



FACULTAD DE PSICOLOGÍA

INSTRUMENTOS PARA MEDIR DEPRESIÓN VALIDADOS EN POBLACIÓN

PERUANA: UNA REVISIÓN SISTEMÁTICA

Línea de investigación: Evaluación psicológica y psicométrica

Tesis para optar el Título Profesional de Licenciado en Psicología con
mención en psicología clínica

Autor

Mayanga Aliaga, Manuel Humberto

Asesora

Tantalean Terrones de Callohuanca, Lizley Janne

ORCID: 0000-0001-7439-5699

Jurado

Díaz Hamada, Luis Alberto

Mayorga Falcon, Luz Elizabeth

Aguilar Mori De Oporto, Karim

Lima - Perú

2026



INSTRUMENTOS PARA MEDIR DEPRESIÓN VALIDADOS EN POBLACIÓN PERUANA: UNA REVISIÓN SISTEMÁTICA

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

16%

FUENTES DE INTERNET

10%

PUBLICACIONES

8%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1

hdl.handle.net

Fuente de Internet

2%

2

www.researchgate.net

Fuente de Internet

1%

3

repositorio.unfv.edu.pe

Fuente de Internet

<1%

4

www.scielo.edu.uy

Fuente de Internet

<1%

5

Ana L. Vilela-Estrada, David Villarreal-Zegarra, Mauricio Toyama, Adriana Carbonel et al.
"Psychometric properties of the patient health questionnaire-8 and general anxiety disorder-7 in adolescents and young adults from three Latin American cities: Internal structure, invariance, internal consistency and divergent validity", Journal of Affective Disorders, 2025

Publicación

<1%

6

es.wikipedia.org

Fuente de Internet

<1%

7

Submitted to Universidad Nacional Federico Villarreal

Trabajo del estudiante

<1%

8

repositorio.ucv.edu.pe

Fuente de Internet

<1%



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

FACULTAD DE PSICOLOGÍA

**INSTRUMENTOS PARA MEDIR DEPRESIÓN
VALIDADOS EN POBLACIÓN PERUANA:
UNA REVISIÓN SISTEMÁTICA**

Línea de Investigación:
Evaluación psicológica y psicométrica

Tesis para optar el título profesional de Licenciado en Psicología con mención en Psicología
Clínica

Autor

Mayanga Aliaga, Manuel Humberto

Asesora:

Tantalean Terrones de Callohuanca, Lizley Janne
ORCID: 0000-0001-7439-5699

Jurado:

Díaz Hamada, Luis Alberto
Mayorga Falcon, Luz Elizabeth
Aguilar Mori De Oporto, Karim

Lima – Perú

2026

Dedicatoria

A mi familia, por su apoyo incondicional y por ser el refugio al que siempre regreso.

A todos aquellos sufrientes quienes cargan en silencio el peso de la depresión, que combaten
sus propios fantasmas sin que nadie lo sepa y, aun así, siguen adelante.

A Mario Bunge, quien me enseñó que la ciencia es, ante todo, una forma de amar al mundo
con rigor y honestidad.

Agradecimientos

A la Universidad Nacional Federico Villarreal, mi alma máter, por brindarme la formación académica y los espacios necesarios para desarrollarme como investigador.

A la Mg. Lizley Tantalean, asesora y docente, por su guía experta, su paciencia y su compromiso durante todo el proceso de elaboración de esta tesis. Su orientación fue fundamental para que este trabajo alcanzara el rigor que merece.

A mi familia, por su presencia constante, su comprensión en los momentos de mayor exigencia y por creer en mí incluso cuando yo dudaba.

ÍNDICE

RESUMEN	viii
ABSTRACT.....	ix
I. INTRODUCCIÓN.....	1
1.1. Descripción y formulación del problema	1
1.2. Antecedentes	3
<i>1.2.1. Antecedentes internacionales</i>	<i>3</i>
<i>1.2.2. Antecedentes nacionales</i>	<i>5</i>
1.3. Objetivos	7
<i>1.3.1. Objetivo general.....</i>	<i>7</i>
<i>1.3.2. Objetivos específicos</i>	<i>7</i>
1.4. Justificación.....	7
II. MARCO TEÓRICO	10
2.1. Bases teóricas sobre el tema de investigación.....	10
<i>2.1.1. Definición y concepto de la depresión</i>	<i>10</i>
<i>2.2.1. Clasificación y tipos de depresión según criterios diagnósticos</i>	<i>12</i>
<i>2.3.1. Principales aspectos de la evaluación psicométrica de la depresión.....</i>	<i>14</i>
III. MÉTODO	19
3.1. Tipo de investigación	19
3.2. Ámbito temporal y espacial.....	20
3.3. Variables.....	20
<i>3.3.1. Medición de la depresión</i>	<i>20</i>

3.4. Población y muestra	22
3.5. Instrumento.....	23
3.5.1. Rúbrica de adaptaciones y validaciones de instrumentos psicométricos	23
3.6. Procedimientos	25
3.7. Análisis de datos.....	29
3.8. Consideraciones éticas	30
IV. RESULTADOS	31
V. DISCUSIÓN DE RESULTADOS	93
VI. CONCLUSIONES.....	117
VII. RECOMENDACIONES	119
VIII. REFERENCIAS.....	121
IX. ANEXOS	143
Anexo A. Matriz de consistencia	143
Anexo B. Protocolo PRISMA 2020	144
Anexo C. Artículos evaluados.....	146

ÍNDICE DE TABLAS

Tabla 1 Operacionalización de la variable medición de la depresión.....	21
Tabla 2 Rúbrica de adaptaciones y validaciones de instrumentos psicométricos.....	23
Tabla 3 Términos de búsqueda según componente de método PICO.....	27
Tabla 4 Muestra total por cada instrumento validado.....	33
Tabla 5 Características de los estudios y muestras de los instrumentos para medir depresión validados en población peruana	36
Tabla 6 Procedimientos de adaptación cultural de los instrumentos para medir depresión validados en población peruana	41
Tabla 7 Fuentes de evidencia de confiabilidad reportados en los estudios de validación de instrumentos de depresión en población peruana	44
Tabla 8 Fuentes de evidencia de validez reportadas en los estudios de validación de instrumentos de depresión en población peruana	46
Tabla 9 Indicadores normativos en los estudios de validación de instrumentos de depresión en población peruana	48

ÍNDICE DE FIGURAS

Figura 1 Diagrama de flujo de búsqueda y selección - PRISMA 2020	31
Figura 2 Frecuencia de instrumentos según la documentación revisada	32
Figura 3 Característica principal del grupo muestral empleado	33
Figura 4 Tipo de muestreo empleado	34
Figura 5 Frecuencia de año de cada publicación	35

RESUMEN

La depresión constituye un problema relevante de salud mental, por lo que disponer de instrumentos válidos y confiables resulta fundamental para su evaluación. El objetivo de esta investigación fue analizar las características psicométricas de los instrumentos para medir la depresión validados en población peruana. Se realizó una revisión sistemática siguiendo PRISMA 2020, con búsquedas en Scopus, Web of Science, SciELO, PubMed, BVS y Google Scholar en enero de 2026. Para la extracción de datos se elaboró una rúbrica ad hoc basada en los estándares AERA, APA y NCME (2018) y en las directrices de la ITC (2013), organizada en cuatro criterios: adaptación cultural, confiabilidad, validez y normas. De 3676 registros identificados, se incluyeron 35 estudios que reportaron 39 validaciones de 13 instrumentos, con una muestra acumulada de 221 958 participantes. El PHQ, en sus versiones PHQ-9, PHQ-8 y PHQ-2, fue el instrumento con mayor evidencia acumulada, seguido del BDI en sus distintas versiones. Los hallazgos mostraron que la adaptación cultural fue el criterio menos desarrollado, que la evidencia de validez se concentró principalmente en la estructura interna y en las relaciones con otras variables, y que la confiabilidad se evaluó casi siempre mediante consistencia interna, así mismo, el desarrollo normativo presentó gran variabilidad entre instrumentos. En conjunto, la evidencia disponible muestra un avance importante en la validación de instrumentos para depresión en el Perú, pero también revela brechas en adaptación cultural, diversidad muestral y elaboración de normas locales, especialmente en población infanto-juvenil y en contextos rurales e indígenas.

Palabras clave: depresión, propiedades psicométricas, revisión sistemática, validación de instrumentos, población peruana

ABSTRACT

Depression is a significant mental health problem, making the availability of valid and reliable assessment instruments essential. The aim of this study was to analyze the psychometric properties of depression measurement instruments validated in the Peruvian population. A systematic review was conducted following PRISMA 2020 guidelines, with searches in Scopus, Web of Science, SciELO, PubMed, BVS, and Google Scholar in January 2026. Data extraction was carried out using an ad hoc rubric based on the AERA, APA, and NCME (2018) standards and the ITC (2013) guidelines, organized around four criteria: cultural adaptation, reliability, validity, and norms. From 3,676 identified records, 35 studies were included, reporting 39 validations of 13 instruments, with a cumulative sample of 221,958 participants. The PHQ, in its PHQ-9, PHQ-8, and PHQ-2 versions, was the instrument with the greatest accumulated evidence, followed by the BDI in its various versions. Findings showed that cultural adaptation was the least developed criterion, that validity evidence was concentrated primarily in internal structure and relationships with other variables, and that reliability was almost exclusively assessed through internal consistency; additionally, normative development showed considerable variability across instruments. Overall, the available evidence reflects meaningful progress in the validation of depression instruments in Peru, while also revealing gaps in cultural adaptation, sample diversity, and the development of local norms, particularly for child and adolescent populations and in rural and indigenous contexts.

Keywords: depression, psychometric properties, systematic review, instrument validation, Peruvian population

I. INTRODUCCIÓN

1.1. Descripción y formulación del problema

La depresión es uno de los problemas en salud mental más importantes a nivel mundial; en este sentido, se calcula que aproximadamente un 5% de los adultos de todo el globo padecen depresión (alrededor de 280 millones), afectando significativamente más a mujeres que a hombres (50% más probabilidades), pudiendo desembocar en conductas autolesivas o, directamente, en el suicidio, mientras que en países con bajos recursos económicos más del 75% de personas que padecen depresión no reciben un tratamiento adecuado (Organización Mundial de la Salud [OMS], 2023).

En el Perú, la depresión es el problema de salud mental que ocupa el primer lugar entre los pacientes con enfermedades crónicas no transmisibles, con una prevalencia del 30.3% entre pacientes adultos pertenecientes a centros de salud y hospitales de Lima Metropolitana (Saavedra, 2018), así mismo, se atendió a más de 300 000 personas por casos de depresión en el sistema de salud público peruano, lo que revela un aumento del 12% de casos en comparación con las atenciones antes del estado de emergencia, según el Ministerio de Salud (MINSA, 2022).

