



UNIVERSIDAD PERUANA
CAYETANO HEREDIA
FACULTAD DE PSICOLOGÍA

ANÁLISIS DE LAS PROPIEDADES
PSICOMÉTRICAS DE LA VERSIÓN
PERUANA DEL CUESTIONARIO DE
SALUD DEL PACIENTE PHQ-9

TESIS PARA OPTAR EL TÍTULO DE
LICENCIADO EN PSICOLOGÍA

KENNEDY LUIGUI QUISPILAYA CAPCHA

LIMA – PERÚ

2021

MIEMBROS DEL JURADO

Dra. Elizabeth Dany Araujo Robles

Presidente

Lic. Mey Ling Rosa Chang Marcovich

Vocal

Mg. Chiara Fiorella Branizza Colarossi

Secretaria

ASESOR DE TESIS

Lic. Giuliana Salazar Alvarez

DEDICATORIA

A mi difunto padre Dario Ángel y a mi madre Albina por ser la base de mi educación. A mis hermanos, Darky y Sadith por su apoyo incondicional y motivarme a ser mejor.

AGRADECIMIENTOS

A mi familia por todo el apoyo brindado, a la universidad y beca-18 por brindarme la oportunidad de estudiar una carrera profesional, a mis profesores por compartir sus conocimientos, a mi asesora Giuliana por acompañarme en todo el proceso y a todas las personas que me motivaron a culminar la tesis.

TABLA DE CONTENIDOS

RESUMEN	
ABSTRACT	
INTRODUCCIÓN	11
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	13
IDENTIFICACIÓN DEL PROBLEMA	13
JUSTIFICACIÓN E IMPORTANCIA DEL PROBLEMA.....	16
LIMITACIONES DE LA INVESTIGACIÓN.....	17
OBJETIVOS DE LA INVESTIGACIÓN	18
OBJETIVO GENERAL	18
OBJETIVOS ESPECÍFICOS	18
CAPÍTULO II: MARCO TEÓRICO	19
ASPECTOS CONCEPTUALES PERTINENTES	19
1.1. LA DEPRESIÓN.....	19
1.2. PROPIEDADES PSICOMÉTRICAS	26
2. ANTECEDENTES.....	34
2.1 INVESTIGACIONES NACIONALES.....	34
2.2 INVESTIGACIONES INTERNACIONALES	36
3. DEFINICIONES CONCEPTUALES Y OPERACIONALES DE LAS VARIABLES.....	38
HIPÓTESIS	40
HIPÓTESIS GENERAL	40
HIPÓTESIS ESPECÍFICOS	40
CAPÍTULO III: METODOLOGÍA	41

1. NIVEL Y TIPO DE INVESTIGACIÓN	41
2. DISEÑO DE LA INVESTIGACIÓN	41
3. NATURALEZA DE LA MUESTRA	41
3.1. DESCRIPCIÓN DE LA POBLACIÓN	41
3.2. MUESTRA Y MÉTODO DE MUESTREO	42
3.3. CRITERIOS DE INCLUSIÓN	42
3.4 CRITERIOS DE EXCLUSIÓN	43
4. INSTRUMENTO	43
5. PROCEDIMIENTO	44
6. CONSIDERACIONES ÉTICAS.....	45
7. PLAN DE ANÁLISIS DE DATOS	46
CAPITULO IV: RESULTADOS	47
CAPITULO V: DISCUSIÓN	58
CAPÍTULO VI: CONCLUSIÓN	69
RECOMENDACIONES	70
REFERENCIAS BIBLIOGRÁFICAS	71
ANEXOS.....	83
1. Consentimiento informado	83
2. Prueba PHQ-9 obtenida de la encuesta demográfica y de salud familiar ENDES 2019	84

RESUMEN

En Perú, durante la pandemia por COVID-19 y el tiempo de toque de queda, se obtuvo como resultado que el 28.5% de encuestados indicaron presentar sintomatología depresiva, de este subgrupo, el 41% de los participantes presentaron síntomas asociados a la depresión moderada y severa; y el 12.8% refirió pensamientos suicidas. Por lo que se sugiere que la depresión sea evaluada en todas las fases de la enfermedad y tratamiento, más aún post-pandemia. Se analizó las propiedades psicométricas de la versión peruana del cuestionario sobre la salud del paciente (PHQ-9) con una muestra de 33613 personas, obtenidas de una base de datos secundaria de la ENDES 2019. La metodología es de nivel básico no experimental y con diseño instrumental. Los resultados indican buena calidad de los ítems ($IHC > .30$, $Id < 0.05$ y $h^2 > .50$) con capacidad para medir el mismo constructo y diferenciar a las personas con diferentes niveles de depresión. Se identificó un modelo factorial de dos dimensiones (somático y cognitivo) con una adecuada bondad de ajuste ($RMSEA = .097$; $CFI = .932$; $NFI = .932$; $TLI = .906$). Además, se obtuvo un alto nivel de confiabilidad ($\omega = .892$). Finalmente, se establecen las normas percentilares con 5 niveles para medir la sintomatología depresiva.

Palabras claves: Depresión, PHQ-9, propiedades psicométricas, población peruana, tamizaje

ABSTRACT

In Peru, during the COVID-19 pandemic and the curfew time, it was obtained that 28.5% of respondents indicated presenting depressive symptoms, of this subgroup, 41% of the participants showed symptoms associated with moderate depression and severe; and 12.8% reported suicidal thoughts. Therefore, it is suggested that depression is in all phases of the disease and treatment, even more post-pandemic. The psychometric properties of the Peruvian version of the questionnaire on patient health (PHQ-9) were analyzed with a sample of 33,613 people, obtained from a secondary database of the ENDES 2019. The methodology is of a non-experimental basic level and with instrumental design. The results indicate good quality of the items ($IHC > .30$, $Id < 0.05$ and $h^2 > .50$) with the capacity to measure the same construct and differentiate people with different levels of depression. A two-dimensional factorial model (somatic and cognitive) with adequate goodness of fit was identified ($RMSEA = .097$; $CFI = .932$; $NFI = .932$; $TLI = .906$). In addition, a high level of reliability was obtained ($\omega = .892$). Finally, the percentile norms were achieved with 5 levels to measure depressive symptoms.

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

INTRODUCCIÓN

Como se sabe los trastornos psiquiátricos son los que ocupan el primer lugar dentro de los indicadores de carga por enfermedad, por ejemplo, la depresión, que además es la que mayor porcentaje de prevalencia tiene en pacientes con patologías no transmisibles, también sigue siendo subdiagnosticada y con brechas amplias de atención, a pesar del enfoque de atención primaria (INS, 2018). Es por ello que, Spitzer, Kroenke y Williams (1999) construyeron el cuestionario sobre la salud del paciente (PHQ-9) como instrumento de tamizaje para la depresión.

Por consiguiente, este estudio tiene como objetivo analizar las propiedades psicométricas de la versión peruana del cuestionario sobre la salud del paciente PHQ-9, para ello se recurrió a la base de datos de acceso libre realizado por la ENDES 2019.

La presente investigación consta de 6 capítulos que se detallan en las líneas siguientes:

En el primer capítulo se detalla el planteamiento del problema, así como la justificación e importancia motivo por la cual se realizó la investigación, posterior a ello se explica las limitaciones y se plantea los objetivos de estudio.

En el segundo capítulo se aborda los aspectos conceptuales de la investigación, antecedentes nacionales e internacionales en relación al tema y la definición conceptual y operacional de la variable en cuestión.

En el tercer capítulo se expone el método de la investigación, el nivel y diseño, asimismo se realiza una descripción de la población, muestra y se explica

los criterios de inclusión y exclusión. Por último, se detalla el procedimiento y las consideraciones éticas.

En el cuarto capítulo se presentan los resultados respondiendo a cada objetivo propuesto, para posterior pasar al quinto capítulo, donde se desarrolla la discusión de dichos resultados. Finalmente, en el sexto y último capítulo se da a conocer las conclusiones y recomendaciones de la presente investigación.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

IDENTIFICACIÓN DEL PROBLEMA

En el contexto peruano, los padecimientos neuropsiquiátricos representan el grupo de daños con mayor carga de enfermedad y años de calidad de vida perdidos, considerando de mayor prevalencia la depresión, el trastorno de ansiedad y el abuso del consumo de alcohol, entre otras. En los últimos datos publicados, se pone en manifiesto que la depresión será el principal problema de salud mental sobre la población. (OMS, 2018).

En cuanto a la prevalencia, la sintomatología depresiva oscila entre el 14% y el 12.5% en las zonas urbanas y rurales respectivamente (Villarreal-Zegarra, Copez-Lonzoy, Bernabe-Ortiz, Melendez-Torres, Bazo-Alvarez. 2020), asimismo, pueden variar por diferentes factores, biológicos, sociodemográficos o de estilo de vida. Es así que, la Organización Panamericana de la Salud (2017) indicó que la presencia de la depresión es más común en las mujeres (5.1%) que en los hombres (3.6%) considerando la población mundial (322 millones), también se evidencia que es más frecuente en adultos mayores y personas con nivel socioeconómico bajo. En el contexto peruano, el MINSA (2019) indicó que el 20% de la población adulta y adulta mayor sufre de un trastorno mental especialmente de depresión, y que, en la población mayor de 12 años, los trastornos más frecuentes son los episodios depresivos.

Además, en Perú, los hallazgos previos de la encuesta poblacional sobre salud mental durante la pandemia por COVID-19 y el tiempo de toque de queda,

donde participaron 58,349 personas, se obtuvo como resultado que el 28.5% de todos los encuestados indicaron presentar sintomatología depresiva.

De ese subgrupo, el 41% de los participantes presentaron síntomas asociados a la depresión moderada y severa; y el 12.8% refirió pensamientos suicidas. Los hombres en un 23.4% reportaron sintomatología depresiva, mientras que las mujeres, en un 30.8%. Lo que resalta la vulnerabilidad del sexo femenino frente a la depresión, así como el grupo con mayor frecuencia de afectación depresiva fue el de 18 a 24 años (Antiporta, Cutipe, Mendoza, Celentano, Stuart y bruni. 2021).

Lo más alarmante de la depresión son las consecuencias, el cual comienza con un nivel de angustia que afecta en la capacidad para realizar tareas diarias, incluso las más simples. Asimismo, tiene un impacto negativo en las relaciones familiares, amicales y en el trabajo (Morales, 2017). Por ejemplo, aumenta el ausentismo, el deterioro del desempeño, la baja por enfermedad, entre otros (Correia, De Souza, Susser, y Rossi. 2017). Por otro lado, lo más extremo de la depresión es el suicidio, el cual, según la OMS (2017) aumenta el riesgo para la conducta suicida en un 21% (OMS, 2018). Es por ello que es necesario contar con instrumentos para el diagnóstico temprano de depresión.

Respecto a las pruebas psicométricas que se utilizan en el primer nivel de atención, se han reportado que el inventario de depresión de Beck es el más aceptado por la comunidad de profesionales de la salud mental, además por tener alta validez y confiabilidad en diferentes contextos y poblaciones, sin embargo la versión reducida BDI-PC, por sus siglas en inglés, contiene 7 ítems y solo evalúa

síntomas psicológicos, por lo que no se muestra su eficacia en poblaciones con alguna enfermedad crónica (Lam, Contreras, Mori, Gil, Espinoza, Córdova, 2008). Esto podría afectar las intervenciones que se plantean a partir de los resultados de un instrumento.

Por otra parte, la versión española del PHQ-9 que fue estudiada en Perú y en los otros países presenta contenidos que no son apropiadas para el contexto peruano a escala nacional, ya que se podría emplear versiones no adaptadas, adaptadas parcialmente o versiones adaptadas en otros países, lo que afectaría la confiabilidad y validez de los resultados (Calderón et al., 2012; Gil, 2019).

De esta manera, en el presente estudio se plantea como pregunta de investigación ¿Cuáles son las propiedades psicométricas del cuestionario de salud del paciente PHQ-9 para la población peruana?

JUSTIFICACIÓN E IMPORTANCIA DEL PROBLEMA

A nivel teórico la prueba Patient Questionnaire Health (PHQ-9), cuenta con una validez por el criterio de jueces para el contexto peruano (Calderón et al. 2012), además cuenta con estudios que evaluaron la invarianza y la estructura factorial, sin embargo, dicho estudio tiene una antigüedad de 5 años y se trabajó con un corte diferente de población, por lo que es necesario ser actualizada con información cercana a la pandemia generada por el covid-19. Considerando las consecuencias generadas por el covid-19 a nivel de la salud mental en la población peruana, es apropiado realizar esta investigación, ya que de esta manera se contará con un instrumento válido para el diagnóstico temprano de depresión en el contexto peruano.

Además, a nivel social y político, los resultados del presente estudio contribuirán con el plan nacional de fortalecimiento de servicios de salud mental comunitaria, implementada por el ministerio de salud desde el 2018 al 2021, mediante evaluaciones y diagnósticos tempranos en los pacientes del primer nivel de atención. Esto, mejorará el pronóstico y evitará la presencia de comorbilidades que pueden agravar los síntomas de la depresión, ya que se implementarían intervenciones a nivel individual o grupal de manera oportuna. De esta manera, se estaría contribuyendo con el aumento de la calidad de vida a nivel de la salud mental.

