significativos. Estos hallazgos indicaron la presencia de una notable variabilidad entre los coeficientes de fiabilidad obtenidos en las distintas aplicaciones de la escala, lo que subraya la importancia de analizar las variables que podrían estar vinculadas a esta heterogeneidad.

Se puede considerar como limitaciones del estudio el tiempo utilizado en la selección de los artículos, ya que solo se tomó el período 2020-2022; asimismo, hubo estudios que incluyeron muestras combinadas por etapas de vida, por ejemplo, adolescentes-adultos, adultos-adultos mayores. Por tanto, se recomienda considerar otros tipos y coeficientes de fiabilidad.

## CONCLUSIONES

La presente investigación permite afirmar que el PHQ-9 brinda valores apropiados de fiabilidad para su uso con fines de investigación, según lo señalado por Nunnally y Bernstein (84). Los datos encontrados vinculados a su corta longitud (de tan solo nueve ítems) establecen una alternativa importante, especialmente para aquellas investigaciones en las

que se apliquen instrumentos para la medición de estudios multivariados, ya que este tipo de pruebas presenta ventajas, como tiempos de aplicación breves o tasas de respuesta más elevadas, en comparación con cuestionarios con un mayor número de reactivos (86).

**Conflicto de intereses:** Los autores declaran no tener conflictos de intereses.

**Financiamiento:** MOM, JVV e IGCA recibieron incentivo económico de la Universidad Nacional Francisco Villarreal como parte del concurso de proyectos de investigación, con incentivo aprobado con Resolución Rectoral n.º 1343-2023-UNFV.

## Contribución de autoría:

**MOM:** conceptualización, metodología, análisis formal, supervisión, redacción del borrador original, redacción (revisión y edición).

**JVV:** conceptualización, supervisión, administración de proyecto.

**IGCA:** conceptualización, administración de proyecto, supervisión.

**JARR:** curación de datos, administración de proyecto, redacción del borrador original, redacción (revisión y edición).

**EWLB:** curación de datos, validación.

**JLS:** curación de datos, análisis formal, redacción del borrador original, redacción (revisión y edición), validación.

**Agradecimientos:** A la Universidad Nacional Federico Villarreal por el apoyo brindado para la realización de este artículo.

## REFERENCIAS

1. Organización Mundial de la Salud. Depresión [Internet]. OMS; 2023, 31 de marzo. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/depression>
2. Organización Mundial de la Salud. Plan de Acción Integral sobre Salud Mental 2013-2030 [Internet]. OMS; 2022. Disponible en: <https://iris.who.int/bitstream/handle/10665/357847/9789240050181-spa.pdf?sequence=1>
3. Instituto Especializado de Salud Mental. Estudio epidemiológico metropolitano en salud mental 2002. Informe general. Anales de Salud Mental [Internet]. 2002; 18(1-2). Disponible en: <https://www.inism.gob.pe/investigacion/archivos/estudios/2002-ASM-EESM-M/files/res/downloads/book.pdf>