Es debido a esta realidad que la medición, evaluación y estudio de esta variable desde una perspectiva cuantitativa y estandarizada es vital, a fin de identificar oportunamente casos probables, estimar la severidad de los síntomas y monitorear la evolución clínica y la respuesta a intervenciones en distintos niveles de atención (Siniscalchi et al., 2020), ya que contribuye a dimensionar la carga psicopatológica y orientar decisiones de salud pública basadas en evidencia (Kilbourne et al., 2018). No obstante, para que estas mediciones sean válidas, los instrumentos deben demostrar propiedades psicométricas adecuadas y pertinencia cultural y lingüística en la población donde se aplican, pues de lo contrario se incrementa el riesgo de

subestimar o sobreestimar la depresión y de tomar decisiones clínicas o preventivas poco precisas (Bibi et al., 2020).

En particular, en nuestro país las escalas, test o instrumentos para medir la depresión son ampliamente utilizados por psicólogos y psiquiatras, así como por otros profesionales relacionados a la salud mental (Robles, 2008), sin embargo, la validación y adaptación de dichos instrumentos puede ser más que cuestionable, teniendo en cuenta la gran cantidad de realidades, culturas y poblaciones diametralmente diferentes que coexisten en nuestro territorio (Alarcón, 2020).

Contribuyendo a esta problemática, el MINSA (2008) recomienda el uso de escalas como el SRQ 18 para la evaluación de síntomas relacionados a la depresión en el primer nivel de atención, un instrumento que solo posee una única validación peruana en una institución militarizada realizada hace más de 20 años (Véliz, 1998), lo que no ofrece resultados concluyentes para otras poblaciones y contextos (Vega-Dienstmaier, 2022), y recomienda la *Hamilton Depression Rating Scale* (HDRS) para la exploración específica de la depresión en pacientes ya identificados, la cual carece de estudios de validación propiamente dichos en el territorio nacional (Cjuno et al., 2022), por lo que no se conoce cuáles son sus propiedades psicométricas dentro de la población peruana.

Este es un problema muy grave, ya que el MINSA deposita la salud mental de los peruanos en instrumentos sin el mínimo grado de evidencia que cabe esperar para herramientas psicométricas, más aún frente a una variable tan importante como la depresión, tal y como mencionan Cjuno et al. (2022) “Lo anterior es preocupante, pues en la selección de instrumentos de tamizaje se debe realizar considerando las mejores evidencias de sus propiedades psicométricas.” (p. 360), algo parcialmente corregido en la guía de práctica clínica más reciente (MINSA, 2023).

Es por esto es que la revisión sistemática (RS) de la literatura es importante (Villasís-Keever et al., 2020), sobre todo respecto a la medición de la depresión, ya que permite conocer cuáles son los instrumentos más óptimos para la cuantificación de sus síntomas en la población peruana y que los estamentos gubernamentales y profesionales clínicos puedan establecer adecuados estándares de evaluación clínica, de acuerdo con los criterios de cientificidad propios de las disciplinas de la salud mental.

Debido al contexto problemático anteriormente descrito, la presente investigación formula la siguiente interrogante de investigación: ¿Cuáles son las características psicométricas de los instrumentos para medir depresión validados en población peruana?

1.2. Antecedentes

1.2.1. Antecedentes internacionales

Bjelland et al. (2002) realizaron una revisión de la literatura sobre las validaciones de la Escala Hospitalaria de Ansiedad y Depresión (HADS) a través de investigaciones ubicadas en Medline, ISI y PsycINFO hasta mayo del 2000. Se identificaron 1403 artículos en la búsqueda inicial, 71 artículos después de la depuración. Los resultados revelaron la adecuación de la estructura bifactorial de la operacionalización original, con correlaciones interdimensiones entre .4 y .74, coeficientes de alfa de Cronbach entre .68 y .93 para ansiedad y .67 y .90 para depresión, con una sensibilidad y especificidad alrededor de .80, la validez de criterio convergente fue de .49 y .83. Se concluye que la HADS es un instrumento con adecuadas propiedades psicométricas en pacientes clínicos y no clínicos.

Cabañero-Martínez et al. (2007) evaluaron las propiedades psicométricas y criterios de aplicación de pruebas para medir depresión en adultos mayores españoles (≥ 55 años) mediante una revisión sistemática en MEDLINE, PsyInfo, CINAHL y Embase, identificando 30 artículos que analizaron 6 escalas, aunque no se reportó el año límite de la búsqueda. Hallaron que la Escala de Depresión Geriátrica (GDS) mostró las mejores propiedades y mayor

bibliografía, pero el uso de múltiples versiones dificulta un análisis unificado. En contraste, las demás escalas presentaron adaptaciones culturales y lingüísticas deficientes, con reportes escasos y estándares poco adecuados; además, los resultados fueron heterogéneos y poco detallados, limitando un análisis global.

González-Forteza et al. (2012) llevaron a cabo una investigación a fin de determinar la cantidad de estudios sobre la Escala de Depresión del Centro de Estudios Epidemiológicos (CESD) llevados a cabo en México en el período de años comprendido entre 1986 hasta julio de 2011, en bases de datos como EBSCOhost y Google Académico. Se culminó con 80 artículos publicados mayormente en las revistas Salud Mental (27), Salud Pública de México (11) y la Revista Mexicana de Psicología (5), de los cuales, el 36% fueron realizadas en estudiantes, 33% en población general, 20% en pacientes, y el resto en pacientes-general y trabajadores. No se detallaron las propiedades psicométricas de las validaciones realizadas, no se enlistaron las investigaciones encontradas y sus características individuales.

Wang y Gorenstein (2013) revisaron estudios sobre las propiedades psicométricas del Inventario de Depresión de Beck-II (BDI-II) publicados en MEDLINE y PsycINFO entre 1996 y octubre de 2012, sin restricción poblacional de idioma ni edad. De 2611 registros iniciales, incluyeron 118 estudios, reuniendo evidencias de validación en más de 60 000 participantes. El BDI-II fue traducido a 17 idiomas y se usó principalmente en inglés (65%), mostrando alfa de Cronbach promedio de .90, confiabilidad test-retest (r) entre .73 y .96, correlaciones ítem-test $> .50$ en la mayoría de los ítems, validez convergente entre .66 y .86 y sensibilidad $> .70$; no obstante, las adaptaciones interculturales evidenciaron variaciones en la validez de constructo y en varios ítems (hasta 12 en la versión turca), sugiriendo diferencias culturales entre versiones.

Terol-Cantero et al. (2015) realizaron una revisión sistemática sobre las propiedades psicométricas de la Escala de Ansiedad y Depresión Hospitalaria (HAD) en población

española, a partir de artículos de ScienceDirect, Medline, PsycInfo (WOK), Psycodoc e ISOC, aunque no se reportó el periodo temporal de búsqueda. De 2384 registros iniciales, se incluyeron 15 tras la depuración, con validaciones mayormente en muestras >300 participantes, principalmente clínicas. En general, se halló consistencia interna $>.80$ para ansiedad y depresión, correlaciones ítem-resto aceptables, confirmación frecuente de la estructura bifactorial, varianza explicada entre 46% y 60% y sensibilidad/especificidad $>70\%$ en la mayoría; por ello, se concluye que la HAD presenta amplia evidencia de validación en España y propiedades psicométricas adecuadas.

Vega-Dienstmaier (2022) realizó una RS sobre las características y propiedades psicométricas de instrumentos en español para medir trastornos mentales comunes (depresión y ansiedad) en PubMed y Google Scholar, hasta octubre de 2021. El estudio no precisa el número total de artículos recuperados, pero analiza 37 investigaciones. El General Health Questionnaire (GHQ) fue el más utilizado en cuatro países (España, Colombia, Argentina y Perú), con valores $\alpha > .829$ en la versión de 28 ítems; le siguió la Hospital Anxiety and Depression Scale (HADS), validada en España, Colombia, Chile y México, con coherencia interna $> .76$ para ansiedad y $> .83$ para depresión, además de validez convergente y discriminante; en tercer lugar, se ubicó la Escala de Malestar Psicológico de Kessler en Argentina, México y Perú, con $\alpha > .88$.

1.2.2. Antecedentes nacionales

Navarro-Loli y Moscoso (2017) realizaron una revisión sistemática sobre investigaciones de depresión en adolescentes peruanos publicadas en Scielo, Latindex, Redalyc, Scopus, PubMed y Google entre 2000 y 2016 (en español e inglés). Tras identificar 31 artículos, incluyeron 21 luego de la depuración, hallando que menos de la mitad reportó el diseño de estudio (43%), que más del 50% se desarrolló en Lima y el norte del Perú, y que solo 43% estimó la confiabilidad de los instrumentos. Asimismo, concluyeron que la adolescencia

constituye una población vulnerable a la depresión (con mayor riesgo en mujeres) y que la ansiedad fue la variable más asociada; sin embargo, señalaron heterogeneidad en los reportes, incumplimiento frecuente de estándares APA y el uso de diversos instrumentos sin validación en la población objetivo.

Cjuno et al. (2021) realizaron una revisión narrativa sobre investigaciones de depresión en poblaciones quechua hablantes, publicadas en PubMed, Web of Science y Scopus hasta junio de 2021. De 24 registros iniciales, se incluyeron 7 artículos; todos emplearon instrumentos psicométricos, pero solo uno incorporó validación instrumental (HSCL-25). La evidencia provino principalmente de Perú (5 estudios) y, en menor medida, de Ecuador, y se reportó mayor intensidad de síntomas depresivos en mujeres; además, se describió el denominado *llaqui* como un constructo culturalmente cercano a la depresión, con alta concordancia con criterios DSM-III-R en un subgrupo evaluado. En conjunto, los autores subrayan que la investigación instrumental sobre depresión en quechua hablantes es escasa y que se requiere operacionalizar y adaptar culturalmente el *llaqui* como un problema clínicamente evaluable.

Cjuno et al. (2022) revisaron estudios psicométricos realizados en Perú sobre validación de escalas breves o tamizajes para depresión, mediante una revisión sistemática en Scopus, Web of Science, PubMed y SciELO (1993 a 2022). De 537 investigaciones, incluyeron 22 y hallaron que la escala más validada fue la Patient Health Questionnaire (PHQ-9) (9 estudios), seguida por la Edinburgh Postnatal Depression Scale (EPDS), el Inventario de Depresión de Beck (BDI-II) y el State-Trait Depression Inventory (ST-DEP) (2 estudios cada una). La mayoría de las validaciones se realizó en Lima (13) y por la Universidad Peruana Cayetano Heredia (5), en adultos, mujeres adultas y gestantes, pacientes con depresión y universitarios; sin embargo, se reporta limitada especificación de los análisis empleados (p. ej., validez,

fiabilidad, sensibilidad/especificidad e invariancia) y de los valores obtenidos, lo que dificulta juzgar su adecuación para las poblaciones estudiadas.

1.3. Objetivos

1.3.1. Objetivo general

- Analizar las características psicométricas de los instrumentos para medir la depresión validados en población peruana.