La revisión de la validez, confiabilidad y la creación de nuevos baremos de este instrumento para tamizaje de depresión ofrecerá múltiples beneficios. Es así que, a nivel práctico, emplear el PHQ-9 en el primer nivel de atención reducirá el tiempo en las evaluaciones para tamizaje de depresión, de esa manera se puede

evaluar a más personas en poco tiempo, esto considerando que el tiempo de las consultas en el primer nivel de atención son cortos y con alta demanda de pacientes.

Por otra parte, este estudio servirá como fuente bibliográfico y herramienta para desarrollar más investigaciones dentro del área de la salud mental, esto debido a que en el repositorio Nacional Digital de Ciencia, Tecnología e Innovación (Alicia) existen muchos estudios que toman en cuenta el PHQ-9 como instrumento de medición, sin embargo, solo se apoyan de la validez por criterio de jueces o se enfocan en validar para grupos de poblaciones pequeñas y específicas. Con este estudio se cubrirá esa necesidad de emplear instrumentos válidos en el contexto peruano para las investigaciones.

LIMITACIONES DE LA INVESTIGACIÓN

Debido a la pandemia y la necesidad del aislamiento social para este estudio se recurrió a una fuente de datos secundaria, por lo que no se puede analizar la validez de contenido, concurrente, la predictibilidad, sensibilidad y especificidad del cuestionario PHQ-9.

OBJETIVOS DE LA INVESTIGACIÓN

OBJETIVO GENERAL

- Analizar las propiedades psicométricas de la versión peruana del cuestionario de salud del paciente PHQ-9.

OBJETIVOS ESPECÍFICOS

- Identificar la calidad de los ítems de la versión peruana del cuestionario de salud del paciente PHQ-9.
- Identificar la validez relacionada al constructo mediante la validez factorial confirmatorio de la versión peruana del cuestionario de salud del paciente PHQ-9.
- Identificar la confiabilidad por consistencia interna y método de dos mitades de la versión peruana del cuestionario de salud del paciente PHQ-9.
- Elaborar los baremos del cuestionario de salud del paciente PHQ-9 para la población peruana.

CAPÍTULO II: MARCO TEÓRICO

ASPECTOS CONCEPTUALES PERTINENTES

1.1. LA DEPRESIÓN

A lo largo de la historia la mayor parte de los trastornos psiquiátricos han cambiado de significado, es así como la depresión se asociaba a la melancolía. El término depresión, como tal, se consideró en el siglo XIX, con el objetivo de indicar un estado de tristeza. Por su parte, Kraepelin consideraba los trastornos psiquiátricos como patologías basadas en modelos neurológicos, médicos, con etiología orgánica específica, por ejemplo, creía que la locura maníaco - depresiva estaba desligada del estrés psicológico. De esta manera es como clasificaba a la depresión (Paykel, 2008).

Posterior a ello, Meyer a inicios del siglo XX, después de sustituir el término melancolía por depresión, indicó que las enfermedades mentales, como la depresión, no tenían una etiología orgánica ni específica, tampoco se consideraría como consecuencia de una patología neurológica (Meyer, 1910). Por el contrario, él concibió la depresión como una consecuencia afectiva a sucesos vitales estresantes en los que intervienen factores psicológicos, biológicos y sociales (Ban, 2014). Por lo tanto, los planteamientos de Meyer serían un precursor del modelo biopsicosocial moderno.

El concepto moderno sobre la depresión, tomando en cuenta el CIE 10 y el DSM -V, es considerado esencialmente un síndrome clínico, caracterizado por una serie de síntomas clínicos, además no requiere una etiología específica y se

reconocen la posibilidad de factores biológicos como psicológicos. Los síntomas de los episodios depresivos de la CIE 10 es la misma que el DSM V, con la diferencia de que se agrega: un síntoma, que es la pérdida de confianza o autoestima; la consideración de la cantidad de síntomas dependiendo de la gravedad del episodio; y un síntoma como elegible, que es la fatiga. En ambos casos la duración mínima es de 2 semanas (Paykel, 2008).

Tabla 01. Criterios diagnósticos sobre la depresión según el CIE-10 Y DSM.

	CIE 10	DSM V
SÍNTOMAS ESENCIALES	1. Pérdida de la capacidad de disfrutar e interés de las cosas.	1. Estado de ánimo depresivo la mayor parte del día.
	2. Estado de ánimo depresivo.	2. Disminución de la capacidad o interés para el disfrute de las actividades diarias.
	3. Disminución de la vitalidad que lleva a una reducción del nivel de actividad, cansancio exagerado y fatigabilidad.	
SÍNTOMAS SECUNDARIOS	4. Disminución de la atención	3. Pérdida o aumento de peso, sin la necesidad de seguir un protocolo de

	dieta, así como un aumento o pérdida de apetito.
5. Sentimiento de inferioridad y pérdida de confianza en sí mismo	4. Hipersomnias o insomnio
6. Ideas de ser inútil o culpa	5. Agitación o retraso psicomotor (observable por un tercero)
7. Percepción sombría del futuro	6. Pérdida de energía o fatigabilidad
8. Autoagresiones con actos suicidas o pensamientos	7. Sentimiento excesivo de culpa o inutilidad
9. Trastornos de sueño	8. Disminución de la atención y la toma de decisiones.
10. Pérdida de apetito	9. Pensamientos recurrentes de muerte o ideación suicida sin especificar un plan

Fuente: OMS (2017) y APA (1995)

1.1.1. Salud mental en hospitales generales o centro de salud

Como se sabe los trastornos psiquiátricos son los que ocupan el primer lugar dentro de los indicadores de carga de enfermedad, por ejemplo, la depresión, que además es la que mayor porcentaje de prevalencia tiene en pacientes con patologías no transmisibles, también sigue siendo subdiagnosticada y con brechas amplias de atención, a pesar del enfoque de atención primaria (INS, 2018).

Asimismo, en el Perú el 85% de centros de salud pertenecen al estado y la gran mayoría organiza su atención a los pacientes por orden de llegada. El motivo de esta forma de atención se debe a que el personal de salud labora 6 horas al día, de los cuales se dividen en 4 horas para atención a pacientes y 2 para labores administrativas o asistenciales.

Al mismo tiempo, existe una directiva del MINSA que recomienda atender a los pacientes en 20 minutos, por lo que al día se atendería a 12 pacientes por cada consultorio. El gran problema es que el personal de salud prefiere tener a disposición 12 pacientes para atenderlos lo más rápido y así cumplir con sus objetivos del día, sin embargo, esto genera incomodidad en los pacientes y por ende la calidad de atención (Pacheco, 2018). En este sentido los instrumentos de screening podrían mejorar la calidad de los diagnósticos en las atenciones.

1.1.2. Modelo Teórico del PHQ-9

Según el DSM IV, Las características principales de la depresión son el ánimo deprimido o pérdida de interés o placer para realizar las actividades en un periodo de al menos 2 semanas. Los pacientes también pueden presentar otros 4 síntomas de una lista que incluyen cambios en el apetito o peso, alteraciones del

sueño y de la actividad motora, disminución de la energía, sentimientos de infravaloración o culpa, desconcentración o incapacidad para tomar decisiones y finalmente pensamientos de muerte o ideación suicida. La frecuencia debe presentarse durante la mayor parte del día, casi todos los días, durante al menos 2 semanas consecutivas (APA, 1995, como se citó en De la Flor, 2020).

Los criterios son:

A1. El estado de ánimo es descrito por el paciente como deprimido, triste, desesperanzado, desanimado o “como en un pozo”. En algunos casos la tristeza puede ser negada, en otros suele presentarse como irritabilidad o alguna molestia somática.

A2. Generalmente se presenta la pérdida de interés y capacidad para el disfrute. Las personas pueden encontrarse menos interesados en sus aficiones o dejar de disfrutar las actividades que antes consideraban placenteras.

A3. Normalmente el apetito disminuye, por lo tanto, el paciente tiene que esforzarse para comer. Sin embargo, también puede presentar aumento del apetito, con preferencias a ciertas comidas (dulces u otros hidratos de carbono). Como consecuencia puede presentarse pérdida o ganancia de peso.

A4. La alteración de sueño suele presentarse como insomnio medio, que es la dificultad para volver a dormir cuando se despierta por la noche, o el insomnio tardío que se presenta como pocas horas de sueños durante la noche. Por otra parte, es menos frecuente el aumento de sueño (hipersomnia) diurno o por tiempos prolongados.

A5. Los cambios psicomotores pueden presentarse como agitación (incapacidad para mantenerse quieto) o enlentecimiento (lenguaje, pensamiento y movimientos enlentecidos). Estos síntomas tienen que ser lo bastante graves como para ser observados por los demás.

A6. Es habitual la falta de energía, el cansancio y la fatiga, esto se presenta incluso cuando la persona no realiza ejercicios que demandan alto desgaste energético o realice mínimos trabajos.

A7. El sentimiento de inutilidad o de culpa puede presentarse como valoraciones negativas y no realistas del propio valor o preocupaciones en exceso de pequeños eventos. Las personas suelen interpretar erróneamente acontecimientos neutros o triviales, asimismo pueden presentar un alto grado de responsabilidad por las adversidades.

A8. Muchos de los pacientes presentan una disminución para pensar, concentrarse o tomar decisiones. A nivel conductual suelen quejarse de la falta de memoria o de distraerse con facilidad. Como consecuencia son incapaces de funcionar adecuadamente en sus labores.

A9. Son frecuentes las ideas suicidas o las tentativas de muerte, estas ideas varían desde la creencia consiente de que los demás estarían mejor si ellos mueren, hasta los pensamientos transitorios del hecho de suicidarse o realizar planes para comértelos. La frecuencia, letalidad e intensidad pueden ser muy variables.

1.1.3. Modelos Bio-Psico-Social sobre la salud y enfermedad

El modelo biopsicosocial pone de relieve la influencia de factores psicosociales y sociales en la salud de la población. En general los factores de riesgo

para los desórdenes mentales están relacionados con la desigualdad social, por ejemplo, la pobreza. Asimismo, se conoce que las experiencias de vida y las condiciones en las primeras etapas de la personalidad son relevantes en la determinación de riesgos para enfermar, es así que el 50% de los problemas de salud que afectan a las personas adultas comienzan antes de los 14 años. Por otro lado, los determinantes sociales también se presentan con las creencias, prejuicios y estigmatizaciones sobre los tratamientos y actitudes respecto a los servicios de salud (INS, 2018).

1.1.4. Diferencias según el sexo y la edad

Considerando la etapa evolutiva de los seres humanos, se ha identificado que es en la etapa de la adolescencia cuando se inicia el desarrollo de los síntomas depresivos (Huberty, 2012). Es por ello que, las personas de esta etapa tienen mayor vulnerabilidad y se encuentran más propensas a sufrir esta enfermedad mental (OMS, 2014).

Según la National Institute of Mental Health NIH (2016), la depresión afecta de manera distinta a las personas, por ejemplo:

- Las mujeres sufren de depresión con más frecuencia a comparación de los hombres. Esto depende del ciclo de vida y hormonal que ellas experimentan. Además, las características más comunes son la tristeza, la falta de autoestima y los sentimientos de culpa.
- Los hombres no suelen reconocer la depresión, por lo tanto, no buscan ayuda. En cuanto a las características de la depresión en los hombres, ellos experimentan cansancio, irritabilidad, y a menudo, enojo, además pierden

el interés por las actividades que disfrutaban y por las responsabilidades del trabajo. Por otra parte, puede ser un factor de riesgo, ya que pueden llegar a consumir drogas o alcohol, como una manera de disminuir los síntomas.

- Las personas adultas mayores pueden experimentar síntomas menos evidentes o ser menos propensos a aceptar que poseen sentimientos de tristeza o pena. Además, son más propensos a presentar trastornos médicos, como la hipertensión, diabetes, entre otras, esto puede contribuir o causar la depresión.
- Los adolescentes que experimentan depresión pueden estar malhumorados e irritables o negarse asistir a sus centros de estudios, además puede presentar comorbilidad con la ansiedad, trastornos de alimentación o drogadicción.
- Los niños con depresión suelen negarse a asistir a las escuelas o fingir estar enfermos, aferrándose a sus padres o tener una preocupación excesiva en que uno de sus padres pueda fallecer.

1.2. PROPIEDADES PSICOMÉTRICAS

Las propiedades psicométricas están reguladas por una serie de guías producidas por la American Psychological Association (APA), la American Educational Research Association (AERA) y el National Council of Measurements in Education (NCME).

En la historia de la validez, desde 1974 hasta 1999, se identificaron 3 fases de validación (Elosua, 2003). Para elaborar esta investigación se tomó en cuenta la última edición de los Standards publicada en el 2014 (American Educational

Research Association, American Psychological Association, y National Council on Measurement Education, 2014).