4. Instituto Especializado de Salud Mental. Estudio epidemiológico de salud mental en la sierra peruana 2003. Informe general. Anales de Salud Mental [Internet]. 2003; 19(1-2). Disponible en: <https://www.insm.gob.pe/investigacion/archivos/estudios/2003-ASM-EESM-SP/files/res/downloads/book.pdf>
5. Instituto Especializado de Salud Mental. Estudio epidemiológico de salud mental en la selva peruana 2004. Informe general. Anales de Salud Mental [Internet]. 2005; 21(1-2). Disponible en: <https://www.insm.gob.pe/investigacion/archivos/estudios/2004-ASM-EESM-SP/files/res/downloads/book.pdf>
6. Instituto Nacional de Salud Mental. Estudio epidemiológico de salud mental en fronteras 2005. Informe general. Anales de Salud Mental [Internet]. 2006; 22(1-2). Disponible en: <https://cdn.www.gob.pe/uploads/document/file/3661227/Estudio%20Epidemiológico%20de%20Salud%20Mental%20en%20Fronteras%202005.pdf.pdf>
7. McCarron RM, Shapiro B, Rawles J, Luo J. Depression. *Ann Intern Med* [Internet]. 2021; 174(5): ITC65-ITC80. Disponible en: <https://doi.org/10.7326/aitc202105180>
8. Motlhatlhedhi K, Molebatsi K, Wambua GN. Prevalence of depressive symptoms in urban primary care settings: Botswana. *Afr J Prim Health Care Fam Med* [Internet]. 2021; 13(1): a2822. Disponible en: <https://doi.org/10.4102/phcfm.v13i1.2822>
9. Michaud CM, Murray CJ, Bloom BR. Burden of disease—Implications for future research. *JAMA* [Internet]. 2001; 285(5): 535-539. Disponible en: <https://doi.org/10.1001/jama.285.5.535>
10. Johnston DA, Harvey SB, Glozier N, Calvo RA, Christensen H, Deady M. The relationship between depression symptoms, absenteeism and presenteeism. *J Affect Disord* [Internet]. 2019; 256(1): 536-540. Disponible en: <https://doi.org/10.1016/j.jad.2019.06.041>
11. Sivertsen H, Bjørkløf GH, Engedal K, Selbæk G, Helvik A. Depression and quality of life in older persons: a review. *Dement Geriatr Cogn Disord* [Internet]. 2015; 40(5): 311-319. Disponible en: <https://doi.org/10.1159/000437299>
12. König H, König HH, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci* [Internet]. 2019; 29: e30. Disponible en: <https://doi.org/10.1017/S2045796019000180>
13. Egede LE. Failure to recognize depression in primary care: issues and challenges. *J Gen Intern Med* [Internet]. 2007; 22: 701-703. Disponible en: <https://doi.org/10.1007/s11606-007-0170-z>
14. Kroenke K. Depression screening and management in primary care. *Fam Pract* [Internet]. 2018; 35(1): 1-3. Disponible en: <https://doi.org/10.1093/fampra/cmz129>
15. World Health Organization. The World Health Report 2001—Mental health: new understanding, new hope. *Bull World Health Organ* [Internet]. 2001; 79(11): 1085. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566704/>
16. Kramer T, Als L, Garralda ME. Challenges to primary care in diagnosing and managing depression in children and young people. *Br Med J* [Internet]. 2015; 350: h2512. Disponible en: <https://doi.org/10.1136/bmj.h2512>
17. Lecrubier Y. Is depression under-recognised and undertreated? *Int Clin Psychopharmacol* [Internet]. 1998; 13(Suppl 5): S3-S6. Disponible en: <https://doi.org/10.1097/00004850-199809005-00002>
18. Wittchen HU, Höfler M, Meister W. Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *Int Clin Psychopharmacol* [Internet]. 2001; 16(3): 121-135. Disponible en: <https://doi.org/10.1097/00004850-200105000-00001>
19. Grupo de trabajo de la Guía de Práctica Clínica sobre el Manejo de la Depresión en el Adulto. Guía de Práctica Clínica sobre el Manejo de la Depresión en el Adulto [Internet]. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad (ES); Agencia de Evaluación de Tecnologías Sanitarias de Galicia (avalia-t) (ES); 2014. Disponible en: [https://portal.guiasalud.es/wp-content/uploads/2018/12/GPC\\_534\\_Depresion\\_Adulto\\_Avaliat\\_compl.pdf](https://portal.guiasalud.es/wp-content/uploads/2018/12/GPC_534_Depresion_Adulto_Avaliat_compl.pdf)
20. Prieto G, Delgado AR. Fiabilidad y validez. *Pap Psicol* [Internet]. 2010; 31(1): 67-74. Disponible en: <https://www.papelesdelpsicologo.es/pdf/1797.pdf>
21. Quispe LV, Lévano JR. Uso de test psicológicos: perspectivas de psicólogos del sector público. *Rev Investig Psicol* [Internet]. 2021; 25: 81-97. Disponible en: [http://www.scielo.org.bo/pdf/rip/n25/n25\\_a08.pdf](http://www.scielo.org.bo/pdf/rip/n25/n25_a08.pdf)
22. Moscoso MS. Avances de la medición psicométrica de la depresión. *LIBERABIT* [Internet]. 2014; 20(1): 29-39. Disponible en: <http://www.scielo.org.pe/pdf/liber/v20n1/a04v20n1.pdf>