1.3.2. Objetivos específicos

- Identificar los instrumentos para medir la depresión validados en población peruana y las características de las muestras en las que fueron aplicados.
- Describir los procedimientos de adaptación cultural de los instrumentos para medir depresión validados en población peruana.
- Describir la evidencia de confiabilidad de los instrumentos para medir depresión validados en población peruana.
- Describir la evidencia de validez de los instrumentos para medir depresión validados en población peruana.
- Describir las normas de las puntuaciones de los instrumentos para medir depresión validados en población peruana.

1.4. Justificación

La presente investigación aporta al ODS 3: Salud y bienestar propuesto por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, 2019), porque fortalece la comprensión conceptual de la depresión como problema prioritario de salud mental y permite identificar qué instrumentos psicométricos la operacionalizan con mayor coherencia en la población peruana. Su valor teórico radica en que integra y compara críticamente las bases conceptuales, las dimensiones evaluadas y los criterios de severidad empleados por las escalas validadas, lo que ayuda a precisar qué medida representa mejor el

constructo en el contexto nacional. En consecuencia, la investigación no solo ordena la evidencia disponible, sino que también clarifica cómo se ha entendido y medido la depresión en el Perú, contribuyendo a una discusión teórica más sólida y contextualizada.

Metodológicamente, esta revisión sistemática fortalece el ODS 3 al reunir y valorar evidencia psicométrica que permite decidir qué instrumentos ofrecen resultados más confiables, válidos y comparables para la evaluación de depresión. Su ganancia específica consiste en sintetizar indicadores como validez de constructo, estructura factorial, confiabilidad, puntos de corte y sensibilidad/especificidad, facilitando la identificación de medidas con mejor respaldo empírico para su uso en investigación y evaluación clínica. Así, el estudio aporta un criterio metodológico más riguroso para seleccionar escalas y evita que futuros trabajos utilicen instrumentos con evidencia dispersa, insuficiente o poco pertinente para la realidad peruana.

En el plano práctico, la revisión ofrece una base útil para que profesionales de salud mental, investigadores y equipos asistenciales elijan instrumentos ya validados según el tipo de población y el contexto de aplicación, lo que favorece el tamizaje, la estimación de severidad y el seguimiento de la depresión en escenarios reales. Su aporte concreto es convertir una evidencia dispersa en una guía sintetizada que facilita decisiones clínicas e institucionales más informadas, especialmente en atención primaria, hospitales, universidades y otros espacios de detección temprana. De este modo, la investigación contribuye a mejorar la calidad de la evaluación y a promover un uso más consistente y pertinente de las escalas disponibles.

Socialmente, el estudio se vincula con el ODS 3 porque una medición adecuada de la depresión es clave para reconocer necesidades reales de la población, priorizar recursos y orientar acciones de prevención y atención en salud mental. La ganancia de esta revisión es que puede reducir el uso de instrumentos poco adecuados o mal contextualizados, disminuyendo el riesgo de subestimar o sobreestimar el problema en distintos grupos peruanos. Al sintetizar la evidencia disponible, la investigación también puede servir como insumo para

lineamientos, políticas y programas de salud mental más ajustados a la realidad nacional, con mayor potencial de impacto en bienestar y acceso oportuno a intervenciones.

II. MARCO TEÓRICO

2.1. Bases teóricas sobre el tema de investigación

2.1.1. Definición y concepto de la depresión

Existe un consenso generalizado que establece, de manera más o menos homogénea, las características básicas de la depresión (Caballo et al., 2014) como un estado anímico de afectación ponderado por tristeza y bajo interés general, con reducción de las actividades físicas y psicofisiológicas, que afectan tanto a factores volitivos como vivenciales; a nivel clínico, este cuadro se expresa mediante un patrón persistente de síntomas emocionales, cognitivos, conductuales y somáticos, que interfiere con el funcionamiento personal, social y ocupacional, y se asocia a deterioro de la calidad de vida.

En este sentido, se destaca que el principal síntoma de la depresión es la tristeza, entendida como estado de ánimo negativo, decaimiento anímico, dolor emocional o estado de aflicción; dicho estado de ánimo no es necesariamente patológico y es inherente al vivir cotidiano de los seres humanos, especialmente en situaciones traumáticas o de pérdida (Coto-Lesmes et al., 2020). Cuando la tristeza sobrepasa las capacidades psicológicas del individuo para afrontarla y las estrategias para retomar el equilibrio emocional, dicho estado ánimo se torna patológico y es posible entrar en depresión (Parada et al., 2022). Es por lo que la tristeza es un estado de ánimo, mientras que la depresión es un trastorno del estado de ánimo (Caballo et al., 2014).

Esto se ve reflejado en lo que, a su vez, menciona la *American Psychological Association* (APA, 2023), definiendo la *depresión* como:

A negative affective state, ranging from unhappiness and discontent to an extreme feeling of sadness, pessimism, and despondency, that interferes with daily life. Various physical, cognitive, and social changes also tend to co-occur, including altered eating or sleeping habits, lack of energy or motivation, difficulty concentrating or making

decisions, and withdrawal from social activities. It is symptomatic of a number of mental health disorders. (p. 1).

Dentro de esta definición podemos notar las características de la depresión ya mencionadas (predominio de la tristeza e incapacidad de afrontamiento), adicionando la presencia de gradualidad en este trastorno, pudiendo variar desde la infelicidad y el descontento, hasta el predominio en grado extremo de la tristeza y la sensación de pesimismo y abatimiento, lo que revela la alta complejidad de este problema (Cala y Agudelo, 2018; Montes, 2004).

Así mismo, se destaca la interferencia con la vida diaria, factor altamente relevante al momento de considerar a estos sentimientos como un problema potencial y no como un síntoma pasajero que se expresa de manera momentánea, sin afectar significativamente otras áreas de la vida de las personas (Pérez-Padilla et al., 2017).

Por último, se destaca la presencia de la depresión en otros trastornos, característica crucial de ser considerada, debido a las altas tasas de comorbilidad que presenta la depresión con otros desórdenes, ya sea de forma principal (a modo de problemática con cierta independencia) o secundaria (como consecuencia aparente de otro trastorno) (Aguado y Del Álamo, 2020; Alonso et al., 2019; Aragonès et al., 2009; García-Batista et al., 2019; González et al., 2011).

Sin embargo, cabe recalcar que la tristeza no es el único síntoma que conforma la depresión, sino el más típico o característico, tal y como menciona la APA (2023) al definir el *trastorno depresivo* como “any of the mood disorders that typically have sadness as the predominant symptom. They primarily include major depressive disorder and dysthymic disorder.” (p. 1), lo que destaca nuevamente a la tristeza, pero la enfatiza como síntoma predominante, no como el único, y también nos introduce a la discusión de los trastornos depresivos, como categoría plural y multidimensional.

2.2.1. Clasificación y tipos de depresión según criterios diagnósticos

Es así como dicho trastorno está conformado por un conjunto de síntomas que pueden variar según el sistema de clasificación y diagnóstico empleado. Los principales sistemas están proporcionados por el *Manual diagnóstico y estadístico de los trastornos mentales* en su quinta edición (DSM-5), la *Clasificación Internacional de Enfermedades* en su décima edición (CIE-10) y, con menos difusión, pero más recientemente, la décimo primera edición de la última mencionada (CIE-11).

2.2.1.1. Manual diagnóstico y estadístico de los trastornos mentales en su quinta edición (DSM-5). Primeramente, en el DSM-5, de la *American Psychiatric Association* (APA, 2014), encontramos a los *trastornos depresivos*, los cuales agrupan al trastorno de desregulación disruptiva del estado de ánimo, trastorno de depresión mayor, trastorno depresivo persistente (distimia), trastorno disfórico premenstrual, trastorno depresivo inducido por sustancia/medicamento, trastorno depresivo debido a otra afección médica, otro trastorno depresivo especificado y otro trastorno depresivo no especificado. Dichos trastornos han sido separados del *trastorno bipolar y relacionados*, que típicamente se consideraban dentro de los *trastornos del estado del ánimo*, cúmulo que sí estaba considerado dentro de la cuarta edición de dicho manual (DSM-IV) (APA, 1994). Recordemos que la codificación alfanumérica para estos y todos los trastornos dentro del DSM-5 es la misma que la catalogada en la CIE-10 debido a temas de estandarización.

Esta ha sido considerada la mejor clasificación de la depresión, debido a una mayor especificidad y operativización de la misma, así como a un mayor sustento empírico con descripciones más detalladas y extensas frente a la CIE-10 (Caballo et al., 2014).

2.2.1.2. Clasificación Internacional de Enfermedades en su décima edición (CIE-10). Por otro lado, la CIE-10, de la OMS (2000), clasifica a la depresión dentro de los *trastornos del humor (afectivos)*. Entre dichos trastornos encontramos al episodio maníaco (F30),

trastorno bipolar (F31), episodios depresivos (F32), trastorno depresivo recurrente (F33), trastornos del humor (afectivos) persistentes (F43), otros trastornos del humor (afectivos) (F38) y trastorno del humor (afectivo) sin especificación (F39). En esta clasificación aún encontramos la conjunción de los *episodios depresivos* con el *episodio maníaco* y, por ende, con el *trastorno bipolar*.

Actualmente, este es el sistema de referencia universal dentro del ámbito peruano para la evaluación y diagnóstico de la depresión, debido a que el MINSA (2008, 2021) oficializa su uso como sistema de clasificación estandarizada a nivel nacional.

2.2.1.3. Clasificación Internacional de Enfermedades en su undécima edición (CIE-11). Finalmente, un sistema más actual y de gran referencia es la CIE-11, aprobada por la OMS (2019) en la 72° reunión de la Asamblea Mundial de la Salud. Esta clasificación retoma la categoría *trastornos del estado de ánimo*, incluyendo a los trastornos bipolares u otros trastornos relacionados (6A6), trastornos depresivos (6A7), presentaciones sintomáticas y de curso para episodios del estado de ánimo en trastornos del estado de ánimo (6A80), otros trastornos del estado de ánimo especificados (6A8Y) y trastornos del estado de ánimo, sin especificación (6A8Z).

Actualmente, este sistema de clasificación no es considerado el estándar para el diagnóstico de la depresión en el Perú y en la mayoría de los países del mundo, debido a su reciente elaboración y publicación, situación que puede mantenerse hasta que las autoridades de salud determinen el cambio a este nuevo sistema, según la Organización Panamericana de la Salud (OPS, 2019).

Es debido a esta gran diversidad de clasificaciones, tipos y subtipos de depresión o de trastornos relacionados a la misma que, históricamente, algunos autores han decidido denominarla como *las depresiones* durante sus disquisiciones o clasificaciones terminológicas (Polaino-Lorente y Ascaso, 1988; Retamal, 1998; Tamayo et al., 2011)

2.3.1. Principales aspectos de la evaluación psicométrica de la depresión

La depresión es un trastorno complejo con múltiples manifestaciones, las cuales pueden generar cierto grado de disonancia entre los criterios utilizados para evaluarla, esta situación debe ser tomada en cuenta cuando se elabora un instrumento psicométrico para medirla, si es que se desea cuantificarla de la forma más eficiente posible.