En primer lugar, se encuentra la fase operacional, dominada por el concepto de validez extrema y pragmatismo, donde se resaltan las correlaciones de la prueba con otras variables para determinar la validez de un test. Se cuenta con tres tipos de validez: contenido, constructo y criterio, esta última engloba la validez concurrente y predictiva.

En segundo lugar, se encuentra la fase teórica, el cual se caracteriza por evaluar la teoría psicológica que fundamenta la prueba. por lo cual el análisis de validez de constructo se convierte en el aspecto principal, antes que la validez de contenido y de criterio.

Finalmente, la fase contextual, caracterizada por generar las puntuaciones, integrando el concepto y la interpretación.

1.2.1 Validez

La validez consiste en juntar evidencia relevante para proveer sustento científico que respalde la interpretación de las puntuaciones de un test que evalúa una variable en específico (American Educational Research Association et al. 2014). En esta investigación, debido a la naturaleza de los datos, se evaluará la validez del constructo.

1.2.2 Validez de contenido

La validez del contenido se define como el grado en el cual el propósito del test es congruente con el contenido del instrumento psicométrico (American

Educational Research Association et al., 2014; Sireci y Faulkner-Bond, 2014). De esa manera, la validez de contenido comprende cuatro elementos de la calidad de un test: Representación del dominio, definición del dominio, relevancia del dominio, adecuación del proceso de desarrollo del test. Por dominio se entiende como la definición operacional mediante la cual se mide el constructo teórico y el rango que abarca la variable. Esto se realiza con una revisión sistemática y minuciosa de la literatura preexistente y mediante el juicio de expertos (Sireci y Faulkner-Bond, 2014).

1.2.3 Validez relacionado al constructo

La validez basada en el constructo se refiere a la estructura interna del test, es decir se define el grado en que los ítems y componentes se relacionan con el constructo (American Educational Research Association et al., 2014). Existen diferentes métodos con relación a la obtención de la validez de este tipo, no obstante, el más utilizado es el análisis factorial. Según Hernandez (2014) este método muestra el grado en que las categorías se combinan con determinado número de dimensiones.

El análisis factorial AF busca hallar constructos ocultos, cuya presencia se acepte para mostrar lógica con alguna teoría que sustente la variable. El AF tiene dos grandes categorías: El Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC), en la primera se busca descubrir la composición profunda de un conjunto de elementos y en la segunda sólo comprueba los factores previamente establecidos en la teoría (López y Gutiérrez, 2019).

Previamente a la evaluación factorial se necesitan los permisos estadísticos, es decir el grado de adecuación de las variables, para ello se utiliza el índice de Kaiser-Meyer Olkin (KMO) (Lloret, Ferrer, Hernández y Tomas, 2014). Estos mismos autores encontraron que Kaiser propuso una matriz con valores para una mejor interpretación, estos son: puntajes inferiores a 50 son considerados inadecuados, valores entre 60 y 69 se consideran mediocres y superior a 80 satisfactorio para el análisis factorial.

Para el análisis factorial confirmatorio se necesitan estadísticos de ajuste para hallar la adecuación del modelo. Así tenemos el índice de bondad de ajuste (GFI), el cual evalúa si el modelo debe ser ajustado, entre mas cercano a cero se encuentra, representa un mayor desajuste. Asimismo, el error de aproximación cuadrático medio (RMSEA), representa el ajuste anticipado con el valor total de la población y no con la muestra. Cuando el RMSEA es menor o igual a 0.05 indica error de aproximación del modelo con la realidad. Dentro de este mismo orden de ideas, tenemos el índice No normalizado de ajuste (NNFI) o índice de Tucker Lewis, tiene mas parámetros que el NFI, ya que considera los grados de libertad del modelo propuesto. Los valores van de 0 a 1, siendo recomendable valores mayores a 0.9. Por otra parte, tenemos el índice normado de ajuste (NFI), este índice compara el modelo nulo con el modelo propuesto, el valor aceptable es mayor a 0.9. (Escobedo, Hernandez, Estebané y Martinez. 2016)

Por otro lado, las medidas de ajuste de la parsimonia están relacionadas a la calidad de ajuste de los modelos con la cantidad de coeficientes estimados. Dentro de este grupo se tiene el criterio de información de Akaike (AIC) el cual es una medida que compara los modelos con diferentes números de constructos. Cuanto

mas cercano a cero se encuentra indica mayor ajuste y una mayor parsimonia (Escobedo, Hernandez, Estebané y Martinez. 2016).

1.2.4. Confiabilidad

En cualquier proceso de investigación y ante la gran cantidad de fuentes de potenciales de error, es recomendable que los responsables de las investigaciones proporcionen una mayor confianza en los resultados y conclusiones. Es decir, un instrumento es confiable reproducible o preciso, cuando las mediciones generan el mismo resultado en diferentes escenarios, poblaciones y no cambia en el tiempo, siempre y cuando se aplica en las mismas condiciones (Manterola, Grande, Otzen, Garcia, Salazar y Quiroz. 2018).

El estadístico mas conocido cuando se aborda la confiabilidad es el Alfa de Cronbach, por ser el más utilizado en estudio psicométricos. Este estadístico es una estimación de consistencia interna que indica la covarianza y en que medida el constructo esta presente en los ítems. A pesar de ser el más empleado el Alfa de Cronbach tiene limitaciones en su cálculo, por ejemplo, se ve afectado por el número de ítems, la cantidad de alternativas en las respuestas y la proporción de la varianza del test. De esta manera, según Lozano, Garcia y Muñiz (2008) demostraron que la disminución de las alternativas en las respuestas, afectaba en la variabilidad de la escala, por ende, el coeficiente alfa se alteraba. Generalmente, este estadístico trabaja con variables continuas, la cual no se emplea en las ciencias sociales, por lo que también minusvalora la confiabilidad (Ventura-Leon y Caycho-Rodriguez. 2017)

Una alternativa frente a esta limitación es el coeficiente omega, la diferencia es que este estadístico utiliza las cargas factoriales, los cuales no son más que las sumas ponderadas de las variables estandarizadas. La ventaja es que esos cálculos son más estables y refleja el verdadero nivel de fiabilidad. Por otra parte, es una alternativa cuando las cargas factoriales no cumplen con el principio de equivalencia, es decir cuando sus valores son muy diferentes. El valor aceptable para el coeficiente omega oscila entre 0.70 y 0.90, aunque en otras circunstancias pueden ser validos valores superiores a 0.65 (Ventura-Leon y Caycho-Rodriguez. 2017).

1.2.5 Análisis de ítems

Determina el grado que cada ítem tiene para medir el constructo del instrumento (Abad, 2006). Así tenemos, el índice de homogeneidad que distingue si los reactivos evalúan de forma igual un constructo (Rodero, Molina, Fernández y Redondo, 2016); el índice de discriminación que distingue o diferencia a las personas de baja, media y alta puntuación, y por el último, el índice de dificultad que representa si lo ítems son difíciles para los examinados (Hurtado, 2018). Sin embargo, este último solo se aplica para pruebas de rendimiento, sobre todo en el campo educativo (Pérez-Sala, Sirlopú, Cobo, Awad. 2019).

Por otra parte, los valores aceptables del índice de homogeneidad oscilan entre 0.50 y 0.71 (Salazar y Del Castillo, 2018). No obstante, algunos autores como Likert 1993(como se citó en Domínguez, (2013); Ebel 1965 (como se citó en Rodriguez, 2014) y Kline 1999 indican que valores mayores o iguales a 0,20 y 0.30 son adecuados para hallar la homogeneidad. Asimismo, los valores permitidos para las comunalidades deben ser mayores o iguales a 0.50, considerándose este puntaje

como el mínimo requerido (Hair, Anderson, Tatham y Black, 2004). Por último, el índice de discriminación no maneja ningún valor numérico, la obtención del índice se realiza mediante el estadístico T de Student, el cual indica que si la significancia es menor a 0.05, los ítems pueden diferenciar los grupos de comparación. Pero si la significancia es mayor a 0.05, los ítems no encuentran diferencias entre los grupos (Molina, Ochoa y Ortega. 2020)

1.2.6. Baremos

La función de los baremos en las pruebas psicométricas es transformar los puntajes brutos o directos en puntuaciones derivadas como percentiles, cuartiles, entre otras. En otras palabras, permite ubicar al individuo en una posición con relación a un grupo. Además, los baremos son necesarios para todo instrumento que se construya o adapte (Chan y Quezada. 2013).

Las puntuaciones del test pueden ser interpretadas de 2 formas. La primera es según la norma, el cual permite comparar la puntuación de una persona en relación a un grupo que desarrolló la misma prueba. La segunda es según el criterio, esta forma de interpretación toma como referencia la definición de un dominio, y permite medir la ejecución del individuo ya no en comparación de un grupo, sino en función de un criterio establecido por un dominio (Meneses, Barrios, Bonillo, Cosculluela, lozano, Turbany y Valero. 2013).

Dentro de las interpretaciones normativas, una vez obtenido los puntajes directos de una persona en un test, se suelen transformar en otros tipos de puntuaciones con el objetivo de facilitar la comprensión para los evaluadores. Uno de ellos es los percentiles, el cual consiste en asignar a cada puntuación directa el

porcentaje de individuos que obtienen puntuaciones inferiores a ella. De esta manera, se da a conocer de manera rápida e intuitiva la posición relativa de una persona en el grupo. Por otra parte, se tiene las puntuaciones típicas, el cual consiste en transformar las puntuaciones directas en otras llamadas “típicas” (Z), se realiza mediante una conversión lineal, consistente en restarle la media y dividir por la desviación típica. Es decir, se puede saber si una persona se encuentra por debajo o por encima de la media y en que grado se separa de ella (Muñiz, 2018).

1.2.7 Modelo Unidimensional

La unidimensionalidad implica que un conjunto de ítems pertenezca a un solo factor latente o constructo, es decir un instrumento de evaluación psicométrica será unidimensional si las respuestas están basadas en un único atributo (Burga, 2006). En términos psicométricos se entiende como la varianza común es explicada por un solo factor (Ferrando y Anguiano, 2010).

Según Béjar (como se cito en Burga, 2010) indica que la unidimensionalidad no implica que necesariamente las respuestas a los ítems se deban a un solo proceso psicológico, sino que puede existir un amplio conjunto de procesos psicológicos pertenecientes en dichas respuestas, pero en la medida que los mismos procesos afecten de la misma manera las respuestas a un conjunto de ítems, de esta manera se podrá inferir que existe unidimensionalidad intrínsecamente en dicho instrumento de evaluación.

Asimismo, indicó que las puntuaciones obtenidas, según la Teoría Clásica de los Test siguen un modelo monotónico lineal, es decir existe una relación lineal entre el puntaje directo y el rasgo o atributo que se está midiendo.

1.2.8 Modelo Bifactorial y Bidimensional

Ambos modelos pueden sugerir que son similares en cuanto a su estructura, sin embargo, son totalmente diferentes, ya que el modelo o estructura bifactorial se utiliza para representar empíricamente los modelos teóricos y de medición con el supuesto de que las dimensiones se explican mejor por la presencia de un factor G (Reise, 2012, como se citó en Lapuente, 2018), es decir que existe un factor general independientemente de las dimensiones.

Por otra parte, los modelos bidimensionales explican directamente la variación de los ítems, es decir se basa en el análisis de primer orden. Asimismo, cada dimensión debe medirse como una combinación formativa, aunque cada dimensión puede ser medido por separado. Además, los modelos bidimensionales tienen menos puntajes de error y logran una medición más precisa (Alper y Demirtasli. 2012).

2. ANTECEDENTES

2.1 INVESTIGACIONES NACIONALES

En Perú, Calderon et al. (2012), hicieron un estudio de validación de la versión peruana del PHQ-9 para diagnosticar la depresión de los pacientes que asisten al primer nivel de atención de salud. Específicamente se realizó la validación de contenido por juicio de expertos, para ello participaron 23 médicos psiquiatras, tres psicólogos y una enfermera. Se realizaron sugerencias y modificaciones que mejorarían la versión peruana de la PHQ-9. Se obtuvo como conclusión que el

instrumento es válido para diagnóstico de depresión, y que puede ser empleado con facilidad en el ámbito de atención primaria de la salud en el Perú.

Asimismo, en el Perú, Zhong et al. (2014) estudiaron la validez de la versión española del PHQ-9 empleando 1520 participantes mujeres embarazadas, comprendidas entre las edades de 18 y 49 años, y con 16 semanas de gestación. Según el análisis de componentes principales de los residuos se determinó que los datos se ajustan a un modelo de un solo factor, además se identificó independencia local evaluando las correlaciones inter - ítem ($r < .3$). Por otra parte, se obtuvo una adecuada consistencia interna ($PSI = 9.72$). Finalmente, se obtuvo que la invarianza de medida no fue significativa, debido a que dos ítems presentaron diferencias de acuerdo al nivel educativo de las personas participantes.