23. Livia J, Ortiz M. Los test psicológicos en el Perú: investigación, uso y abuso. *Psicología Actual*. 2014; 7: 45-49.
24. Cjuno J, Moya A, Calderón-Pérez E, Quispe-Illizarbe C, Mayon L, Livia J. Producción científica sobre validación y adaptación de instrumentos de tamizaje de la depresión en población peruana. *Rev Peru Med Exp Salud Pública* [Internet]. 2022; 39(3): 357-361. Disponible en: <https://doi.org/10.17843/rpmpesp.2022.393.11197>
25. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire*. *JAMA* [Internet]. 1999; 282(18): 1737-1744. Disponible en: <https://doi.org/10.1001/jama.282.18.1737>
26. Spitzer RL, William JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol* [Internet]. 2000; 183(3): 759-769. Disponible en: <https://doi.org/10.1067/mob.2000.106580>
27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* [Internet]. 2001; 16(9): 606-613. Disponible en: <https://doi.org/10.1046%2Fj.1525-1497.2001.016009606.x>
28. Galindo-Vázquez O, Núñez P, Vidal S, Bargalló E, Lerma A, Sánchez Y, et al. Evidencia psicométrica del Cuestionario sobre la Salud del Paciente (PHQ-9) y Escala de Ansiedad Generalizada (GAD-7) en pacientes con asesoramiento genético en oncología. *Psicooncología* [Internet]. 2023; 20(2): 267-281. Disponible en: <https://doi.org/10.5209/psic.91526>
29. Escrig VJ, Lluca JA, Granel L, Bellver M. Metaanálisis: una forma básica de entender e interpretar su evidencia. *Rev Senol Patol Mamar* [Internet]. 2021; 34(1): 44-51. Disponible en: <https://doi.org/10.1016/j.senol.2020.05.007>
30. Ones DS, Viswesvaran C, Schmidt FL. Realizing the full potential of psychometric meta-analysis for cumulative science and practice of human resource management. *Hum Resour Manag Rev* [Internet]. 2017; 27(1): 201-215. Disponible en: <https://doi.org/10.1016/j.hrmmr.2016.09.011>
31. Kaggwa MM, Najjuka SM, Ashaba S, Mamun MA. Psychometrics of the Patient Health Questionnaire (PHQ-9) in Uganda: a systematic review. *Front Psychiatry* [Internet]. 2022; 13: 781095. Disponible en: <https://doi.org/10.3389/fpsy.2022.781095>
32. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord* [Internet]. 2020; 279: 473-483. Disponible en: <https://doi.org/10.1016/j.jad.2020.09.131>
33. Sánchez-Meca J, Alacid de Pascual I, López-Pina J, Sánchez-Jiménez JC. Meta-análisis de generalización de la fiabilidad del inventario de obsesiones de Leyton versión para niños auto-aplicada. *Rev Esp Salud Pública* [Internet]. 2016; 90: e1-e14. Disponible en: <https://www.redalyc.org/articulo.oa?id=17043728056>
34. Cabedo-Peris J, Martí-Vilar M, Merino-Soto C, Ortiz-Morán M. Basic Empathy Scale: a systematic review and reliability generalization meta-analysis. *Healthcare* [Internet]. 2022; 10(1): 29. Disponible en: <https://doi.org/10.3390/healthcare10010029>
35. Sánchez-Meca J, López-Pina JA. El enfoque meta-analítico de generalización de la fiabilidad. *Acción Psicológica* [Internet]. 2008; 5(2): 37-64. Disponible en: <https://www.redalyc.org/pdf/3440/344030760005.pdf>
36. Vacha-Haase T, Kogan LR, Thompson B. Sample compositions and variabilities in published studies versus those in test manuals: validity of score reliability inductions. *Educ Psychol Meas* [Internet]. 2000; 60(4): 509-522. Disponible en: <https://doi.org/10.1177/00131640021970682>
37. Badenes-Ribera L, Rubio-Aparicio M, Sánchez-Meca J. Meta-análisis de generalización de la fiabilidad. *Información Psicológica* [Internet]. 2020; 119: 17-32. Disponible en: <https://doi.org/10.14635/IPSIC.2020.119.6>
38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Rev Esp Cardiol* [Internet]. 2021; 74(9): 790-799. Disponible en: <https://doi.org/10.1016/j.recresp.2021.06.016>
39. Sánchez-Meca J, Marín-Martínez F, López-López JA, Núñez-Núñez RM, Rubio-Aparicio M, López-García JJ, et al. Improving the reporting quality of reliability generalization meta-analyses: the REGEMA checklist. *Res Synth Methods* [Internet]. 2021; 12(4): 516-536. Disponible en: <https://doi.org/10.1002/jrsm.1487>

40. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. *J Educ Behav Stat* [Internet]. 2002; 27(4): 335-340. Disponible en: <https://doi.org/10.3102/10769986027004335>
41. Aslan J, Cova F, Saldivia S, Bustos C, Inostroza C, Rincon P, et al. Psychometric properties of the Patient Health Questionnaire-9 in elderly Chilean primary care users. *Front Psychiatry* [Internet]. 2020; 11: 555011. Disponible en: <https://doi.org/10.3389/fpsy.2020.555011>
42. Cumbe VF, Muanido A, Manaca MN, Fumo H, Chiruca P, Hicks L, et al. Validity and item response theory properties of the Patient Health Questionnaire-9 for primary care depression screening in Mozambique (PHQ-9-MZ). *BMC Psychiatry* [Internet]. 2020; 20: 382. Disponible en: <https://doi.org/10.1186/s12888-020-02772-0>
43. Dajpratham P, Pukrittayakamee P, Atsariyasing W, Wannarit K, Boonhong J, Pongpirul K. The validity and reliability of the PHQ-9 in screening for post-stroke depression. *BMC Psychiatry* [Internet]. 2020; 20: 291. Disponible en: <https://doi.org/10.1186/s12888-020-02699-6>
44. Degefa M, Dubale B, Bayouh F, Ayele B, Zewde Y. Validation of the PHQ-9 depression scale in Ethiopian cancer patients attending the oncology clinic at Tikur Anbessa specialized hospital. *BMC Psychiatry* [Internet]. 2020; 20: 446. Disponible en: <https://doi.org/10.1186/s12888-020-02850-3>
45. Kigozi G. Confirmatory factor analysis of the Patient Health Questionnaire-9: a study amongst tuberculosis patients in the Free State province. *S Afr J Infect Dis* [Internet]. 2020; 35(1): 242. Disponible en: <https://doi.org/10.4102/sajid.v35i1.242>
46. Leung DY, Mak YW, Leung SF, Chiang VC, Loke AY. Measurement invariances of the PHQ-9 across gender and age groups in Chinese adolescents. *Asia Pac Psychiatry* [Internet]. 2020; 12(3): e12381. Disponible en: <https://doi.org/10.1111/appy.12381>
47. Molebatsi K, Motlathledi K, Wambua GN. The validity and reliability of the Patient Health Questionnaire-9 for screening depression in primary health care patients in Botswana. *BMC Psychiatry* [Internet]. 2020; 20: 295. Disponible en: <https://doi.org/10.1186/s12888-020-02719-5>
48. Killian MO, Sánchez K, Eghaneyan BH, Cabassa LJ, Trivedi MH. Profiles of depression in a treatment-seeking Hispanic population: psychometric properties of the Patient Health Questionnaire-9. *Int J Methods Psychiatr Res* [Internet]. 2020; 30(1): e1851. Disponible en: <https://doi.org/10.1002/mpr.1851>
49. Sebera F, Nickenig JR, Umwiringirwa J, Teuwen DE, Boon PE, Dedeken P. Validity, reliability and cut-offs of the Patient Health Questionnaire-9 as a screening tool for depression among patients living with epilepsy in Rwanda. *PLoS ONE* [Internet]. 2020; 15(6): e0234095. Disponible en: <https://doi.org/10.1371/journal.pone.0234095>
50. Shin C, Ko Y, An H, Yoon H, Han C. Normative data and psychometric properties of the Patient Health Questionnaire-9 in a nationally representative Korean population. *BMC Psychiatry* [Internet]. 2020; 20: 194. Disponible en: <https://doi.org/10.1186/s12888-020-02613-0>
51. Sun Y, Fu Z, Bo Q, Mao Z, Ma X, Wang C. The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. *BMC Psychiatry* [Internet]. 2020; 20: 474. Disponible en: <https://doi.org/10.1186/s12888-020-02885-6>
52. Ye X, Shu HL, Feng X, Xia D, Wang Z, Mi W, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire-9 (C-PHQ-9) in patients with psoriasis: a cross-sectional study. *BMJ Open* [Internet]. 2020; 10(7): e033211. Disponible en: <http://dx.doi.org/10.1136/bmjopen-2019-033211>
53. Mwangi P, Nyongesa MK, Koot HM, Cuijpers P, Newton CR, Abubakar A. Validation of a Swahili version of the 9-item Patient Health Questionnaire (PHQ-9) among adults living with HIV compared to a community sample from Kilifi, Kenya. *J Affect Disord* [Internet]. 2020; 1: 100013. Disponible en: <https://doi.org/10.1016/j.jad.2020.100013>
54. Jiraniramai S, Wongpakaran T, Angkurawaranon C, Jiraporncharoen W, Wongpakaran N. Construct validity and differential item functioning of the PHQ-9 among health care workers: rasch analysis approach. *Neuropsychiatr Dis Treatment* [Internet]. 2021; 17: 1035-1045. Disponible en: <https://doi.org/10.2147/NDT.S271987>
55. Kianimoghadam AS, Arani AM, Mohraz M, Bakhtiari M, Manshadi SA, Alinaghi SA, et al. Psychometric properties of Persian version of Patient Health Questionnaire (PHQ-9) in an Iranian HIV-infected patients. *J Organ Behav Res* [Internet]. 2021; 6(2): 46-57. Disponible en: <https://doi.org/10.51847/OBS3N8Qdtx>
56. Pilunthanakul T, Goh TJ, Fung DS, Sultana R, Allen JC, Sung M. Validity of the Patient Health Questionnaire 9-item in autistic youths: a pilot