2.3.1.1. Anomalías diagnósticas. Por ejemplo, la presentación de síntomas con manifestaciones tanto negativas (por defecto) como positivas (por exceso) dentro de las depresiones conocidas como *atípicas* o *paradógicas* (Caballo et al., 2014) puede generar inconvenientes en la evaluación de la depresión, dejando de lado una importante muestra de pacientes depresivos si es que no se consideran dichos síntomas. Estos indicadores se han denominado anomalías diagnósticas y se presentan dentro de los siguientes criterios diagnósticos:

Actividad (CIE-10, episodio depresivo leve, criterio C5; DSM-5, trastorno de depresión mayor, criterio A5): Alteraciones en la actividad psicomotora, manifestada a través de dos polos, *agitación* (exceso de actividad, inquietud) o *retraso* (enlentecimiento) psicomotor. Los pacientes presentan alteraciones de su estado normal de actividad psicomotora durante los episodios depresivos, pudiendo evidenciar letargo, enlentecimiento y una marcada disminución de la actividad, frente a otros que muestran síntomas completamente opuestos, tales como un exceso de actividad, gran inquietud y activación psicomotora. Esta última forma del síntoma puede enmascarar la presencia del trastorno depresivo frente a familiares, amigos y personal sanitario no entrenado. La mayor diferencia entre los manuales diagnósticos es que la CIE-10 admite que dicho cambio en la actividad pueda ser objetiva (referida por otros) o subjetiva (percibida por uno mismo), mientras que el DSM-5 solo admite la manifestación objetiva, descartando la sensación subjetiva como único síntoma aceptable.

Sueño (CIE-10, episodio depresivo leve, criterio C6; DSM-5, trastorno de depresión mayor, criterio A4): Otro síntoma propio de la depresión es la alteración del patrón de sueño, manifestado a través de dificultades para conciliar el sueño o para volver a dormir cuando se han despertado a mitad o antes de cumplir las horas acostumbradas (*insomnio* inicial, medio o tardío). Por otro lado, existen pacientes que manifiestan un gran deseo de dormir o excesiva somnolencia en el día y/o noche (*hipersomnia* diurna o nocturna), con gran facilidad para conciliar el sueño en lugares insólitos y dificultad para despertarse.

Apetito (CIE-10, episodio depresivo leve, criterio C7; DSM-5, trastorno de depresión mayor, criterio A3): Otro síntoma paradójico de la depresión es la alteración del apetito. Ésta puede presentarse como el *aumento* anormal del apetito, impulsando a los pacientes con depresión a comer grandes cantidades de comida, sobre todo de hidratos de carbono; mientras, en el polo opuesto, otros manifiestan una marcada disminución o *pérdida* del apetito, con evitación o restricción de las situaciones que involucran el consumo de alimentos. Consecuentemente, tanto el aumento o pérdida del apetito generan un drástico aumento o pérdida del peso, sin embargo, esto último está en controversia ya que cada vez se diagnostica a los pacientes con depresión con mayor prontitud y en muchos casos no llegan a manifestar pérdida o aumento significativo de peso debido a los efectos terapéuticos benéficos, por lo que no funcionaría como un criterio diagnóstico tan relevante como otros (Beck et al., 2006).

Todas estas manifestaciones pueden presentar síntomas considerados como inversos a los esperados en pacientes depresivos típicos, teniendo como base el supuesto de que la depresión genera el enlentecimiento psicobiológico del individuo, reduciendo drásticamente su actividad tanto cognitiva, emocional y motora (Caballo et al., 2014). Dicho aspecto atípico de la sintomatología depresiva puede generar dificultades al diseñar escalas para medir este trastorno, ya que se provoca que se deban considerar reactivos con un solo síntoma, pero con respuestas tanto positivas (por exceso de síntoma) como inversas (por defecto de síntoma). Por

ejemplo, para un ítem orientado a medir una alteración del sueño se debe formular una pregunta o afirmación que identifique insomnio o hipersomnia, algo que puede generar dificultades o confusión.

Si bien esta característica es poco considerada en la medición psicométrica específica de la depresión, existen instrumentos como el Inventario de depresión de Beck en su segunda edición (BDI-II) (Beck et al., 2006), que buscan superar esta dificultad. Es así como la primera edición de la escala:

...se refería a la pérdida de apetito pero no a su aumento. Sin embargo, algunas personas con depresión, especialmente aquellas con depresiones “atípicas”, experimentan incremento del apetito. Existe una situación similar con respecto a la variación de sueño; ya que algunos pacientes con depresión duermen considerablemente más de lo habitual. Moran y Lambert (1983) también notaron que el BDI carecía de ítems explícitos referidos a la agitación y la actividad psicomotriz, a pesar de que los criterios del DSM-III incluyen a ambas en el diagnóstico de Trastorno Depresivo Mayor. (p. 44).

Es por eso que el manual de la segunda edición de la escala menciona dichas falencias y las corrige, reformulando los ítems de *cambios en los hábitos del sueño* (ítem 16) y *cambios en el apetito* (ítem 18) para cuantificar tanto el aumento como la disminución de dichas necesidades fisiológicas, aunque el criterio de *actividad* no se formula de la misma manera, manteniendo el ítem sobre *cansancio o fatiga* (ítem 20) y agregando el de *agitación* (ítem 11) que a primera vista pueden considerarse opuestos. Es por esta característica de la depresión que los instrumentos para medirla deben ser complejos y correctamente elaborados, tomando en cuenta esta situación, e interpretados únicamente por profesionales entrenados en la evaluación de la depresión.

2.3.1.2. Sensibilidad de las pruebas de tamizaje. Todo instrumento psicológico intenta optimizar la sensibilidad (s) y especificidad (e) de sus puntuaciones (Tornimbeni et al.,

2008), a fin de identificar con más precisión aquellos con mayores o menores probabilidades de padecer depresión según los criterios diagnósticos. Es por ello que ambos estadísticos deben mantener cierto grado de equilibrio, a fin de no provocar falsos positivos (alta sensibilidad, baja especificidad) o falsos negativos (baja sensibilidad, alta especificidad) (Beck et al. 2006).

Sin embargo, en el caso puntual de la medición psicométrica de la depresión se debe tomar en cuenta que será preferible identificar con más ahínco a los casos positivos que a los negativos, debido a la necesidad clínica de priorizar la atención de los pacientes con depresión o posibilidad de la misma. Esta situación provoca que los instrumentos enfocados a medir la depresión deban priorizar la sensibilidad frente a especificidad, lo que puede facilitar la presentación de falsos positivos y dificultar la de falsos negativos, situación que, realmente, es deseable, tal y como expresan Beck et al. (2006):

Con el propósito de desarrollar un instrumento que permita identificar depresión mayor con fines clínicos (*screening*), se consideró más importante la sensibilidad del test que su especificidad. Es decir, el clínico querrá adoptar un umbral más bajo para detectar depresión a fin de disminuir la probabilidad de cualquier falso negativo (p. 54).

Esta situación, si bien puede generar falsos positivos, también permitirá que más personas con indicios de depresión puedan ser identificados y monitorizados con prontitud, pudiendo implementar programas preventivos en aquellos que lo requieran, antes de manifestar un trastorno depresivo como tal.

Es por esto que los instrumentos cuantitativos sobre esta variable deben tomar en cuenta una mayor sensibilidad que especificidad dentro de la evaluación clínica de la depresión ya que, si bien esto redundará en una mayor identificación de casos positivos y, por ende, en mayores esfuerzos para la diferenciación y descarte diagnóstico debido a una mayor cantidad de falsos positivos (Tornimbeni et al., 2008), también redundará en mayores beneficios para los pacientes, que podrán ser identificados oportuna y, tal vez, preventivamente, diferente

situación a la encontrada en instrumentos destinados a la investigación, en la que se requiere identificar grupos diagnósticos más puros, con una diferenciación lo más exacta posible entre depresivos y no depresivos, a fin de establecer descripciones y análisis más fieles de la realidad epidemiológica (Beck et al., 2006).

III. MÉTODO

3.1. Tipo de investigación

Una revisión sistemática implica el diseño y la ejecución de un procedimiento explícito, ordenado y reproducible orientado a identificar, seleccionar, evaluar críticamente y sintetizar la evidencia disponible sobre una temática o variable específica (Villasís-Keever et al., 2020). En tal sentido, se considera una forma de investigación secundaria, en tanto integra estudios primarios que han analizado directamente la realidad mediante la recolección y el procesamiento de datos de primera mano (Driscoll, 2011).

De acuerdo con Ato et al. (2013), la presente pesquisa se inscribe en un diseño teórico, ya que se orienta a recopilar y sistematizar avances metodológicos vinculados con una problemática de investigación determinada. Este tipo de diseño comprende revisiones y actualizaciones de la literatura que no emplean datos empíricos originales, los cuales provienen de investigaciones primarias, sino que se sustentan en la valoración y síntesis de la evidencia previamente publicada.

Las revisiones narrativas, revisiones sistemáticas y metanálisis comparten similitudes, pero también diferencias. Una revisión narrativa es no estructurada, no sistemática, no sigue un protocolo de búsqueda y selección determinado y tiene una alta probabilidad de sesgo, mientras que la revisión sistemática compensa todos estos puntos débiles utilizando una estructura o sistema, con un protocolo de búsqueda y selección específico, con mucha menor probabilidad de sesgo. Por otro lado, un metanálisis utiliza todas las ventajas de una revisión sistemática (la toma como base), pero condensa los resultados utilizando técnicas de análisis estadísticos que arrojan resultados cuantitativos e integran dichos resultados a fin de establecer estadísticos globales, es decir, todo metanálisis tiene como requisito previo una revisión sistemática (Villasís-Keever et al., 2020).

3.2. Ámbito temporal y espacial

La presente investigación se llevó a cabo entre los meses de enero de 2026 a abril de 2026, estableciendo la primera toma de artículos durante la segunda semana de enero y realizando el análisis y depuración de los artículos durante los meses del periodo señalado, utilizando bases de datos tales como Scopus, Web of Science, Scielo, PubMed, Biblioteca Virtual en Salud (BVS) y Google Scholar.

3.3. Variables

3.3.1. Medición de la depresión

La medición de la depresión se conceptualiza como el proceso de cuantificar y evaluar los síntomas y la gravedad de la depresión en una persona a través del uso de herramientas psicométricas, tales como escalas o cuestionarios específicos diseñados para evaluar el estado emocional, mental y conductual de un individuo en relación con la depresión (Caballo et al., 2014; Moscoso, 2014).

Se operacionaliza la medición de la depresión a través de la rúbrica de adaptaciones y validaciones de instrumentos psicométricos que considera los 5 aspectos fundamentales para la adecuada medición psicométrica de dicha variable: adaptación cultural, confiabilidad, validez, y normas (AERA, APA y NCME, 2018; Luján-Tangarife y Cardona-Arias, 2015; Muñoz et al., 2013).