Además, De la Flor (2020) realizó una investigación para evaluar la evidencia de validez de las puntuaciones de la versión peruana del PHQ-9 para tamizaje de depresión en adultos con cáncer de un hospital de atención especializada de Lima metropolitana. El estudio se llevó a cabo con 500 personas y se obtuvo como resultado la evidencia basada en el contenido ($V > 0.8$), además se identificó una estructura interna simple de dos factores que explican el 47% de la varianza con una adecuada bondad de ajuste ($CFI = .942$; $TLI = .92$; $AGFI = .946$; $SRMR = .044$; $RMSEA = .058$), donde se presentaron una moderada dimensión somática ($\omega = .69$) y una adecuada confiabilidad de la dimensión Cognitiva/afectiva ($\omega_1 = .76$; $\omega_2 = .72$). Finalmente se encontró que el adecuado punto de corte es mayor o igual a 8 ($Se = .87$; $Sp = .77$) y su aplicación puede ser con un algoritmo ($Se = .70$; $Sp = .86$).

Por otra parte, también se han evaluado diferentes modelos factoriales del Cuestionario de salud del paciente. Es así que, Villarreal, et al. (2019), en su trabajo de investigación evaluaron diferentes modelos, unidimensional y bidimensional, este último tiene diferentes versiones en cuanto a la composición de las dimensiones. Sin embargo, debido a que obtuvieron mejores puntajes de la bondad de ajuste eligieron el modelo unidimensional, con un adecuado nivel de confiabilidad ($\omega=0.87$). Cabe resaltar que las dimensiones obtenidas en los modelos estructurales de dos dimensiones se correlacionaban con un nivel muy alto.

Por su parte Huarcaya, et al. (2020) en su trabajo de investigación tuvo como objetivo evaluar las propiedades psicométricas y la estructura factorial del PHQ-9, concluyendo que el análisis factorial confirmatorio reveló que el modelo bifactorial, el cual toma un factor general y dos factores latentes, muestra un mejor ajuste para el PHQ-9.

2.2 INVESTIGACIONES INTERNACIONALES

También se emplearon el PHQ-9 para realizar investigaciones en Sudamérica.

En Chile se realizó un estudio sobre la validez y utilidad del PHQ-9 para el diagnóstico de depresión, para ello se trabajó con 1327 pacientes asistentes al primer nivel de atención. Según un análisis de componentes principales se determinó un modelo unidimensional que explica 47% de la varianza total, además se obtuvo una confiabilidad por consistencia interna aceptable ($\alpha= .835$). Finalmente, se encontró una alta superposición de los resultados categóricos del

DHS, PHQ-9 Y CIE-10, evidenciando una elevada precisión diagnóstica (Baader et al., 2012).

En Ecuador, Fabara (2013) estudió la validez del PHQ-9 y el GHQ-12 para la evaluación de depresión, se empleó 189 personas como muestra de la investigación y como resultado se obtuvo una adecuada precisión considerando un punto de corte ≥ 10 (Se=84.2%; Es=91.2%; LR+=9.53; LR-=.17).

De la misma manera en Argentina se estudió la utilidad del PHQ-9 con una muestra de 152 pacientes. Se evidenció la validez convergente, mediante el HADS Y PHQ-9, además se evidenció la validez de criterio concurrente mediante el índice de Barthel y la presencia de depresión evaluada por el PHQ-9. Finalmente, se encontró una adecuada precisión diagnóstica (Montenegro et al., 2017).

Finalmente, en Colombia, Cassiani-Miranda et al. (2017) evaluaron la confiabilidad y las dimensiones del PHQ-9 para el diagnóstico de sintomatología depresiva en estudiantes de la carrera de ciencias de la salud, por lo que contaron con 550 participantes de 20 años como edad media. Se obtuvo como resultado, mediante un AFE, dos dimensiones que explicaban en un 42.8% la varianza total, y se obtuvo una alta confiabilidad mediante el estadístico alfa de Cronbach ($\alpha=.83$) y omega de McDonald ($\omega=.89$), evidenciando la validez de constructo, sin embargo, el autor resalta que los modelos bidimensionales o unidimensionales pueden variar según la naturaleza del tipo de población.

3. DEFINICIONES CONCEPTUALES Y OPERACIONALES DE LAS VARIABLES

Tabla 2. Definiciones conceptuales y operacionales

Variables	Definición conceptual	Definición operacional	Indicadores	Ítem
Depresión	Conjunto de síntomas caracterizado fundamentalmente por el humor depresivo, pérdida de la capacidad de interesarse y disfrutar de las cosas	<p>A. ¿Pocas ganas o interés de hacer las cosas?</p> <p>B. ¿Sentirse demasiado(o), deprimida(o), triste o sin esperanza?</p> <p>C. ¿Problemas para dormir o mantenerse dormida(o), o en dormir demasiado?</p> <p>D. ¿Sentirse cansada(o) o tener poca energía sin motivo que lo justifique?</p> <p>E. ¿Poco apetito o comer en exceso?</p> <p>F. ¿Dificultad para poner atención o concentrarse en las cosas que hace?</p>	<ul style="list-style-type: none"> ● Poco interés ● Tristeza ● Sueño excesivo ● Cansancio ● Falta de apetito ● Dificultad para concentrarse ● Movimientos lentos 	9 ítems

-
- G. ¿Moverse más lento o hablar más lento de lo normal o sentirse más inquieta(o) o intranquila(o) de lo normal?
- H. ¿Pensamientos de que sería mejor estar muerta(o) o que quisiera hacerse daño de alguna forma buscando morir?
- I. ¿Sentirse mal acerca de si misma(o) o sentirse que es una(un) fracasada(o) o que se ha fallado a sí misma(o) o a su familia?
-

HIPÓTESIS

HIPÓTESIS GENERAL

- Existen adecuadas propiedades psicométricas de la versión peruana del cuestionario de salud del paciente PHQ-9.

HIPÓTESIS ESPECÍFICOS

- Existe calidad de los ítems mediante los índices de dificultad, discriminación y homogeneidad de la versión peruana del cuestionario de salud del paciente PHQ-9.
- Existe validez de constructo la validez factorial de la versión peruana del cuestionario de salud del paciente PHQ-9.
- Existe confiabilidad por consistencia interna y método de dos mitades de la versión peruana del cuestionario de salud del paciente PHQ-9.

CAPÍTULO III: METODOLOGÍA

1. NIVEL Y TIPO DE INVESTIGACIÓN

El nivel y tipo de la investigación es básico, no experimental de corte transversal (Hernández, Fernández y Baptista, 2014)

2. DISEÑO DE LA INVESTIGACIÓN

Esta investigación es de diseño instrumental, pues está orientado al análisis de las propiedades psicométricas de instrumentos de medida del área psicológica (Ato, López & Benavente, 2013)

3. NATURALEZA DE LA MUESTRA

3.1. DESCRIPCIÓN DE LA POBLACIÓN

La muestra de estudio será representativa en función de la población peruana, conformada por hombres y mujeres, de entornos urbanos y rurales de origen multiétnico y diferente nivel socioeconómico. Se incluirán datos desde los 15 años de las personas que hayan pernoctado la noche anterior, a la encuesta, en la vivienda seleccionada, pero excluyendo a las personas que tenían datos faltantes en las variables de interés (sexo, edad, nivel de educación, personas con hipertensión, diabetes, nivel socioeconómico y los nueve ítems del PHQ). Inicialmente había 34.971 participantes; sin embargo, después de excluir a los participantes con información incompleta, se analizarán los datos de 33 613 participantes.

3.2. MUESTRA Y MÉTODO DE MUESTREO

Se recurrirá a un análisis de datos secundarios, tomando en cuenta la Encuesta de Demografía y Salud del Perú, una encuesta representativa que se realiza anualmente a nivel nacional.

El diseño de la ENDES incluye una técnica de muestreo caracterizado por ser probabilística de tipo equilibrado, bietápica, estratificada e independiente, a nivel de departamentos y por áreas rurales y urbanas. El tamaño de la muestra es de 36760 viviendas, correspondiendo:

- 14780 capitales de departamentos y 43 distritos de la provincia de Lima.
- 9320 viviendas del área urbana.
- 12 660 viviendas del área rural.

Sin embargo, para los resultados de la investigación se tomará en cuenta solo los datos completos, los cuales son 33613.

3.3. CRITERIOS DE INCLUSIÓN

La encuesta de la ENDES 2019 incluyó:

- Los residentes habituales de viviendas particulares de áreas urbanas y rurales del país que hayan pasado la noche anterior en la vivienda seleccionada.

Para la presente investigación se incluyó:

- Personas de género masculino o femenino con datos completos en el cuestionario PHQ-9

- Personas de género masculino o femenino que hayan brindado su consentimiento informado en la encuesta.
- Personas de 15 años a más.

3.4 CRITERIOS DE EXCLUSIÓN

Según la ficha técnica de la encuesta de la ENDES 2019 se excluirán:

- Los residentes habituales que no hayan pasado la noche anterior a la encuesta en la vivienda seleccionada
- Visitantes que pasaron la noche anterior a la encuesta en la vivienda elegida.

4. INSTRUMENTO

El Cuestionario sobre la Salud del Paciente (PHQ-9) fue originalmente desarrollado por Spitzer, Kroenke, and Williams (1999). Es una escala clínica que mide la depresión. Puede ser aplicado individual o colectivamente, a adultos mayores de 17 años, con un tiempo de aplicación aproximado de 5 minutos. Los materiales necesarios son el cuestionario, lápiz y papel.

El PHQ-9 es un instrumento unidimensional que evalúa la depresión de acuerdo a los criterios del DSM-IV mediante 9 ítem con formato de respuesta Likert del 0 (“nunca”) al 3 (“casi todos los días”) que indica la severidad de la sintomatología depresiva, y provee una puntuación total cuantitativa de intervalo del 0 hasta 27 (Kroenke, Spitzer, y Williams, 2001; Kroenke y Spitzer, 2002).

5. PROCEDIMIENTO

El presente estudio psicométrico se desarrolló con fuente de datos secundarias de acceso libre, por consiguiente, no se aplicó el instrumento directamente.

1. En primer lugar, se realizó la búsqueda de fuentes de datos libres para conocer las posibles variables.
2. Luego se identificó los ítems y la variable que se desarrollará en la presente investigación.
3. Después se redactó el proyecto de investigación, el cual consistió en el planteamiento del problema, justificación, objetivos, antecedentes, marco teórico y metodología. Para ello se revisó diferentes fuentes bibliográficas.
4. Luego se envió al comité institucional de ética en investigación CIEI de la UPCH para su revisión y aprobación.
5. Posterior a ello y como parte del primer objetivo específico, se comenzó a evaluar los ítems, obteniendo los índices de homogeneidad, discriminación y las comunalidades. Los cuales fueron procesados en el paquete estadístico SPSS.
6. Seguidamente y como parte del segundo objetivo, se procesaron los datos en los modelos que fueron diseñados en investigaciones pasadas. Cabe resaltar que solo se evaluaron los puntajes del modelo más adecuado para el instrumento en la presente investigación, debido a que dichas estructuras factoriales ya se encontraban re-diseñadas. Estos análisis se realizaron con el paquete estadístico AMOS

7. Después se respondió al tercer objetivo específico, el cual trató sobre la confiabilidad de la prueba. Para ello, se obtuvo los estadísticos de confiabilidad de Omega de Mc Donald, Spearman Brown y Guttman. Este análisis se realizó con el SPSS y su extensión Omega.
8. Después se respondió al cuarto objetivo específico, donde se halló las normas percentilares. Previamente, se evaluó las diferencias según género y edad.
9. Posterior a ello, se redactó la discusión en función de los resultados obtenidos.
10. Finalmente, se redactó las conclusiones y recomendaciones.

6. CONSIDERACIONES ÉTICAS

En cuanto a las consideraciones éticas de la base de datos, la encuesta cuenta con un apartado para informar el nombre del entrevistador y de la institución, así como para dar a conocer el objetivo. Además, la ENDES indicó que se respetará la confidencialidad y se mantendrá en reserva el nombre. Cabe resaltar que las encuestas fueron aplicadas por personal debidamente capacitado y la base de datos es de acceso libre, asimismo el tratamiento de los datos fue transparente sin modificar ningún puntaje o resultado.

Por ende, la presente investigación no representa ningún riesgo ético. Finalmente, debido a la naturaleza de la base de datos el comité de ética de la UPCH lo consideró como proyecto exonerado.

7. PLAN DE ANÁLISIS DE DATOS

En primer lugar, se revisó la base de datos para remover y depurar los datos que no corresponden para el análisis estadístico. Seguidamente, se evaluó la normalidad del PHQ-9 mediante el coeficiente de kolmogorov Smirnov. Asimismo, se analizó la calidad de los ítems, mediante los estadísticos: elemento total corregido, comunalidades y T de Student, para los índices de homogeneidad, comunalidades y discriminación respectivamente.

Después se analizó la validez del constructo, para ello se evaluó la adecuación de los datos mediante el test KMO y el test de esfericidad de Bartlett ($p < .05$). Luego se evaluó los modelos unidimensionales, bidimensionales y bifactor mediante el método de extracción de máxima verosimilitud utilizando el análisis factorial confirmatorio. Los estadísticos que se tomaron en cuenta para la evaluación de cada modelo fueron las medidas de ajuste absoluto (Chi cuadrada y RMSEA), las medidas de ajuste incremental (TLI, CFI y NFI) y las medidas de ajuste de la parsimonia (PRATIO, PCFI, PNFI y AIC).