- study. *BMC Psychiatry* [Internet]. 2021; 21: 564. Disponible en: <https://doi.org/10.1186/s12888-021-03556-w>
57. Quiñonez-Freire C, Vara MD, Tomás JM, Baños RM. Psychometric properties of the Spanish version of the Patient Health Questionnaire-9 in users of the Ecuadorian public health care system. *Rev Lat Am Psicol* [Internet]. 2021; 53: 210-217. Disponible en: <https://doi.org/10.14349/rlp.2021.v53.23>
  58. Wisting L, Johnson SU, Bulik CM, Andreassen OA, Rø Ø, Bang L. Psychometric properties of the Norwegian version of the Patient Health Questionnaire-9 (PHQ-9) in a large female sample of adults with and without eating disorders. *BMC Psychiatry* [Internet]. 2021; 21: 6. Disponible en: <https://doi.org/10.1186/s12888-020-03013-0>
  59. Ma S, Yang J, Yang B, Kang L, Wang P, Zhang N, et al. The Patient Health Questionnaire-9 vs. the Hamilton Rating Scale for Depression in assessing major depressive disorder. *Front Psychiatry* [Internet]. 2021; 12: 747139. Disponible en: <https://doi.org/10.3389/fpsy.2021.747139>
  60. Farrahi H, Gharraee B, Oghabian MA, Zare R, Pirmoradi M, Batouli S, et al. Psychometric properties of the Persian version of Patient Health Questionnaire-9. *Iran J Psychiatry Clin Psychol* [Internet]. 2021; 27(2): 248-263. Disponible en: <http://doi.org/10.32598/ijpcp.27.2.3375.1>
  61. Lopez MB, Goldstein LH, Robinson EJ, Vitoratou S, Chalder T, Carson A, et al. Validation of the PHQ-9 in adults with dissociative seizures. *J Psychosom Res* [Internet]. 2021; 146: 110487. Disponible en: <https://doi.org/10.1016/j.jpsychores.2021.110487>
  62. Miller AP, Espinosa da Silva C, Ziegel L, Mugamba S, Kyasanku E, Bulamba RM, et al. Construct validity and internal consistency of the Patient Health Questionnaire-9 (PHQ-9) depression screening measure translated into two Ugandan languages. *Psychiatry Res Commu* [Internet]. 2021; 1(2): 100002. Disponible en: <https://doi.org/10.1016/j.psycom.2021.100002>
  63. Sun Y, Kong Z, Song Y, Liu J, Wang X. The validity and reliability of the PHQ-9 on screening of depression in neurology: a cross-sectional study. *BMC Psychiatry* [Internet]. 2022; 22: 98. Disponible en: <https://doi.org/10.1186/s12888-021-03661-w>
  64. Yona T, Weisman A, Gottlieb U, Lin E, Masharawi Y. The reliability and validity of the Hebrew Patient Health Questionnaire (PHQ-9) in the general population [preprint]. medRxiv [Internet]. 2021, 16 de julio. Disponible en: <https://doi.org/10.1101/2021.07.13.21260485>