Tabla 1*Operacionalización de la variable medición de la depresión*

Criterio	Propiedad	Definición
Adaptación cultural	Etapas 1: Traducción directa al idioma de interés	Traducción del idioma de origen al idioma de adaptación
	Etapas 2: Síntesis de traducciones y solución de discrepancias	Análisis de las traducciones anteriores y selección de formatos consensuados
	Etapas 3: Traducción inversa (idioma destino - idioma original)	Retrotraducción del idioma de adaptación al idioma de origen
	Etapas 4: Revisión por comité de expertos	Análisis y selección de formatos finales
	Etapas 5: Prueba piloto	Análisis psicométrico previo al central, recolección de impresiones previas y corrección de posibles inconvenientes de aplicación
Confiabilidad/ precisión	Coefficientes de formularios alternativos	Correlación entre dos versiones similares del mismo instrumento
	Coefficientes test-retest	Correlación entre diferentes aplicaciones de un instrumento
	Coefficientes de coherencia interna	Correlación entre los ítems que componen un instrumento
	Índices de coherencia entre evaluadores	Acuerdo entre los resultados obtenidos por diferentes evaluadores
Validez	Evidencia basada en el contenido de la prueba	Análisis de la relación entre el contenido de la prueba y el constructo que se intenta medir
	Evidencia basada en los procesos de respuesta	Análisis de los procesos cognitivos empleados por los evaluados para responder la prueba
	Evidencia basada en la estructura interna	Análisis del grado en que las relaciones entre ítems y componentes se ajustan al constructo
	Evidencia basada en relaciones con otras variables	Análisis de la relación de los puntajes de la prueba con variables externas a la prueba
	Evidencia basada en las consecuencias de las pruebas	Análisis de las implicaciones prácticas y consecuencias del uso de la prueba
Normas	Sensibilidad (s) y especificidad (e)	Capacidad del instrumento de detectar los casos positivos y negativos
	Baremación	Establecimiento de grados o niveles que expresen la intensidad del problema o rasgo cuantificado

3.4. Población y muestra

En la presente investigación se recolectaron todos los documentos publicados sobre la medición psicométrica de la depresión validados en población peruana, dichos elementos serán identificados en las bases de datos señaladas según el criterio de búsqueda formulado a fin de localizar los títulos y palabras clave más relevantes en los artículos. Luego de dicha primera selección inicial, se realizó una depuración en la que se tomarán como criterios de inclusión los siguientes enunciados para todas las investigaciones halladas; como criterios de exclusión, la carencia de uno de más de ellos:

- Investigaciones psicométricas sobre instrumentos que midan depresión.
- Que dichos instrumentos sean cuestionarios autoinformativos.
- Artículos científicos publicados en revistas indexadas.
- Estudios realizados en población peruana de cualquier edad, que resida en Perú al momento de realizar la investigación.
- Estudios primarios que traten con datos directos.
- Estudios cuantitativos o mixtos que muestren los datos psicométricos de forma directa.
- Estudios que reporten como mínimo 2 análisis diferentes (p. e., confiabilidad y validez, confiabilidad y sensibilidad/especificidad, etc.) dentro de la Teoría Clásica de los Test (TCT).
- Redactados en español o inglés.
- Publicados hasta enero de 2026 o antes.

No se tomaron en cuenta criterios relacionados a la antigüedad de los datos, la calidad de la investigación o los estándares nacionales o internacionales empleados. En caso de encontrarse deficiencias metodológicas o técnicas en las investigaciones, serían analizadas y comparadas en la discusión.

3.5. Instrumento

3.5.1. Rúbrica de adaptaciones y validaciones de instrumentos psicométricos

Para la presente investigación se ideó una rúbrica, a fin de clasificar, describir y analizar las propiedades psicométricas obtenidas de las adaptaciones y validaciones que se realicen de los instrumentos (Tabla 2), siguiendo las pautas otorgadas por los estándares de la *American Educational Research Association*, *American Psychological Association* y *National Council on Measurement in Education* (AERA, APA y NCME, 2018) y las directrices de la *International Test Commission* (ITC, 2013). Se ideó una rúbrica, siguiendo el acercamiento de Becker-Haimes et al. (2020) y De Los Reyes y Langer (2018) quienes utilizan una rúbrica desarrollada por Hunsley y Mash (2008) y extendida por Youngstrom et al. (2017), pero sin utilizar criterios calificativos que impidan una descripción extensa de los estadísticos encontrados.

Tabla 2

Rúbrica de adaptaciones y validaciones de instrumentos psicométricos

Criterio	Propiedad
Adaptación cultural	Etapas 1: Traducción directa al idioma de interés
	Etapas 2: Síntesis de traducciones y solución de discrepancias
	Etapas 3: Traducción inversa (idioma destino - idioma original)
	Etapas 4: Revisión por comité de expertos
	Etapas 5: Prueba piloto
Confiabilidad/precisión	Coefficientes de formularios alternativos
	Coefficientes test-retest
	Coefficientes de coherencia interna
	Índices de coherencia entre evaluadores
Validez	Evidencia basada en el contenido de la prueba
	Evidencia basada en los procesos de respuesta
	Evidencia basada en la estructura interna
	Evidencia basada en relaciones con otras variables
	Evidencia basada en las consecuencias de las pruebas
Normas	Sensibilidad (s) y especificidad (e)
	Baremación

La rúbrica cuenta con 4 criterios, conformados por los procesos básicos y fundamentales para determinar la calidad de una adaptación y validación psicométrica, aunque también puede utilizarse para analizar la presencia de adecuados criterios durante la construcción de un instrumento. Los criterios que la conforman son:

Adaptación cultural: Refleja todas las etapas que debe seguir una investigación que desea adaptar un instrumento psicométrico que ha sido diseñado y validado originalmente en territorio y población extranjeros o locales, pero con idiomas y características socioculturales diferentes. Abarca la traducción, retrotraducción, selección de traducciones y ejecución de prueba piloto. Dichos componentes son tomados de Luján-Tangarife y Cardona-Arias (2015) y Muñiz et al. (2013), quienes explicitan los procesos más utilizados y adecuados para esta etapa de una investigación psicométrica.

Confiabilidad: Refleja la presencia de procedimientos para evaluar la coherencia y estabilidad de los resultados que produce un instrumento psicométrico a través del tiempo y en diferentes situaciones. Se recurre a psicométricos tales como coeficientes de formularios alternativos, coeficientes test-retest, coeficientes de coherencia interna e índices de coherencia entre evaluadores. Tales coeficientes son los establecidos por la AERA, APA y NCME (2018), como parte de sus estándares.

Validez: Refleja la presencia de procedimientos para evaluar la variable objetivo de forma verdadera y precisa, con una adecuada construcción del modelo, a través de factores y reactivos pertinentes de acuerdo con la teoría subyacente a la construcción de la escala. Como tal, la validez es unitaria y no posee tipos, mas sí posee fuentes de evidencia, las cuales son evidencia basada en el contenido de la prueba, evidencia basada en los procesos de respuesta, evidencia basada en la estructura interna, evidencia basada en relaciones con otras variables y evidencia basada en las consecuencias de las pruebas. Dichas fuentes son las establecidas y recomendadas por la AERA, APA y NCME (2018), como parte de sus estándares. No se utilizó

la clásica división tripartita que la divide en validez de contenido, criterio y constructo, llamada también la doctrina trinitaria de la validez o la santísima trinidad (Guion, 1980), por ser rechazada y abandonada por los estándares a partir de 1999 (Bademci, 2022).

Normas: Refleja la presencia de procedimientos para evaluar la capacidad de un instrumento psicométrico de diferenciar entre los individuos que poseen un trastorno o característica psicológica, de aquellos que no la poseen. Implica el uso de estadísticos de sensibilidad (s) y especificidad (e), así como sus modificaciones a través del índice de Youden (J) y área bajo la curva Relative Operating Characteristic (abcROC). Así mismo, implica la presencia de baremos que expresen los niveles, grados o cualquier tipo de intensidad del problema o característica evaluada, expresando de forma cuantitativa pero gradual qué tan grave es un problema o qué tan presente está un rasgo determinado. Implica el uso de percentiles, cuartiles, eneatis. puntuaciones Z, entre otros.

Se construye la rúbrica anteriormente mencionada siguiendo las recomendaciones Cochrane (Deeks et al., 2023) sobre la elaboración de revisiones sistemáticas de precisión de pruebas diagnósticas, en las que se especifica que una adecuada recopilación de datos debe contar con un formulario o sistema que pueda representar fielmente la estructura y organización de la información compilada.

3.6. Procedimientos

Primeramente, se utilizó el método PICO para identificar los elementos que compondrán la estrategia de búsqueda, adaptado de intervenciones a investigaciones instrumentales, de la siguiente forma (Higgins et al., 2024):

P (población): peruanos

I (intervención): evaluación de la depresión

C (comparación): sin comparación

O (outcomes/resultados): características psicométricas

En este mismo sentido, puede reinterpretarse este método bajo el acrónimo PIT, en cual se establece los elementos de P (población), I (*Index test*) y T (*Target condition*), que, aunque ideado para revisiones sistemáticas de test, viene a ser una alternativa derivada directamente del método PICO en la que solamente se ha retirado la C (comparación) (Deeks et al., 2023).

Se realizó una selección de los términos más utilizados para el estudio de la depresión, tanto en los manuales de diagnóstico DSM 5 (APA, 2014), CIE 10 (OMS, 2000) y CIE 11 (OMS, 2019), como en el tesoro multilingüe DeCS/MeSH, dirigido por el Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud, la Organización Panamericana de la Salud y la Organización Mundial de la Salud (BIREME, OPS y OMS, 2023), particularmente para los términos relacionados a la intervención. Se excluyó el trastorno bipolar, por ser considerado como una categoría diferente tanto por el DSM 5 (APA, 2014), como por otros autores que la consideran cualitativa y cuantitativamente distinta de la depresión propiamente dicha (Caballo et al., 2014). Se utilizó la herramienta de traducciones de Google y Perplexity Pro, así como las versiones en inglés de los manuales antes mencionados, a fin de traducir los términos con la mayor precisión y relación al campo clínico y psicométrico.