En cuanto a la confiabilidad, se evaluó por el método de consistencia interna, para ello se empleó el estadístico Omega. Finalmente, para la baremación, previamente se halló las diferencias según género y edad mediante el estadístico U de Mann Whitney, luego se continuó hallando los percentiles.

CAPITULO IV: RESULTADOS

En cuanto al primer objetivo específico, se identificó la calidad de los ítems mediante los índices de discriminación, homogeneidad y comunalidades de la versión peruana del cuestionario de salud del paciente PHQ-9, para la cual se emplearon 33613 casos con respuestas completas.

Tabla 3. Índices para el análisis de la calidad de reactivos.

	IHC	h2	Id
Ítem 1: Anhedonia	.656	.542	.000
Ítem 2: Distimia	.715	.623	.000
Ítem 3: Sueño	.646	.533	.000
Ítem 4: Cansancio	.694	.597	.000
Ítem 5: Apetito	.601	.477	.000
Ítem 6: Desconcentración	.635	.527	.000
Ítem 7: Psicomotricidad	.627	.519	.000
Ítem 8: Ideación suicida	.500	.360	.000
Ítem 9: Culpa	.558	.436	.000

En la tabla 3 se puede observar los estadísticos descriptivos para evaluar la calidad de los ítems. Se aprecia que los 9 reactivos presentan adecuado índice de homogeneidad ($>.30$), lo que significa que todos los reactivos miden la misma variable (depresión) de manera selectiva. Por otra parte, los índices de discriminación, obtenidas mediante la T de student, son significativas ($P<0.05$), es decir los reactivos son adecuados para discernir entre los niveles leve, moderado, y severo de depresión. Sin embargo, las comunalidades 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.

Respecto a los resultados del segundo objetivo, se realizó el análisis de la validez factorial confirmatoria, para ello, se evaluó los permisos estadísticos.

Tabla 4. KMO y prueba de esfericidad de Bartlett del PHQ-9.

Estadísticos de prueba	Medida Kaiser-Meyer-Olkin	Bartlett
PHQ-9	0.914	0.000

Los resultados en la tabla 4 indican que la matriz de correlación de los reactivos del Cuestionario de Salud del paciente es ideal para realizar análisis factorial. Se obtuvo un índice KMO favorable (0.914), lo que sugiere que la proporción de varianza explicada por correlaciones parciales es reducida, por ende, la matriz es adecuada. Con respecto al test de esfericidad de Bartlett indica que los reactivos tienen una correlación estadísticamente significativa ($P < 0.05$). Por lo tanto, la prueba cumple con los permisos estadísticos.

Por consiguiente, se emplearon 5 modelos sugeridos por investigaciones previas (Modelo unidimensional, bidimensional y bifactorial) lo cuales son:

Figura 1. Modelo unidimensional

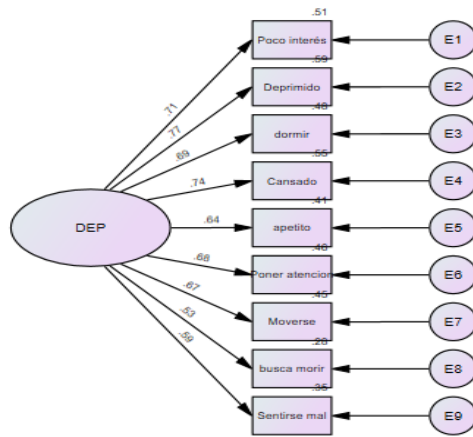


Figura 2. Modelo Bidimensional



Figura 3. Modelo Bidimensional

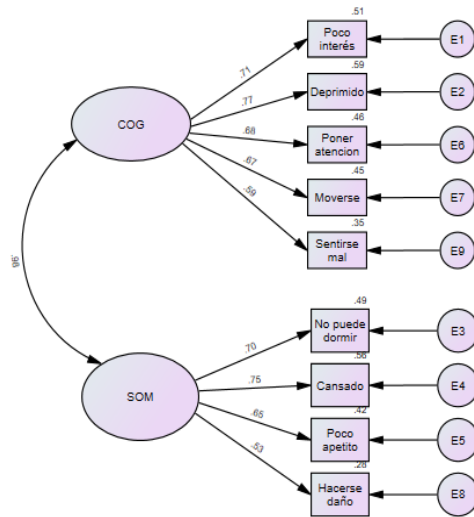


Figura 4. Modelo Bidimensional

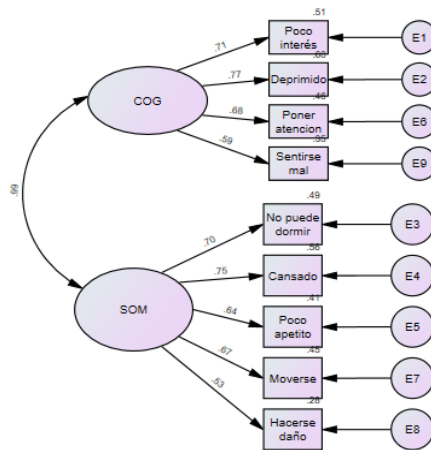


Figura 5. Modelo Bifactorial

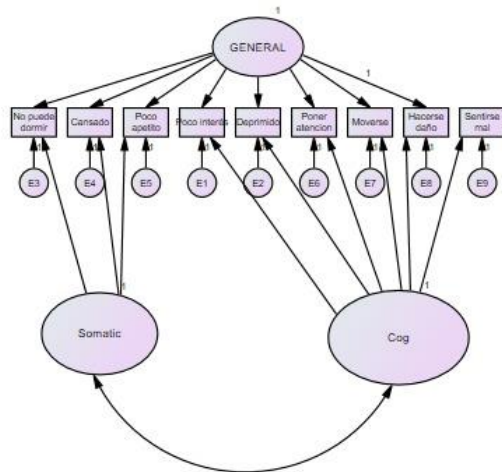


Tabla 5. Índices de bondad de ajustes de los diferentes modelos según el AFC.

	Medida de Ajuste Absoluto		Medidas de Ajuste Incremental			Medida de Ajuste de la Parsimonia			
	Chi-cuadrado	RMSEA	CFI	TLI	NFI	PRATIO	PCFI	PNFI	AIC
Modelo 1	0	0.098	0.929	0.905	0.928	0.750	0.696	0.696	8796.712
Modelo 2	0	0.097	0.932	0.906	0.932	0.722	0.673	0.673	8367.711
Modelo 3	0	1	0.929	0.902	0.929	0.722	0.672	0.671	8736.132
Modelo 4	0	1	0.929	0.901	0.928	0.722	0.671	0.671	8762.992
Modelo 5	0	0.077	0.971	0.942	0.971	0.528	0.512	0.512	3719.817

En la tabla 5 se puede observar los estadísticos de bondad de ajuste de los 5 modelos evaluados, donde la prueba estadística exacta chi-cuadrada χ^2 indicaría que los datos no se ajustan a ningún modelo ($p < .001$). Similarmente, el RMSEA para el modelo 1, 2, 3 y 4 son de 0.098, 0.097, 1 y 1 respectivamente, lo que significa que las medidas de ajuste absolutos presentan un nivel mediocre, no obstante, en el modelo 5 se obtuvo un RMSEA de 0.077, el cual indicaría un mejor ajuste absoluto, sin embargo, no es 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, a comparación de los 4 primeros modelos, el último modelo tiene medidas de Ajuste de la Parsimonia menor, lo que indica un mejor ajuste de los datos.

Para la presente investigación se considera como valido el modelo 2 (bidimensional), debido a que obtuvo, ligeramente, mejores puntajes en la 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 (somático y cognitivo).

Tabla 6. Cargas factoriales del PHQ-9 obtenidas mediante AFC.

	Modelo 1	Modelo 2		Modelo 3		Modelo 4		Modelo 5		
		F1	F2	F1	F2	F1	F2	FG	F1	F2
Ítem1: Anhedonia	0.711	0.714		0.714		0.715		0.214		-0.697
Ítem2: Distimia	0.768	0.773		0.771		0.772		0.258		-0.733
Ítem3: Sueño	0.695		0.713		0.700		0.697	0.212	0.682	
Ítem4: Cansancio	0.744		0.765		0.750		0.747	0.234	0.728	
Ítem5: Apetito	0.643		0.655		0.646		0.644	0.204	0.620	
Ítem6: Desconcentración	0.677	0.680		0.679		0.677		0.233		-0.632
Ítem7: Psicomotricidad	0.669	0.673		0.671			0.670	0.226		-0.627
Ítem8: Ideación suicida	0.528	0.534			0.527		0.527	0.296		-0.424
Ítem9: Culpa	0.591	0.598		0.591		0.591		1.879		0.040

En la tabla 6 se puede observar las cargas factoriales con sus respectivas dimensiones. Se observa que las cargas factoriales de los 4 primeros modelos son idénticas y significativas ($\lambda > 0.50$). De forma particular, sus cargas factoriales de los ítems 1,2,3,4,5,6 y 7 oscilan entre 0.64 y 0.77. Sin embargo, para los ítems 8 y 9, oscilan entre 0.52 y 0.59. No obstante, existe un ligero incremento de las cargas factoriales en los modelos 2, 3 y 4 (bidimensional) respecto al modelo 1 (unidimensional). Por otra parte, en el factor G del modelo 5, las cargas factoriales se encuentran entre 0.204 y 1.879, este último no cumple con el rango de los coeficientes de correlación. Los ítems del factor somático tienen cargas factoriales entre 0.620 y 0.728, lo cual es aceptable. Los ítems del factor cognitivo/afectivo cargan entre -0.424 y -0.733, lo que indica una correlación negativa con el factor, excepto el ítem 9 con 0.040. De cualquier forma, el modelo 5 presenta incoherencias en las cargas factoriales.

Tabla 7. Correlación de las dimensiones de los modelos obtenidos mediante AFC

Modelo	R de Pearson
Modelo1 Unidimensional	No presenta
Modelo 2 Bidimensional	0.952
Modelo 3 Bidimensional	0.985
Modelo 4 Bidimensional	0.991
Modelo 5 Bifactorial	-0.958

En la tabla 7 se observa las correlaciones de las dimensiones, se aprecia que existen correlaciones muy altas ($r > 0.9$). Lo que significaría que las dimensiones se encuentran superpuestas.

Respecto a los resultados del tercer objetivo, se halló la confiabilidad por consistencia interna mediante el estadístico Mc Donald's Omega y por el método de dos mitades, para la cual se emplearon 33613 casos con respuestas completas.

Tabla 8. Estadísticos de confiabilidad Mc Donald's Omega del PHQ-9.

Instrumento	Número de ítems	Mc Donald's Omega
PHQ-9	9	.892

En la tabla 8 se observa la confiabilidad de consistencia interna con un total de 9 ítems mediante el estadístico Omega, obteniendo un puntaje alto ($\omega = .892$). Lo que supone una buena fiabilidad del modelo y buena correlación entre variables, el cual incorpora 9 reactivos.

Tabla 9. Confiabilidad por el método de dos mitades.

Spearman-Brown		Guttman
Misma longitud	Longitud desigual	
.823	.824	.753

En la tabla 9 se puede apreciar que la confiabilidad por el método de dos mitades es adecuada (>0.8) en ambas longitudes. Al contar con 9 ítems, la longitud desigual es la más apropiada para esta investigación. Similarmente el estadístico de Guttman es adecuado (>0.6). Por lo que se infiere una adecuada consistencia interna.

Respecto a los resultados del cuarto objetivo, se halló los baremos con normas percentilares. Previamente se comparó los grupos según sexo y edad, para la cual se emplearon 33613 casos con respuestas completas

Tabla 10. U de Mann Withney según género.

Género	N	Rango promedio	Sig. asintótica (bilateral)	Z
Hombre	14301	15409.99	.000	-24.254
Mujer	19312	17841.52		
Total	33613			

En la tabla 10 se puede observar la diferencia de medias no paramétricas mediante el estadístico U de Mann Whitney, donde se obtuvo una significancia bilateral de 0.000 menor a 0.05, por lo que se rechaza la hipótesis nula, aceptándose la hipótesis alterna, es decir existen diferencias entre los puntajes según el género.

Tabla 11. U de Mann Whitney según edad.

EDAD	N	Sig. asintótica (bilateral)	Z	Suma de rango
Adolescente	3560	.398	-0.846	50850983.00
Adultos	24806			351478178.00
Adolescente	3560	.000	-19.201	13513930.50
Adultos	5247			25272097.50
Mayores	24806	.000	-28.430	357550054,00
Adultos	5247			94056377,00
Mayores				

En la tabla 11 se puede observar la diferencia de medias no paramétricas mediante el estadístico U de Mann Whitney, donde se obtuvo una significancia

bilateral de 0.00 menor a 0.05, por lo que se rechaza la hipótesis nula para los adultos mayores con los adolescentes y adultos, aceptándose la hipótesis alterna, es decir existen diferencia de los puntajes entre los adultos mayores con los adolescentes y adultos. Sin embargo, se obtuvo una significancia bilateral de 0.398 mayor a 0.05, por lo que se rechaza la hipótesis alterna para los adolescentes y adultos, y se acepta la hipótesis nula, es decir no existen diferencias de los puntajes entre la adolescencia y adultez. En consecuencia, los datos sugieren realizar baremos para adolescentes y adultos como uno solo y para los adultos mayores por separado.