  65. Matrángolo GM, Azzollini S, Simkin H. Evidencias de validez de la versión argentina del Cuestionario de Salud del Paciente-9 (CSP-9). *Anu Psicol* [Internet]. 2022; 52(2): 155-167. Disponible en: <https://doi.org/10.1344/ANPSIC2022.52/2.5>
  66. Mohamadian R, Khazaie H, Ahmadi SM, Fatmizade M, Ghahremani S, Sadeghi H, et al. The psychometric properties of the Persian versions of the Patient Health Questionnaires 9 and 2 as screening tools for detecting depression among university students. *Int J Prev Med* [Internet]. 2022; 13(1): 116. Disponible en: [https://doi.org/10.4103/ijpvm.IJPVM\\_213\\_20](https://doi.org/10.4103/ijpvm.IJPVM_213_20)
  67. Errazuriz A, Beltrán R, Torres R, Passi-Solar A. The validity and reliability of the PHQ-9 and PHQ-2 on screening for major depression in Spanish speaking immigrants in Chile: a cross-sectional study. *Int J Environ Res Public Health* [Internet]. 2022; 19(21): 13975. Disponible en: <https://doi.org/10.3390/ijerph192113975>
  68. Rahman MA, Dhira TA, Sarker AR, Mehareen J. Validity and reliability of the Patient Health Questionnaire scale (PHQ-9) among university students of Bangladesh. *PLoS ONE* [Internet]. 2022; 17(6): e0269634. Disponible en: <https://doi.org/10.1371/journal.pone.0269634>
  69. Molebatsi K, Motlathledi K, Wambua GN. The validity and reliability of the Patient Health Questionnaire-9 for screening depression in primary health care patients in Botswana. *BMC Psychiatry* [Internet]. 2020; 20: 295. Disponible en: <https://doi.org/10.1186/s12888-020-02719-5>
  70. Beswick E, Quigley S, Macdonald P, Patrick S, Colville S, Chandran S, et al. The Patient Health Questionnaire (PHQ-9) as a tool to screen for depression in people with multiple sclerosis: a cross-sectional validation study. *BMC Psychol* [Internet]. 2022; 10: 281. Disponible en: <https://doi.org/10.1186/s40359-022-00949-8>
  71. Ghazisaeedi M, Mahmoodi H, Arpaci I, Mehrdar S, Barzegari S. Validity, reliability, and optimal cut-off scores of the WHO-5, PHQ-9, and PHQ-2 to screen depression among university students in Iran. *Int J Ment Health Addict* [Internet]. 2022; 20(3): 1824-1833. Disponible en: <https://doi.org/10.1007/s11469-021-00483-5>
  72. Beaman J, Lawson L, Keener A, Mathews ML. Within clinic reliability and usability of a voice-based Amazon Alexa Administration of the Patient Health Questionnaire 9 (PHQ 9). *J Med*