Tras este proceso, se determinaron los siguientes términos de búsqueda:

Tabla 3*Términos de búsqueda según componente de método PICO*

Componente	Términos en español	Términos en inglés	Términos truncos
Población	perú peruano, peruana	peru peruvian	
Intervención	abatimiento	dejection, despondency	
	depresión	depression	depres*
	depresión unipolar	unipolar depression	
	depresivo	depressive	
	desánimo	discouragement	
	distimia	dysthymia	
	episodio depresivo	depressive episode	
	episodio depresivo mayor	major depressive episode	
	episodios depresivos	depressive episodes	
	melancolía	melancholia, melancholy	
	síntoma depresivo	depressive symptom	
	síntomas depresivos	depressive symptoms	
	sintomatología depresiva	depressive symptomatology	
	trastorno afectivo estacional	seasonal affective disorder	
	trastorno depresivo	depressive disorder	
	trastorno depresivo persistente	persistent depressive disorder	
	trastorno distímico	dysthymic disorder	
	trastornos depresivos	depressive disorders	
Outcomes/ resultados	adaptación	adaptation	
	adaptación cultural	cultural adaptation	
	adaptación transcultural	cross-cultural adaptation	
	análisis discriminante	discriminant analysis	
	análisis factorial	factor analysis	
	características psicométricas	psychometric characteristics	
	confiabilidad, fiabilidad	reliability	
	cuestionario	questionnaire	
	escala	scale	
	especificidad	specificity	
	evaluación	assessment	
	evidencias psicométricas	psychometric evidence	
	instrumento	instrument	
	invarianza	invariance	
	medición	measurement	measur*
	propiedades psicométricas	psychometric properties	
	psicometría	psychometric, psychometrics	psicometr*, psychometr*
	sensibilidad	sensitivity	
	tamizaje, cribado	screening	screen*
	test	test	
	traducción	translation	
	validación	validation	
	validez	validity	valid*
	versión abreviada	abbreviated version	
	versión breve	brief version	
	versión corta	short version	
	versión peruana	peruvian version	

Nota: el indicativo “*” (asterisco) se utiliza para caracterizar a un término trunco dentro de los buscadores de los repositorios e indexadores. Los buscadores no diferencian entre mayúsculas y minúsculas.

A continuación, se formuló la estrategia de búsqueda utilizando los operadores booleanos AND y OR, así como paréntesis, comillas y asteriscos para demarcar los conjuntos de términos, las frases y los términos truncos, respectivamente, con lo que obtuvo la siguiente estrategia:

(abatimiento OR depresión OR "depresión unipolar" OR depresivo OR desánimo OR distimia OR "episodio depresivo" OR "episodio depresivo mayor" OR "episodios depresivos" OR melancolía OR "síntoma depresivo" OR "síntomas depresivos" OR "sintomatología depresiva" OR "trastorno afectivo estacional" OR "trastorno depresivo" OR "trastorno depresivo persistente" OR "trastorno distímico" OR "trastornos depresivos" OR dejection OR despondency OR depression OR "unipolar depression" OR depressive OR discouragement OR dysthymia OR "depressive episode" OR "major depressive episode" OR "depressive episodes" OR melancholia OR melancholy OR "depressive symptom" OR "depressive symptoms" OR "depressive symptomatology" OR "seasonal affective disorder" OR "depressive disorder" OR "persistent depressive disorder" OR "dysthymic disorder" OR "depressive disorders" OR depres*) AND (adaptación OR "adaptación cultural" OR "adaptación transcultural" OR "análisis discriminante" OR "análisis factorial" OR "características psicométricas" OR confiabilidad OR fiabilidad OR cuestionario OR escala OR especificidad OR evaluación OR "evidencias psicométricas" OR instrumento OR invarianza OR medición OR "propiedades psicométricas" OR psicometría OR sensibilidad OR tamizaje OR cribado OR test OR traducción OR validación OR validez OR "versión abreviada" OR "versión breve" OR "versión corta" OR "versión peruana" OR adaptation OR "cultural adaptation" OR "cross-cultural adaptation" OR "discriminant analysis" OR "factor analysis" OR "psychometric characteristics" OR reliability OR questionnaire OR scale OR specificity OR assessment OR "psychometric evidence" OR instrument OR invariance OR measurement OR "psychometric properties" OR psychometric OR psychometrics OR sensitivity OR screening OR test OR translation OR validation OR validity OR "abbreviated version" OR "brief version" OR "short version" OR

"peruvian version" OR measur* OR psicometr* OR psychometr* OR screen* OR valid*) AND (perú OR peruano OR peruana OR peru OR peruvian)

Dicha estrategia se utilizó en los buscadores de las bases de datos Scopus, Web of Science, Scielo, PubMed, Biblioteca Virtual en Salud (BVS) y Google Scholar debido a su gran relevancia dentro en el ámbito de la publicación científica, en los títulos, resúmenes y palabras clave (TITLE-ABS-KEY) a fin de recopilar todos los estudios relacionados a los objetivos de la investigación, con la consiguiente depuración según los criterios de inclusión establecidos. Por último, se analizaron los datos más relevantes pertenecientes a las características hemerográficas y psicométricas de los instrumentos estudiados, utilizando tablas y figuras a fin de presentar los indicadores globales y particulares; se revisaron los artículos en inglés y español, utilizando las herramientas de traducción de Google y Perplexity Pro para agilizar y facilitar su lectura.

Se utilizó el Protocolo PRISMA 2020 para presentar los datos. El protocolo PRISMA 2020 para la publicación y registro de revisiones sistemáticas, elaborado por Page et al. (2021), por sus siglas en inglés *Preferred Reporting Items for Systematic Reviews and MetaAnalyses for Protocols*, consiste en una lista de 27 ítems que señalan las principales características con que debe contar una revisión sistemática rigurosa y objetiva, a fin de aumentar su grado de replicabilidad y disponer de una guía validada por expertos de nivel mundial respecto al adecuado reporte de datos de revisión. Esta lista, a diferencia de su predecesora publicada por Urrútia y Bonfill (2010), incorpora nueva terminología y se retroalimenta de las nuevas evidencias encontradas en revisiones sistemáticas más recientes, tomando criterios empíricos y de expertos para aumentar la utilidad de los criterios adicionales (Anexo 2).

3.7. Análisis de datos

Los artículos fueron extraídos de las bases de datos mencionadas el día 15 de enero del 2026 utilizando la estrategia de búsqueda, a fin de agruparlos, cribarlos, recuperarlos y evaluar

su elegibilidad. Las investigaciones que validaron 2 o más instrumentos se consideraron separadas en algunos casos.

Los datos fueron analizados desglosando las principales características hemerográficas, ubicando el o los artículos analizados, la cantidad y características de la muestra, entre otras, así como relacionadas a la adaptación cultural, validez, confiabilidad, y normas de todos los instrumentos, determinando aquellos que posean propiedades más adecuadas para medir la depresión en sus respectivas muestras. Se utilizó el programa Excel 2016 para codificar y organizar los datos de forma tabular y sistemática, y se procesó lo tabulado con el software de licencia libre Jamovi.

3.8. Consideraciones éticas

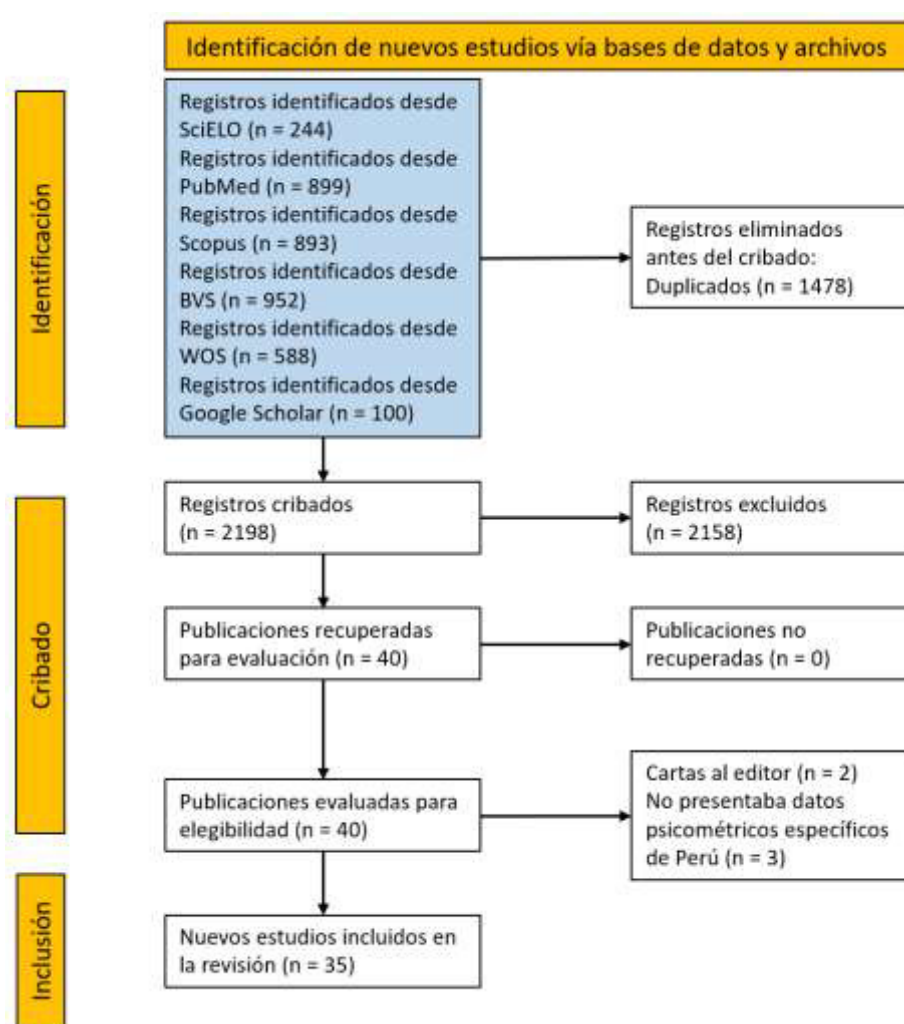
Debido a que no se trataron directamente los datos primarios de las puntuaciones obtenidas de los grupos etarios analizados, no se requirió del tratamiento de información sensible o susceptible a ser tergiversada, ni existió riesgo de exponer la identidad de ninguno de los participantes dentro de las muestras implicadas en cada investigación. Los datos susceptibles de identificación fueron parte de los protocolos de seguridad y privacidad de cada investigador, por lo que la presente investigación no posee injerencia ni potestad sobre ellos.

IV. RESULTADOS

Primeramente, se identificaron 3676 escritos entre todas las bases de datos, de los cuales 1478 fueron eliminados debido a la duplicación de estos, restando un total de 2198. Para la fase de cribado, fueron excluidos 2158, ya que no cumplían los criterios de inclusión previamente acotados para el protocolo de investigación presente, por lo que quedaron 40. Todos los artículos restantes fueron recuperados a fin de ser evaluados detalladamente. Finalmente, se concluyó que 5 no eran elegibles, por lo que la cifra final fue de 35 artículos. Cabe mencionar que algunas de estas investigaciones presentaron la validación de 2 o más instrumentos, por lo que el total validaciones asciende a 39, ver Figura 1.