Tabla 12. Baremos del PHQ-9 para la población peruana.

Niveles	MUJERES		VARONES		PC
	Adolescentes/ Adultas	Adultas Mayores	Adolescentes/ adultos	Adultos Mayores	
Sin síntomas relevantes	0 - 3	0 - 7	0-2	0 - 5	1-74
Leve	4	8 - 9	3	6	75
Moderado	5 - 6	10 - 11	4	7 - 8	81
Moderadamente severo	7 - 8	12 - 14	5	9 - 11	87
severo	>9	> 15	>6	>12	93

En la tabla 12 se pueden observar los puntajes de los baremos, donde existen 5 niveles con percentiles 15, 33, 50 y 70, los niveles van desde ningún síntoma hasta síntomas severos de la depresión. Al ser una variable clínica se tomó en cuenta a la muestra perteneciente al Pc 75.

CAPITULO V: DISCUSIÓN

Debido a que no se analizó la validez relacionada al contenido es necesario aclarar que la prueba ya cuenta con dicha validez en el contexto peruano, el cual fue realizado por el Ministerio de Salud (MINSA) y el Instituto Nacional de Salud (INS) dentro del marco de desarrollo de algoritmos de diagnóstico y tratamiento de Depresión (Calderón et al. 2021). Los cambios sugeridos fueron empleados para la recolección de datos realizada por la ENDES 2019.

Por lo tanto, a partir de los hallazgos encontrados, se acepta la primera hipótesis específica que establece que existen buena calidad de los ítems de la versión peruana del cuestionario de salud del paciente PHQ-9. Para ello, se analizó el índice de homogeneidad a través de la correlación de Pearson entre el ítem y el total corregido, donde se observan valores aceptables de .50 a .71 (Salazar y Del Castillo, 2018), además estos resultados concuerdan con otros criterios de verificación, es así que, según Likert 1993 (como se citó en Domínguez, (2013) indica que el puntaje mínimo requerido es.20. Asimismo, autores como Ebel 1965 (como se citó en Rodríguez, 2014) y Kline 1999 indican que los puntajes mayores a .30 son adecuados para hallar la homogeneidad. Esto refleja que los puntajes se adecuan perfectamente, lo que significa 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 comunalidades, los cuales se encuentran entre 0.519 y 0.623 (51.9% y 62.3%). Estos resultados son aceptables y superan el mínimo requerido 0.5 (Hair, Anderson, Tatham y Black, 2004). No obstante, los ítems 5, 8 y 9 explican la varianza entre 0.36 y 0.47.7 (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 correlación obtenidos con el estadístico elemento-total corregido o índice de homogeneidad. Además, las comunalidades en la psicología generalmente son bajas, ya que las medidas como los test y sobre todo los ítems, contienen intrínsecamente muchos errores de medida (Ferrando y Anguiano-Carrasco, 2010). Por consiguiente, se conservan los 9 ítems.

Dentro de este orden de ideas, la capacidad discriminativa de los ítems se calculó mediante el estadístico T de student, donde se obtuvo puntajes significativos ($p < 0.05$), en otras palabras, todos los ítems pueden diferenciar a las personas con niveles leves, moderados y severos de depresión (Molina, Ochoa y Ortega. 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 necesario una alta discriminación (Masters, 1988, como se citó en Hurtado, 2018).

Del mismo modo se recomienda analizar el índice de dificultad, sin embargo, este índice no es recomendable para pruebas clínicas, sino para test que midan el rendimiento, sobre todo en el campo educativo (Pérez-Sala et al., 2019). Como se pudo observar en esta primera parte, existen buena calidad de los ítems para el diagnóstico de depresión.

Cabe resaltar que la prueba original fue elaborada con los criterios diagnóstico de depresión del DSM IV. Este sistema de clasificación está ampliamente establecido y utilizado internacionalmente para el diagnóstico de depresión en las diferentes edades (niños, adolescentes y adultos), asimismo se

caracteriza por tener un carácter normativo y centrado en los principios psicopatológicos a diferencia del CIE-10 que tiene un fundamento desde la patogenia (Gutierrez, Peña, Santiuste, Garcia, Ochoterena, San Eustaquio y Cánovas, 2019). Por ello, los ítems no solo se deben limitar a una evaluación psicométrica y se recomienda que se evalúe cualitativamente durante las intervenciones, debido a que los reactivos contienen los criterios de diagnóstico del DSM IV y permitirán monitorizar los síntomas de los pacientes.

También se ha verificado la aceptación de la segunda hipótesis, que sostiene la existencia de validez relacionada al constructo, mediante el análisis factorial confirmatorio. Dentro de este marco, para procesar los datos se tomaron 5 modelos diseñadas y re-diseñados por Villarreal-Zegarra, Copez-Lonzoy, Bernabé-Ortiz, Meléndez-Torres y Bazo-Álvarez (2019) y Huarcaya-Victoria, De lama-Moran, Quirós, Bazán, López y Lora (2020). Para ello, previamente se hallaron los permisos estadísticos, es decir, se calculó la medida de adecuación muestral KMO y la prueba de esfericidad de Barlett, los cuales fueron 0.914 y 0.000 respectivamente. Esto indica que se puede continuar con el análisis factorial (Montoya, 2007). 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.

Según los resultados obtenidos en todos los modelos, el chi cuadrado indica un desajuste (0.000), no obstante, este resultado es sensible a la cantidad de muestra que se considera para un estudio (Rose, Markman y Sawilowsky, 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 y 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 (RMSEA) 0.098, considerándose un nivel mediocre. Sin embargo, los índices de ajuste incremental (CFI= .929; NFI= .928; TLI= .905) obtuvieron valores apropiados $\geq 0,90$. Estos resultados concuerdan con el estudio de Wisting, Johnson, Builk, Ro y Bang (2021) donde revelaron un ajuste mediocre para una estructura unidimensional del PHQ-9, independientemente del estado diagnóstico. No obstante, investigadores como Villarreal-Zegarra, et al. (2019) y González-Blanch, Medrano, Muñoz-Navarro, Ruíz-Rodríguez, Moriana, Limonero, Schmitz, Cano-Vindel, y PsicAP Research Group (2018) apoyan la idea de un modelo unidimensional del PHQ-9, debido a que existe una alta correlación entre sus dimensiones, somático y cognitivo, 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 más apropiado, además de ser elegido como el modelo que representa al instrumento. Obtuvo un índice de error cuadrático mediocre (RMSEA) de 0.097, sin embargo, obtuvo valores apropiados en los índices de ajuste incremental (CFI=.932; NFI=.932; TLI=.906) y mejores medidas de ajuste de la parsimonia (Escobedo, Hernandez, Estebané y Martinez. 2016), esto a comparación de las otras estructuras factoriales procesadas.

Los resultados del modelo 2 concuerdan con los estudios de Patel, Oh, Rand, Wu, Cyders, Kroenke y Stewart (2020) quienes emplearon el análisis factorial confirmatorio con una muestra de 31,366 personas y respaldaron un modelo de dos dimensiones. Además, autores como Merz, Malcarne, Roesch, Riley y Sandler (2011) y Saldivia et al (2019) confirmaron una estructura de dos dimensiones del PHQ-9 para las poblaciones latinoamericanas.

Siguiendo con el análisis bidimensional, se halló el RMSEA del modelo 3, el cual es considerado como un nivel muy pobre, no obstante, obtuvo puntajes adecuados en los índices de ajuste incremental (CFI=.929; NFI=.929; TLI=.902), y de modo idéntico el modelo 4 obtuvo un nivel muy pobre en el RMSEA, pero puntajes adecuados en el ajuste incremental (CFI=.929; NFI=.928; TLI=.901). Como se puede observar en ambos modelos el índice de error cuadrático es muy pobre, excepto en los ajustes incrementales (Wisting, Johnson, Builk, Ro y Bang. 2021). Por lo tanto, indicarían que los datos no se adecuan a dichos modelos, y por 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 a comparación de 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. Respecto a los baremos para este modelo, no se hallarán por dimensiones, sino serán la suma de las puntuaciones directas de todos los reactivos, ya que el PHQ-9 es una prueba de tamizaje, esto es posible, ya que las dimensiones tienen una correlación muy alta.

La interpretación de estos resultados fue complicada puesto que los modelos unidimensional y bidimensional mantienen puntajes similares y ambos son respaldados por diferentes investigaciones. Además, en este estudio no se encuentran diferencias relevantes entre los puntajes de la 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, Craske, Roy-Byrne, Stein, Sullivan, Sherbourne, Bystritsky y Chavira. 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).

Por otra parte, es necesario precisar que los índices estadísticos empleados en la presente investigación están basados en estudios empíricos que respaldan su valor, es así que, Jackson, Gillaspay y Purc-Stephenson (2009), indican que los índices de ajuste comparativo (CFI), el índice de ajuste no normado (TLI), la raíz cuadrada media del error de aproximación (RMSEA) y el índice de bondad de ajuste (GFI) son los más utilizados en la literatura especializada que aborda instrumentos psicométricos.

Finalmente, dentro de la validez relacionada al constructo, se procesó el modelo 5 (bifactorial), el cual obtuvo mejores puntajes en la bondad de ajuste (RMSEA= 0.077; CFI= .971; NFI= .971; TLI= .942), 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, Hodge, Wells y Watkins, 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.

Cabe resaltar que los modelos 1, 2, 3 y 4 superan el puntaje mínimo requerido de 0.5 respecto a sus cargas factoriales. Por ende, los resultados son consistentes y coherentes con los índices de bondad de ajuste y los cargas factoriales.

Respecto a la tercera hipótesis, también se acepta su planteamiento que indica que el cuestionario de salud del paciente PHQ-9 tiene confiabilidad por consistencia interna y método de dos mitades. Para tal fin, se evaluó el estadístico de Omega obteniendo como resultado un puntaje de 0.892, este coeficiente, según

Campos-Arias y Oviedo 2008 (citado en Ventura-León y Caycho-Rodriguez. 2017) se considera óptimo. Asimismo, este resultado concuerda con el 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 0.87 de la prueba de PHQ-9. Sin embargo, difiere según el estadístico empleado en la escala de psicopatología depresiva (EPD-6), donde se obtuvo un índice de confiabilidad alfa de Cronbach de 0.736.

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-Leon y Caycho-Rodriguez. 2017).

También, esta hipótesis se ve confirmada con otros valores, para ello se evaluó los coeficientes de Spearman Brown y Guttman, donde se obtuvo 0.824 y 0.753 respectivamente, estos valores son calificados como muy buenos (Garcia, 2006). 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

Por último, para los baremos se analizaron los puntajes directos, obteniéndose una distribución no normal, es decir no paramétrica. Por lo que se empleó la prueba de U de Mann Withney para hallar la diferencia de medianas.

Como resultado se obtuvo que existen diferencias entre los puntajes según el género (Hombre/Mujer) y la edad. En este último se encuentran diferencias

entre todas las edades (adolescentes, adultos y adultos mayores), excepto entre adolescentes y adultos, por ende, se realizaron baremos por género y edad. Estos resultados son similares a los obtenidos en un estudio nacional durante la pandemia, donde indican que para la depresión existen diferencias entre género y grupo etario (Prieto-Molinaria, D., Aguirre, G., de Pierola, I., Luna, G., Merea, L., Lazarte, C., Uribe-Bravo, K. y Zegarra, Á. 2020).

Respecto a los criterios para los puntos de corte fueron tomadas de Gempp, R. Avendaño, C. y Muñoz, C. (2004), quienes indican que los procedimientos basados en percentiles para pruebas clínicas es el cuartil superior, es decir el percentil 75. Es por ello que los baremos presentados en este estudio están determinados a partir del percentil 75 y los niveles se clasificaron en función de los puntos de corte de la prueba original, las cuales son 5, 10, 15 y 20, pero convertidos en percentiles son 15, 33, 50 y 70 respectivamente. Estos cortes se emplearon en el grupo perteneciente al cuartil superior (Pc 75), obteniéndose percentiles 75, 81, 87 y 93.

Debido a que la investigación se desarrolló con base de datos secundarios, no se hallaron los puntajes de la sensibilidad y especificidad de la prueba. Esta limitación según Furukawa 1999 (citado en Gempp, R., Avendaño, C., y Muñoz, C. 2004) vulnera a las normas percentilares frente al efecto de las tasas base del trastorno depresivo en la población, ya que solo nos estamos basando en aproximaciones internas del instrumento. Debido a esta limitación se consideró el criterio normativo para hallar los baremos, pero esto no significa que los puntajes no aportan para un buen criterio en el diagnóstico.