- Syst [Internet]. 2022; 46(6): 38. Disponible en: <https://doi.org/10.1007/s10916-022-01816-0>
73. Vacha-Haase T. Reliability generalization: exploring variance in measurement error affecting score reliability across studies. *Educ Psychol Meas* [Internet]. 1998; 58(1): 6-20. Disponible en: <https://doi.org/10.1177/0013164498058001002>
  74. Badenes-Ribera L, Rubio-Aparicio M, Sánchez-Meca J. Meta-análisis de generalización de la fiabilidad. *Información Psicológica* [Internet]. 2020; 119(6): 17-32. Disponible en: <https://doi.org/10.14635/IPSIC.2020.119.6>
  75. López-Nicolás R, Rubio-Aparicio M, López-Ibáñez C, Sánchez-Meca J. A reliability generalization meta-analysis of the Dimensional Obsessive-Compulsive Scale. *Psicothema* [Internet]. 2021; 33(3): 481-489. Disponible en: <https://doi.org/10.7334/psicothema2020.455>
  76. Sánchez-Meca J, López-Pina JA, López JA. Generalización de la fiabilidad: un enfoque metaanalítico aplicado a la fiabilidad. *Fisioterapia* [Internet]. 2009; 31(6): 262-270. Disponible en: <https://www.um.es/metaanalysis/pdf/5026.pdf>
  77. Saldivida S, Aslan J, Cova F, Vicente B, Inostroza C, Rincón P. Propiedades psicométricas del PHQ-9 (Patient Health Questionnaire) en centros de atención primaria de Chile. *Rev Méd Chile* [Internet]. 2019; 147(1): 53-40. Disponible en: <http://dx.doi.org/10.4067/S0034-98872019000100053>
  78. Pettersson A, Boström KB, Gustavsson P, Ekselius L. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nord J Psychiatry* [Internet]. 2015; 69(7): 497-508. Disponible en: <https://doi.org/10.3109/08039488.2015.1008568>
  79. Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *Br Med J* [Internet]. 2019; 365: 11476. Disponible en: <https://doi.org/10.1136/bmj.11476>
  80. Negeri ZF, Levis B, Sun Y, He C, Krishnan A, Wu Y. et al. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *Br Med J* [Internet]. 2021; 375: n2183. Disponible en: <https://doi.org/10.1136/bmj.n2183>
  81. Bianchi R, Verkuilen J, Toker S, Schonfeld IS, Gerber M, Brähler E, et al. Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. *Psychol Assess* [Internet]. 2022; 34(6): 595-603. Disponible en: <https://doi.org/10.1037/pas0001124>
  82. Ponterotto J, Ruckdeschel DE. An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Percept Mot Skills* [Internet]. 2007; 105(3 Pt 1): 997-1014. Disponible en: <https://doi.org/10.2466/pms.105.3.997-1014>
  83. Carmines EG, Zeller RA. *Reliability and Validity Assessment*. Sage Publications; 1979.
  84. Nunnally JC, Bernstein IH. *The Assessment of Reliability*. Psychom Theory; 1994.
  85. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* [Internet]. 1994; 6(4): 284-290. Disponible en: <https://doi.org/10.1037/1040-3590.6.4.284>
  86. Hidalgo-Fuentes S. Smartphone Application-Based Addiction Scale: un meta-análisis de generalización de la fiabilidad. *Eur J Educ Psychol* [Internet]. 2023; 16(1). Disponible en: <https://doi.org/10.32457/ejep.v16i1.1954>