Figura 1

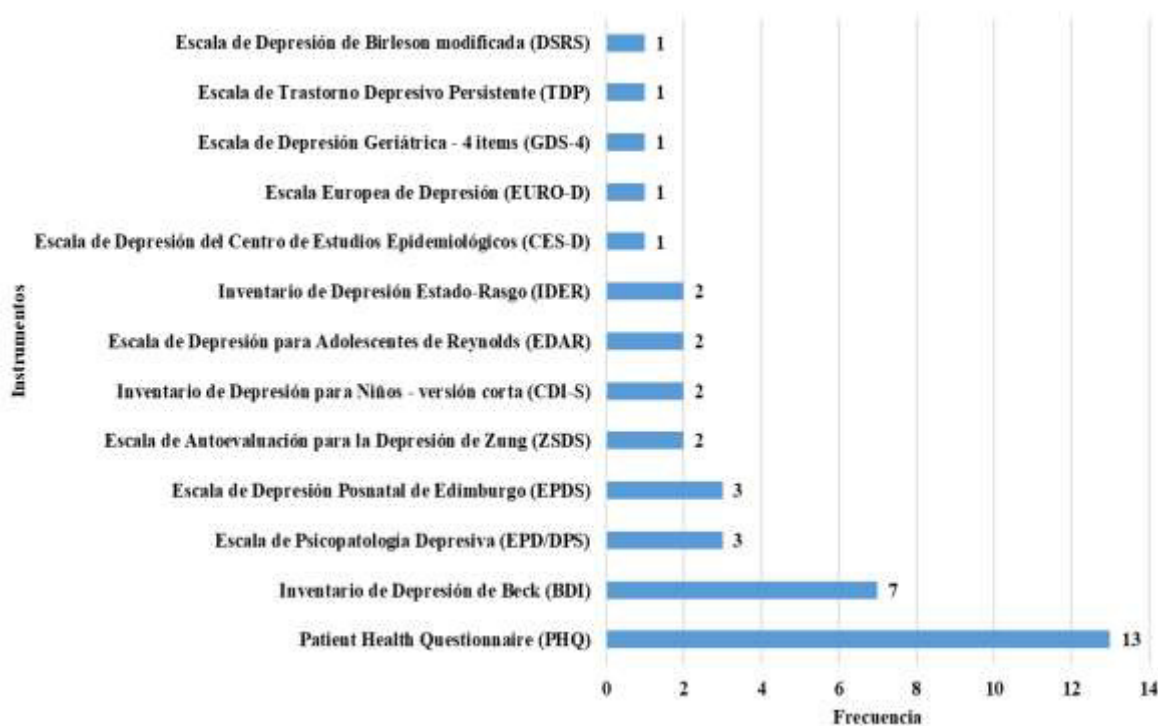
Diagrama de flujo de búsqueda y selección - PRISMA 2020



A continuación, se presenta la frecuencia de instrumentos analizados por los estudios incluidos en la revisión. Como puede observarse, el instrumento más investigado psicométricamente fue el PHQ (en sus 3 variantes, el PH1-9, PHQ-8 y PHQ-2), mientras que el segundo fue el BDI (en sus 4 variantes, BDI, BDI-II, BDI-IIA y BDI-PC), luego del EPD y del EPDS ningún instrumento posee más de 2 validaciones, contabilizando un total de 39 validaciones de instrumentos para medir la depresión en el ámbito peruano, ver Figura 2.

Figura 2

Frecuencia de instrumentos según la documentación revisada

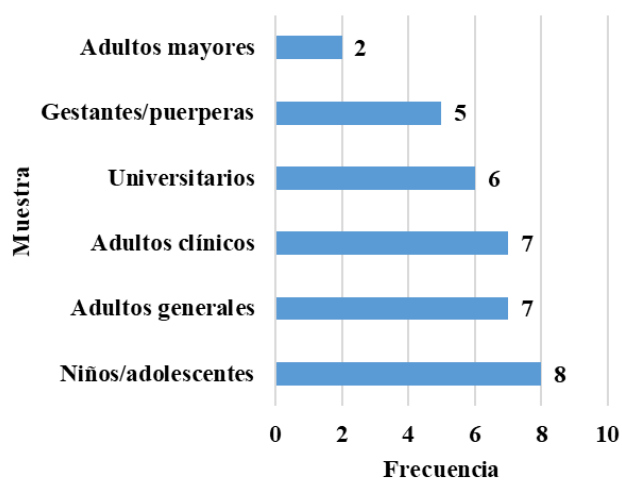


Así mismo, se muestra la cantidad muestral que cada instrumento tuvo en total, sumando las muestras empleadas en las 39 validaciones. El instrumento con mayor muestra total empleada es el PHQ, superando en casi 200 000 participantes al segundo (BDI), a partir del 3ro ninguno supera los 3 000 individuos evaluados. Se consideraron duplicados, ver Tabla 4.

Tabla 4*Muestra total por cada instrumento validado*

Instrumento	Muestra total
PHQ	204 980
BDI	5 139
EDAR	2 880
EPDS	2 431
EURO-D	1 933
CDI-S	1 248
TDP	868
EPD/DPS	708
IDER	565
GDS-4	400
ZSDS	397
DSRS	215
CES-D	194
Total	221 958

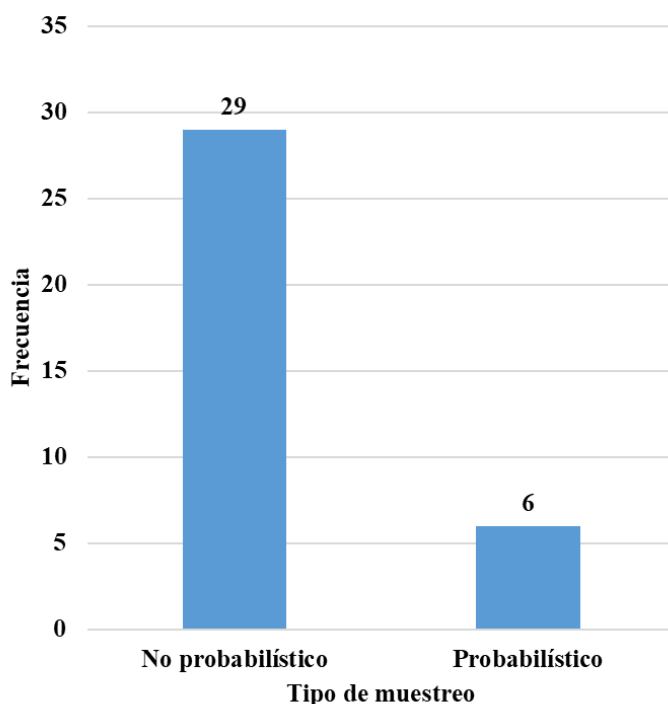
Se presenta la principal característica de las muestras empleadas por cada investigación. La muestra más recurrida fue la de niños y/o adolescentes, seguidos de adultos de población general, luego de adultos con alguna condición o característica clínica, universitarios, madres en etapa gestacional o de puerperio y adultos mayores. No se consideran las repeticiones por instrumento, solo artículos (35 investigaciones), ver Figura 3.

Figura 3*Característica principal del grupo muestral empleado*

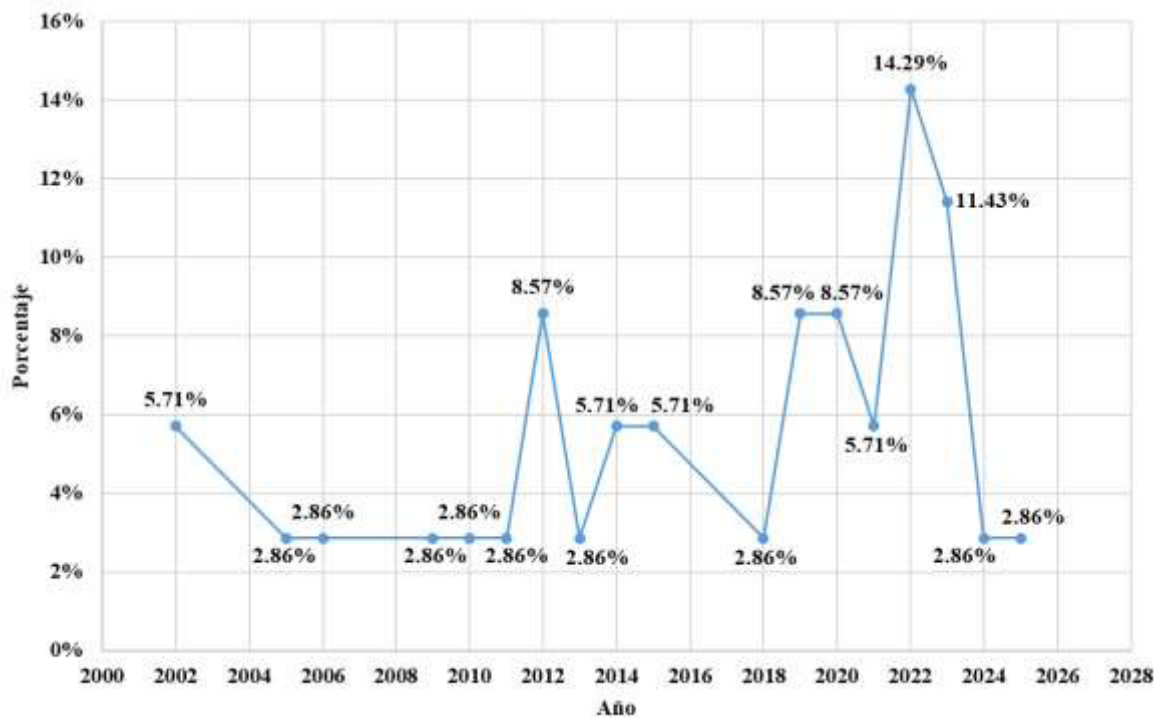
Según el tipo de muestreo empleado, se empleó el tipo no probabilístico por sobre el probabilístico en la mayoría de los estudios. No se consideraron duplicados, ver Figura 4.

Figura 4

Tipo de muestreo empleado



Por otro lado, se presenta la frecuencia de años de publicación para todas las investigaciones revisadas, siendo el 2022 el año con más publicaciones (14.29%, 5 investigaciones), seguido del 2023 (11.43%, 4 investigaciones), el resto de los años no supera las 3 investigaciones. Como puede observarse, existe una clara tendencia hacia el incremento de las investigaciones sobre las propiedades psicométricas de los instrumentos para medir depresión en el Perú en la actualidad. No se consideraron duplicados, ver Figura 5.

Figura 5*Frecuencia de año de cada publicación*

La Tabla 5 presenta las características de los estudios y muestras identificados. En total se incluyeron 35 estudios con un total de 39 validaciones, correspondientes a trece instrumentos. En cuanto al tamaño muestral, este fue variable, con muestras que oscilaron entre 136 participantes y 2005, sin considerar los estudios del PHQ que presentan tamaños considerablemente mayores. En cuanto al método de selección, la gran mayoría de los estudios empleó muestreo no probabilístico, con solo 6 estudios con muestreo probabilístico. Finalmente, en relación con la edad, las muestras fueron heterogéneas: los estudios en niños y adolescentes abarcaron rangos desde los 8 hasta los 18 años, mientras que los estudios en adultos mayores presentaron medias superiores a los 70 años, reflejando la diversidad etaria de la evidencia disponible.