Según De la Flor (2020) existen tres formas para hallar los baremos, la primera se trabaja desde un criterio estadístico tomando en cuenta la puntuación directa para realizar un punto de corte, el cual es 10. Es decir, los puntajes menores a 10 no representan depresión y los puntajes mayores o igual a 10 indican depresión. Esta forma de interpretación fue empleada por investigaciones de análisis multivariado y la ENDES para sus informes (Martina, Ara, Gutierrez, Bolberto y Psicoya. 2017). La segunda, toma en cuenta las recomendaciones del DSM IV para el diagnóstico de depresión, donde indican que, si no puntúan en los ítems principales 1 y 2, se suspende la prueba. Y el tercero considera un criterio más clínico siguiendo el algoritmo diagnóstico de la depresión mayor, donde indican que se debe puntuar en cinco o más ítems, de los cuales al menos uno debe ser el ítem 1 o 2 (anhedonia o animo depresivo).

Como se puede observar el instrumento tiene diferentes formas para su interpretación. En la presente investigación se optó por un criterio normativo, debido a que la naturaleza de la muestra es no clínica y al ser tomado de una base de datos de acceso libre no se halló la especificidad y sensibilidad de la prueba, lo que limita su interpretación bajo el principio de criterio.

Debe suponerse que los índices de discriminación, que por cierto son adecuados en esta investigación, pueden satisfacer para un criterio de baremación clínica o de criterio, ya que dichos índices diferencian a las personas con y sin síntomas de depresión y dentro de este último con diferentes niveles de depresión. Pero lo cierto es que solo evalúan de manera particular cada reactivo y no se pueden comparar con los puntajes de la especificidad y sensibilidad, los cuales evalúan incorporando todos los ítems. Desde luego, es recomendable hallar esos índices

(sensibilidad y especificidad) de la prueba y junto a las normas percentilares obtener cortes que maximicen los resultados.

De esta manera se estaría contribuyendo con el sistema de evaluación del primer nivel de atención en el Perú y brindando herramientas de tamizaje para las investigaciones en situación de pandemia. Según Vega-Dienstmaier (2018), sugiere generar versiones más cortas de los instrumentos para detectar depresión y evaluar su capacidad diagnóstica. Mas aún, en un contexto peruano donde los profesionales de la salud mental son limitados (Huarcaya, 2020).

Así tenemos versiones cortas como CES-D, escala de depresión de Hamilton, la EDZ, el inventario de depresión de Beck y el EPDS. Muchos de ellos son utilizados para evaluar a las personas que atraviesa por la pandemia, por ejemplo, Prieto-Molinaria, D., Aguirre, G., de Pierola, I., Luna, G., Merea, L., Lazarte, C., Uribe-Bravo, K. y Zegarra, Á. (2020), emplearon la escala de psicopatología depresiva (EPD-6) para evaluar a un grupo de adultos residentes de Lima Metropolitana. Sin embargo, a pesar de ser instrumentos de tamizaje aun cuentan con una mayor cantidad de ítems a comparación del PHQ-9, además, sus tipificaciones tienen muchos años de antigüedad.

En síntesis, el PHQ-9 es 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, ya que se encuentra actualizada con datos del 2019.

CAPÍTULO VI: CONCLUSIÓN

Esta tesis ha investigado las propiedades psicométricas de la versión peruana del cuestionario sobre la salud del paciente (PHQ-9) para tamizaje de depresión en la población peruana, apoyado por una base de datos libre.

- El cuestionario de salud sobre la salud del paciente (PHQ-9) presenta propiedades psicométricas significativas para la evaluación de la depresión en la población peruana.
- Mediante los índices de homogeneidad, discriminación y las comunalidades se evidencia buena calidad de los ítems, es decir todos los reactivos miden el mismo constructo y contribuyen para diferenciar a las personas con diferentes niveles de depresión.
- Mediante el análisis factorial confirmatorio se evidencia la validez relacionada al constructo, con una adecuada bondad de ajuste a un modelo bidimensional. La primera dimensión mide los síntomas somáticos, y la segunda los síntomas cognitivos. Además, la evidencia demuestra que también existe una alta correlación entre las dimensiones, por lo que los baremos se interpretarán como un solo puntaje directo.
- Mediante el coeficiente Omega y el método de dos mitades se evidencia una alta confiabilidad para evaluar con el PHQ-9 los síntomas depresivos en la población peruana.
- Mediante la U de Mann Whitney y las normas percentilares se evidencia diferencias entre edad y género, por ello se presentan puntajes por separado. Finalmente, el PHQ-9 cuenta con 5 niveles para evaluar los síntomas depresivos: sin síntomas relevantes, leve, moderado, moderadamente severo y severo.

RECOMENDACIONES

- Cuando toman muestras grandes para el estudio, se sugiere sacar una submuestra para el análisis factorial confirmatorio, ya que la cantidad influye en los resultados del estadístico chi cuadrado.
- Se sugiere tener precaución al momento de emplear el modelo bifactor, ya que se evidencia un sesgo en los resultados, si no consideran otros estadísticos.
- Se sugiere seguir evaluando el modelo factorial más apropiado para el PHQ-9.
- Si se pretende trabajar con fuentes de datos primarias, se sugiere hallar los puntajes de sensibilidad y especificidad para maximizar los puntajes de los baremos.
- Si se trabaja con fuentes de datos primarias, se sugiere emplear los estadísticos descriptivos para el análisis de ítems.
- Es necesario capacitar al personal de salud no especializado en la salud mental para el uso correcto del PHQ-9, ya que la interpretación no sólo es cuantitativa, también es cualitativa.
- Se recomienda continuar con los estudios psicométricos en pruebas de tamizaje sobre la depresión, más aún post pandemia.
- Se recomienda validar el PHQ-9 para su aplicación virtual.

REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association, American Psychological Association, y National Council on Measurement Education. (2014). Estándares para pruebas educativas y psicológicas. Recuperado de https://www.testingstandards.net/uploads/7/6/6/4/76643089/spanish_standards_pdf.pdf
- American Psychiatric Association (1995). Diagnostic and statistical manual of mental disorders (5ª ed). Recuperado de <https://www.eafit.edu.co/ninos/reddelaspreguntas/Documents/dsm-iv-manual-diagnostico-estadistico-trastornos-mentales.pdf>
- American Psychiatric Association. (1995). Trastornos del estado de ánimo. Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-IV-TR (pp. 323–400). Barcelona: Masson
- Antiporta, D., Cutipe, Y., Mendoza, M., Celentano, D., Stuart, E. y Bruni, A. (2021). Síntomas depresivos entre residentes adultos peruanos en medio de un cierre nacional durante la pandemia de COVID-19. Recuperado de <https://pubmed.ncbi.nlm.nih.gov/33602157/>
- Baader, T., Molina, J., Venezian, S., Rojas, C., Farias, R., Fierro-Freixenet, C., Backenstrass, M, y Mundt, C. (2012). Validación y utilidad de la encuesta PHQ-9 en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. Recuperado de https://scielo.conicyt.cl/scielo.php?pid=S0717-92272012000100002&script=sci_arttext

- Ban, T. (2014). From Melancholia to depression: a history of diagnosis and treatment. Recuperado de [https://inhn.org/fileadmin/previews_new/From Melancholia to Depression March 6 2014.pdf](https://inhn.org/fileadmin/previews_new/From_Melancholia_to_Depression_March_6_2014.pdf)
- Campo, A., y Oviedo, C. (2008). Propiedades psicométricas de una escala: la consistencia interna. *Rev Salud Pública*, 10 (5), pp. 831-839.
- Cassiani-Miranda, C., Vargas-Hernández, M., Pérez-Anibal, E., Herazo-Bustos, M., & Hernández-Carrillo, M. (2017). *Biomédica: revista del Instituto Nacional de Salud*, 37(0), 112–120. <https://doi.org/10.7705/biomedica.v37i0.3221>
- Chan, L., y Quezada, J. (2013). Baremos de pruebas psicológicas: niños y adolescentes. Recuperado el 11 de octubre de 2020, de [https://www.unife.edu.pe/publicaciones/psicologia/BAREMOSdePRU EBAS.pdf](https://www.unife.edu.pe/publicaciones/psicologia/BAREMOSdePRU_EBAS.pdf)
- Correia, A., De Souza, C., Susser, E, y Rossi, P. (2017). Depresin relacionada con el trabajo en personal de equipos de atención primaria en Brasil. Recuperado de <https://iris.paho.org/bitstream/handle/10665.2/34367/v41eAPHA22017.pdf?sequence=1&isAllowed=y>
- 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. recuperado de [https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/9023/Validacion DelaFlorValdivia Martin.pdf?sequence=1&isAllowed=y](https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/9023/Validacion_DelaFlorValdivia_Martin.pdf?sequence=1&isAllowed=y)

De la Fuente, S. (2011). Análisis factorial. Recuperado de <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>

De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., Johnson, L., Thankappan, K., y Williams, E. (2021). Are the PHQ-9 and GAD-7 Suitable for Use in India? A Psychometric Analysis. *Frontiers in psychology*, 12, 676398. <https://doi.org/10.3389/fpsyg.2021.676398>

Dominguez, S. (2013). ¿Ítems Politémicos o Dicotómicos? Un estudio empírico con una escala unidimensional. *Revista Argentina de Ciencias del Comportamiento*, 5(3),30-37. [fecha de Consulta 31 de octubre de 2021]. ISSN: Recuperado de: <https://www.redalyc.org/articulo.oa?id=333430131005>

Elosua, P. (2003). Sobre la validez de los tests. Recuperado de <http://www.psicothema.com/psicothema.asp?id=1063>

Escobedo, M., Hernandez, J., Estebané, V. y Martínez, G. (2016). Modelos de ecuaciones estructurales: características, fases, construcción, aplicación y resultados. Recuperado de <https://scielo.conicyt.cl/pdf/cyt/v18n55/art04.pdf>

Escovar, E., Craske, M., Roy-Byrne, P., Stein, M., Sullivan, G., Sherbourne, C., Bystritsky, A. y Chavira, D. (2018). Influencias culturales en los síntomas de salud mental en una muestra de atención primaria de pacientes latinos. Recuperado de <https://www.sciencedirect.com/science/article/abs/pii/S0887618517304577>

Escovar, E., Craske, M., Roy-Byrne, P., Stein, M., Sullivan, G., Sherbourne, C., Bystritsky, A y Chavira, D (2018). Influencias culturales en los síntomas de salud mental en una muestra de atención primaria de pacientes latinos. *Revista de trastornos de*

ansiedad, 55, 39-47. Recuperado de <https://www.paho.org/costa-rica/dmdocuments/Salud.Mental.en.Atencion.Primaria.pdf>

Fabara, E. (2013). Valoración del PHQ-9 y GHQ-12 como herramientas de tamizaje de la depresión en el ámbito de atención primaria en el Ecuador experiencia del uso de versiones electrónicas y autoadministradas en la Clínica Universitaria USFQ sede San Rafael. recuperado de: <https://repositorio.usfq.edu.ec/handle/23000/2794>

Ferrando, P., y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. Papeles del Psicólogo, 31(1),18-33. ISSN: 0214-7823. Recuperado de <https://www.redalyc.org/articulo.oa?id=77812441003>

Gempp, R., Avendaño, C., y Muñoz, C (2004). Normas y punto de corte para la Escala de Depresión del Centro para Estudios Epidemiológicos (CES-D) en población juvenil chilena. Terapia Psicológica, 22(2),145-156. ISSN: 0716-6184. Disponible en: <https://www.redalyc.org/articulo.oa?id=78522205>

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: Recuperado de <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., y 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>

Gutierrez, M., Peña, L., Santiuste, M., Garcia, D., Ochoterena, MM., San Eustaquio, F. y Cánovas, M. (2019). Comparación de los síntomas de clasificación de los trastornos

mentales CIE-10 y DSM-IV. Recuperado de <https://www.atlasvpm.org/wp-content/uploads/2019/06/Comparaci%C3%B3n-de-los-sistemas-de-clasificaci%C3%B3n-de-los-trastornos-mentales-CIE-10-y-DSM-IV.pdf>

Hair, J., Anderson, R., Tatham, R., y Black, W. (2004). Análisis Multivariante. Madrid: Pearson.

Hernández, (2014). Hernández, R. (2014). Metodología de Investigación. Recuperado el 27 de febrero de 2021, de <https://www.uca.ac.cr/wpcontent/uploads/2017/10/Investigacion.pdf>

Huarcaya-Victoria, De lama-Moran, Quiros, Bazan, Lopez y Lora (2020). Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú. Recuperado de http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0034-85972020000200072#B20

Huberty, T (2012). Anxiety and depression in children and adolescents: assessment, intervention, and prevention. Recuperado de [10.1007/978-1-4614-3110-7](https://doi.org/10.1007/978-1-4614-3110-7)

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>

Instituto Nacional de Salud. (2018). Estudio epidemiológico de salud mental en hospitales generales y centros de salud de Lima Metropolitana- 2015. Recuperado de <https://www.inism.gob.pe/investigacion/archivos/estudios/2020/Vol%20XXXIV%202018%20Nro1%20EESM%20en%20Hospitales%20Generales%20y%20Centros%20de%20Salud.pdf>

- Jackson, D. Gillaspay, J. y Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods*, 14(1), 6-23. doi: <http://dx.doi.org/10.1037/a0014694>
- Kline, P. (1999). *El manual de pruebas psicológicas*. Londres. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_nlinks&ref=3461763&pid=S1657-9267200900020001800023&lng=pt
- Lam, N. Contreras, H. Mori, E. Gil, E. Espinoza, E. Córdova, G. (2008). Estudio de la validez interna y concurrente del inventario de depresión de Beck para atención primaria (BDI-PC) en población gestante de Lima Metropolitana, mayo – junio de 2008. Recuperado de https://sisbib.unmsm.edu.pe/BVRevistas/epidemiologia/v13_n2/pdf/a07v13n2.pdf
- Lloret, S., Ferres, A., Hernandez, A., y Tomas, I. (2014). El Análisis Factorial Exploratorio de los Ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. Recuperado de <https://dx.doi.org/10.6018/analesps.30.3.199361>
- López, M., y Gutiérrez, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista d’Innovació i Recerca en Educació*, 12(2), 1–14. <http://doi.org/10.1344/reire2019.12.227057>
- Manterola, C., Grande, L., Otzen, T., Garcia, N., Salazar, P. y Quiroz, G. (2018). Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. Recuperado de <https://scielo.conicyt.cl/pdf/rci/v35n6/0716-1018-rci-35-06-0680.pdf>
- Martina, M., Ara, M., Gutiérrez, C., Nolberto, V., y Piscocoya, J. (2017). Depresión y factores asociados en la población peruana adulta mayor según la ENDES 2014-

2015. Anales de la Facultad de Medicina, 78(4), 393-397. <https://dx.doi.org/10.15381/anales.v78i4.14259>

Medina, A., Arevalo, N., Beltran, S., Chavarro, Y., Herazo, E., y Campo-Arias, A. (2013). Consistencia interna y estructura interna del cuestionario AUDIT en amerindios. Recuperado de <http://www.scielo.org.co/pdf/inan/v16n28/v16n28a02.pdf>

Medrano, L. y 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>

Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., lozano, L., Turbany, J, y Valero, S. (2013). Psicometría. 1 ed. Recuperado de https://www.researchgate.net/profile/Julio-Meneses-2/publication/293121344_Psicometria/links/584a694408ae5038263d9532/Psicometria.pdf

Merz, E., Malcarne, V., Roesch, S., Riley, N. y 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. doi:10.1037/a0023883

Merz, E., Malcarne, V., Roesch, S., Riley, N., y 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>

- Meyer, A. (1910). The nature and conception of dementia praecox. The Journal of Abnormal Psychology, 5(5), 274–285. Recuperado de <https://doi.org/10.1037/h0073298>
- Molina, M., Ochoa, C, y Ortega, E. (2020). Comparación de dos medias. Prueba de la T de Student. Recuperado de <https://evidenciasenpediatria.es/articulo/7699/comparacion-de-dos-medias-pruebas-de-la-t-de-student>
- Montenegro, M. y Kilstein, J. (2017). Utilidad del PHQ-9 en el diagnóstico de trastorno depresivo en una sala de Clínica Médica. Recuperado de <https://www.intramed.net/contenidover.asp?contenidoid=90825>
- Montoya, O. (2007). Aplicación del análisis factorial a la investigación de mercados. caso de estudio. Recuperado de <https://dialnet.unirioja.es>
- Morales, C. (2017). La depresión: Un reto para toda la sociedad del que debemos hablar. Recuperado de <https://www.scielosp.org/pdf/rcsp/2017.v43n2/136-138/es#:~:text=La%20depresi%C3%B3n%20provoca%20un%20nivel,consiguien te%20impacto%20econ%C3%B3mico%20y%20social.>
- Morgan, B., de Bruin, G. y 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. doi: 10.1177/0081246314528834
- Muñiz, J. (2018). Introducción a la psicometría: teoría clásica y TRI. 1ed. Recuperado de <https://es.scribd.com/document/423397957/Muniz-Fernandez-J-Introduccion-a-la-Psicometria-Ediciones-Piramide-2018-pdf>

National Institute of Mental Health NIH. (2016). Depresión: información básica. recuperado de https://www.nimh.nih.gov/health/publications/espanol/depression-sp/19-mh-8079spdf_159009.pdf

Olea, J. Ponsoda, V. Abad, F. y Garrido, J. (2006). Introducción a la psicometría, Teoría clásica de los Test y teoría de la respuesta al ítem. Madrid: Universidad Autónoma de Madrid. Recuperado el 15 de setiembre de 2020, de <http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/39/ARCHIV>

OMS. (2018). Depresión. Recuperado de: <https://www.who.int/es/news-room/fact-sheets/detail/depression>

Organización Mundial de la Salud (2006). Prevencion en el suicidio: un instrumento en el trabajo. Recuperado de https://www.who.int/mental_health/media/general_physicians_spanish.pdf?ua=1

Organización Mundial de la Salud. (2017). Depresión y otros trastornos mentales comunes: estimación de salud mundial. recuperado de https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

Organización panamericana de la salud y Organización mundial de la salud (2017). Depresión y otros trastornos mentales comunes: estimaciones sanitarias mundiales. Recuperado de: <https://iris.paho.org/bitstream/handle/10665.2/34006/PAHONMH17005-spa.pdf?sequence=1&isAllowed=y>

Pacheco, E. (2018). El tiempo de espera en la atención en salud. *El peruano*. Recuperado de <https://elperuano.pe/noticia/65112-el-tiempo-de-espera-en-la-atencion-en-salud>

- Patel, J., Oh, Y., Rand, K., Wu, W., Cyders, M., Kroenke, K., y 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>
- Paykel, E. (2008). Basic concepts of depression. recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181879/>
- Pérez-Sala, C., Sirlopú, D., Cobo, R. y Awad, A. (2019). Análisis Bifactorial de la Escala de Participación Escolar en una Muestra de Estudiantes Chilenos. Recuperado de https://www.redalyc.org/journal/4596/459661296004/html/#redalyc_459661296004_ref31
- Petersen, J., Paulitsch, M., Hartig, J., Mergenthal, K., Gerlach, F., y Gensichen, J. (2015). Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *Journal of affective disorders*, 170, 138–142. <https://doi.org/10.1016/j.jad.2014.08.053>
- Prieto-Molinaria, D., Aguirre, G., de Pierola, I., Luna, G., Merea, L., Lazarte, C., Uribe-Bravo, K. y Zegarra, Á. (2020). Depresión y ansiedad durante el aislamiento obligatorio por el COVID-19 en Lima Metropolitana. *Liberabit*, 26(2),425. <https://dx.doi.org/10.24265/liberabit.2020.v26n2.09>
- Rodero, C., Molina, M., Fernández, M., y Redondo, M. (2016). Análisis de fiabilidad y validez de un cuestionario docente. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=5814961>

- Rodriguez, V. (2014). Medición y evaluación de aprendizaje. Recuperado de [Teoría Clásica del Test \(TCT\) – Medición y Evaluación del aprendizaje: \(wordpress.com\)](#)
- 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.
- Salazar, C. y Del Castillo, S. (2018). Fundamentos básicos de estadística. Primera edición. Recuperado de <http://www.dspace.uce.edu.ec/bitstream/25000/13720/3/Fundamentos%20B%C3%A1sicos%20de%20Estad%C3%ADstica-Libro.pdf>
- Saldivia, S., Aslan, J., Cova, F., Vicente, B., Inostroza, C., y Rincón, P. (2019). Psychometric characteristics of the Patient Health Questionnaire (PHQ-9). 113 revista Médica de Chile, 147(1), 53–60. doi:10.4067/S0034- 98872019000100053
- Sausa, M. (2017). Ministerio de Salud: Riesgo de suicidio es 8 veces mayor entre las mujeres que sufren violencia. Perú 21. Recuperado de <https://peru21.pe/vida/salud/ministerio-salud-riesgosuicidio-8-veces-mayor-mujeres-sufren-violencia-375090>
- SEP. (2015). La depresión es, más que nunca, un reto para la salud pública en Europa. Recuperado de: [http:// www.sepsiq.org/informacion/noticia/la-depresion-esmas-que-nunca-un-reto-para-la-salud-publica-eneuropa/392](http://www.sepsiq.org/informacion/noticia/la-depresion-esmas-que-nunca-un-reto-para-la-salud-publica-eneuropa/392)
- Sireci, S. y Faulkner-Bond, M. (2014). Evidencia de validez basada en el contenido de la prueba. Recuperado de <http://www.psicothema.com/psicothema.asp?id=4167>

- Vega-Dienstmaier, J. (2018). Construcción de versiones cortas de la escala de psicopatología depresiva para tamizaje de depresión mayor y sus características psicométricas. Recuperado de <https://revistas.upch.edu.pe/index.php/RNP/article/view/3383/3593>
- Ventura J. y Caycho, T. (2017). El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. Recuperado de <https://www.redalyc.org/pdf/773/77349627039.pdf>
- Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabe-Ortiz, A., Melendez-Torres G. y Bazo-Alvarez, J. (2019). Se pueden hacer comparaciones de grupo válidas con el Cuestionario de salud del paciente (PHQ-9): un estudio de invariancia de medición entre grupos por características demográficas. Recuperado de <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221717#pone.0221717.ref016>
- Zhong, Q., Gelaye, B., Fann, J., Sanchez, S. y Williams, M. (2014). Cross -Cultural validity of the Snaish version of PHQ-9 among pregnant Peruvian women. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4004697/>

ANEXOS

1. Consentimiento informado

REPÚBLICA DEL PERÚ								
INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA								
ENCUESTA DEMOGRÁFICA Y DE SALUD FAMILIAR								
ENDES - 2019								
CUESTIONARIO DE SALUD								
PRIMER SEMESTRE								
AMPARADO POR EL D. L. N° 604-SECRETO ESTADÍSTICO								
CONGLOMERADO			VIVIENDA			HOGAR		

PRESENTACIÓN	
<p>Señor / Señora / Señorita, mi nombre es _____ y soy trabajadora del Instituto Nacional de Estadística e Informática, institución que por especial encargo del Ministerio de Salud está realizando un estudio sobre la salud de las personas de 15 años a más y de los niños menores de 12 años, a nivel nacional; con el objeto de evaluar y orientar la futura implementación de los diversos programas de salud, dirigidos a mejorar las condiciones de salud de la población en el país. La información que nos brinde es estrictamente confidencial y permanecerá en absoluta reserva.</p>	

NOMBRE Y NÚMERO DE ORDEN DE LA PERSONA SELECCIONADA SEGÚN CUESTIONARIO DEL HOGAR: _____	<table border="1"><tr><td></td><td></td></tr></table>		

2. Prueba PHQ-9 obtenida de la encuesta demográfica y de salud familiar ENDES 2019

En los últimos 14 días, es decir, entre el _____ y el día de ayer, usted ha sentido molestias o problemas como:					
	PARA NADA	VARIOS DÍAS (1 A 6 DÍAS)	LA MAYORÍA DE DÍAS (7 A 11 DÍAS)	CASI TODOS LOS DÍAS (12 A + DÍAS)	
NO HACER INFERENCIAS DE CAUSALIDAD					
SI RESPONDE "SÍ", PREGUNTE: ¿Cuántos días?					
SI RESPONDE "NO", CIRCULE "0" (PARA NADA)					
A. ¿Pocas ganas o interés en hacer las cosas? (DE SER NECESARIO LEA: Es decir, no disfruta sus actividades cotidianas)	A. POCO INTERES.....	0	1	2	3
B. ¿Sentirse desanimada(o), deprimida(o), triste o sin esperanza?	B. DEPRIMIDA(O).....	0	1	2	3
C. ¿Problemas para dormir o mantenerse dormida(o), o en dormir demasiado?	C. DORMIR.....	0	1	2	3
D. ¿Sentirse cansada(o) o tener poca energía sin motivo que lo justifique?	D. CANSADA(O).....	0	1	2	3
E. ¿Poco apetito o comer en exceso?	E. APETITO.....	0	1	2	3
F. ¿Dificultad para poner atención o concentrarse en las cosas que hace? (DE SER NECESARIO LEA: Como leer el periódico, ver televisión, escuchar atentamente la radio o conversar con otras personas)	F. PONER ATENCIÓN.....	0	1	2	3
G. ¿Moverse mas lento o hablar más lento de lo normal o sentirse más inquieta(o) o intranquila(o) de lo normal?	G. MOVERSE.....	0	1	2	3
H. ¿Pensamientos de que sería mejor estar muerta(o) o que quisiera hacerse daño de alguna forma buscando morir?	H. MORIR.....	0	1	2	3
I. ¿Sentirse mal acerca de si misma(o) o sentir que es una(un) fracasada(o) o que se ha fallado a sí misma(o) o a su familia?	I. SENTIRSE MAL.....	0	1	2	3