Tabla 5

Características de los estudios y muestras de los instrumentos para medir depresión validados en población peruana

Instrumento	Versión validada	Autor	Titulo	Muestra			
				Característica principal	Cantidad	Selección	Edad
Escala de Depresión de Birleson modificada (DSRS)	DSRS	Vivar et al. (2005)	Validación de la Escala de Birleson modificada para trastornos depresivos en niños y adolescentes peruanos	Niños/adolescentes	215	No probabilístico	Rango 8-18 años
Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	De la Torre et al. (2006)	Validación de la versión reducida de la escala de depresión geriátrica en el consultorio externo de geriatría del Hospital Nacional Cayetano Heredia	Adultos mayores	400	No probabilístico	Media 70,85; rango 60-98 años
Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Ruiz-Grosso et al. (2012)	Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A Comparative Validation Study	Adultos clínicos	194	No probabilístico	Rango 18-65 años
Escala Europea de Depresión (EURO-D)	EURO-D	Guerra et al. (2015)	Psychometric properties of EURO-D, a geriatric depression scale: a cross-cultural validation study	Adultos mayores	1,933	Probabilístico	Media 75,0 años (urbano); 74,1 años (rural)
Escala de Trastorno Depresivo Persistente (TDP)	TDP	Elejabo-Pacheco et al. (2025)	Evidencias preliminares de validez de la Escala de Trastorno Depresivo Persistente (TDP) en estudiantes universitarios peruanos	Universitarios	868	No probabilístico	Media 22,34 años; rango 18-34 años
Inventario de Depresión Estado-Rasgo (IDER)	IDER	Soto et al. (2012)	Análisis factorial exploratorio del Inventario de Depresión Estado-Rasgo (ST-DEP) en adolescentes	Niños/adolescentes	314	No probabilístico	Media 14,39 años; rango 11-18 años
	IDER	Sotelo et al (2012)	Propiedades Psicométricas del Inventario de Depresión Estado-Rasgo (IDER) en Adultos de Lima	Adultos generales	251	No probabilístico	Media 33,65 años (consultantes); 2

Instrumento	Versión validada	Autor	Titulo	Muestra			
				Característica principal	Cantidad	Selección	Edad 0,44 años (no consultantes)
Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Ramos-Vera et al. (2023)	Factorial and network structure of the Reynolds Adolescent Depression Scale (RADS-2) in Peruvian adolescents	Niños/adolescentes	917	No probabilístico	Media 15,24 años; rango 13-18 años
	EDAR	Chávez y Escurra (2002)	Adaptación psicométrica de la Escala de Depresión para Adolescentes de Reynolds (EDAR) en estudiantes de secundaria de Lima metropolitana	Niños/adolescentes	1963	No probabilístico	Rango 12-18 años
Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Merino-Soto et al. (2022)	Effects of anonymity versus examinee name on a measure of depressive symptoms in adolescents	Niños/adolescentes	189	No probabilístico	Media 13,23 años; rango 11-17 años
	CDI-S	Ventura-León et al. (2020)	Evidencias de validez y fiabilidad de la versión en español del Children's Depression Inventory-Short en población peruana	Niños/adolescentes	1059	No probabilístico	Media 9,73 años; rango 7-12 años
Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Zhong et al. (2014)	Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression	Gestantes/puérperas	1517	No probabilístico	Media 28 años
	EPDS	Lam et al. (2010)	Factores psicosociales y depresión antenatal en mujeres gestantes. Estudio multicéntrico en tres hospitales de Lima, Perú. Abril a junio de 2008	Gestantes/puérperas	593	Probabilístico	Media 28,6 años; rango 20-45 años
	EPDS	Vega-Dienstmaier et al. (2002)	Validación de una versión en español de la Escala de Depresión Postnatal de Edimburgo	Gestantes/puérperas	321	No probabilístico	Media 25,05 años
Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Ruiz-Grosso et al. (2012)	Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A Comparative Validation Study	Adultos clínicos	194	No probabilístico	Rango 18-65 años
	ZSDS	Del Aguila et al. (2021)	Escala de depresión de Zung: propiedades psicométricas en portadores del virus de la inmunodeficiencia humana	Adultos clínicos	203	No probabilístico	≥18 años

Instrumento	Versión validada	Autor	Titulo	Muestra			
				Característica principal	Cantidad	Selección	Edad
Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Vega-Dienstmaier (2018)	Construcción de versiones cortas de la Escala de Psicopatología Depresiva para tamizaje de depresión mayor y sus características psicométricas	Adultos clínicos	226	No probabilístico	Media 28,49 años; rango 18-60 años
	EPD/DPS	Vega-Dienstmaier et al. (2011)	The Depressive Psychopathology Scale: presentation and initial validation in a sample of Peruvian psychiatric patients	Adultos clínicos	226	No probabilístico	Media 28,49 año; rango 18-60 años
	EPD/DPS	Lee-Diaz et al. (2024)	Validación de la Escala de Psicopatología Depresiva de 4 ítems (EPD-4) aplicada en línea en Perú	Adultos generales	256	No probabilístico	Media 37,60 años
Inventario de Depresión de Beck (BDI)	BDI	Vega-Dienstmaier et al. (2014)	Validez de una versión en español del Inventario de Depresión de Beck en pacientes hospitalizados de medicina general	Adultos clínicos	136	No probabilístico	Media 36,5 años
	BDI-II	Sánchez-Villena y Farfán (2019)	Análisis Factorial Exploratorio del Inventario de Depresión de Beck (BDI-II) en Universitarios Cajamarquinos	Universitarios	326	No probabilístico	Media 19,7 años; rango 15-31 años
	BDI-II	Sánchez-Villena et al. (2022)	Estructura factorial y datos normativos del Inventario de Depresión de Beck (BDI-II) en población general peruana	Adultos generales	1665	No probabilístico	Media 24,99 años; rango 18-54 años
	BDI-II	Sánchez-Pachas (2019)	Propiedades Psicométricas del Inventario de Depresión de Beck-II (IDB-II) en una muestra clínica	Adultos clínicos	400	No probabilístico	Media 33 años; rango 17-71 años
	BDI-IIA	Rodríguez y Farfán (2015)	Propiedades Psicométricas del Inventario de Depresión de Beck en estudiantes de 5to año de secundaria de Huancayo, 2013	Niños/adolescentes	300	No probabilístico	Rango 16-18 años
	BDI-IIA	Carranza (2013)	Propiedades psicométricas del inventario de depresión de Beck universitarios de lima	Universitarios	2005	No probabilístico	Rango 16-36 años
	BDI-PC	Lam et al. (2009)	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	Gestantes/púérperas	307	No probabilístico	Media 26 años; rango 14-46 años

Instrumento	Versión validada	Autor	Titulo	Muestra			
				Característica principal	Cantidad	Selección	Edad
Patient Health Questionnaire (PHQ)	PHQ 2	Baños-Chaparro et al. (2021)	Evidencias psicométricas del Patient Health Questionnaire-2 para la evaluación de síntomas depresivos en adultos peruanos	Adultos generales	283	No probabilístico	Media 31 años; rango 18-64 años
	PHQ 2	Villarreal-Zegarra et al. (2023)	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population	Adultos clínicos	1,347	No probabilístico	≥18 años
	PHQ 8	Villarreal-Zegarra et al. (2023)	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population	Adultos clínicos	1,347	No probabilístico	≥18 años
	PHQ 9	Cjuno et al. (2023)	Adaptación cultural al Quechua y análisis psicométrico del Patient Health Questionnaire (PHQ-9) en población peruana	Adultos generales	970	No probabilístico	Rango 18-68 años
	PHQ 9	Renteria et al. (2022)	Efectos de estilos de crianza sobre la depresión en adolescentes escolarizados de una muestra peruana	Niños/adolescentes	308	No probabilístico	Media 14,5 años; rango 12-17 años
	PHQ 9	Huarcaya-Victoria et al. (2020)	Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú	Universitarios	200	Probabilístico	Media 20,84 años; rango 16-31 años
	PHQ 9	Villarreal-Zegarra et al. (2020)	Trends in the prevalence and treatment of depressive symptoms in Peru: A population-based study	Adultos generales	161,061	Probabilístico	≥15 años; media 40,3 años
	PHQ 9	Villarreal-Zegarra et al. (2019)	Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics	Adultos generales	30449	Probabilístico	Media 40,5 años; rango 18-98 años
	PHQ 9	Alvarez et al. (2022)	Validación del Cuestionario sobre la Salud del Paciente-9 (PHQ-9) en internos de medicina humana de una universidad de referencia del Perú durante la pandemia Covid-19	Universitarios	343	Probabilístico	No declara

Instrumento	Versión validada	Autor	Titulo	Muestra			
				Característica principal	Cantidad	Selección	Edad
	PHQ 9	Zhong et al. (2014)	Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression	Gestantes/puerperas	1517	No probabilístico	Media 28 años; rango 18-49 años
	PHQ 9	Lingán-Huamán et al. (2023)	Psychometric Properties of the Patient Health Questionnaire-9 in Peruvian University Students	Universitarios	409	No probabilístico	Media 22,15 años; rango 16-45 años
	PHQ 9	Villarreal-Zegarra et al. (2023)	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population	Adultos clínicos	1,347	No probabilístico	≥18 años
	PHQ 9	Smith et al. (2022)	Validation of the patient health Questionnaire-9 (PHQ-9) for detecting depression among pregnant women in Lima, Peru	Gestantes/puerperas	5,399	No probabilístico	Media 28,1 años; rango 18-49 años

La Tabla 6 presenta los procedimientos de adaptación cultural reportados en los estudios incluidos, organizados en función de las cinco etapas del proceso recomendado. Los resultados muestran que la adaptación cultural fue el criterio con menor cobertura en el conjunto de la evidencia revisada, la mayoría de los instrumentos no registró ninguna de las etapas formales del proceso, ya sea porque se adoptaron versiones previas en español sin descripción del procedimiento seguido, o porque se trataba de instrumentos desarrollados originalmente en el Perú (como la TDP) para los cuales este criterio no aplica. Entre los estudios que sí reportaron alguna etapa, el procedimiento más frecuente fue la prueba piloto (4 estudios), seguido de traducción directa y revisión por comité de expertos (3 estudios cada uno), finalizando con síntesis de traducciones (1 estudio).

Tabla 6

Procedimientos de adaptación cultural de los instrumentos para medir depresión validados en población peruana

Instrumento	Versión validada	Autor	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural
			Etapa 1: Traducción directa al idioma de interés	Etapa 2: Síntesis de traducciones y solución de discrepancias	Etapa 3: Traducción inversa (idioma destino - idioma original)	Etapa 4: Revisión por comité de expertos	Etapa 5: Prueba piloto
Escala de Depresión de Birleson modificada (DSRS)	DSRS	Vivar et al. (2005)					
Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	De la Torre et al. (2006)	X			X	
Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Ruiz-Grosso et al. (2012)					
Escala Europea de Depresión (EURO-D)	EURO-D	Guerra et al. (2015)					
Escala de Trastorno Depresivo Persistente (TDP)	TDP	Elejabo-Pacheco et al. (2025)					
Inventario de Depresión Estado-Rasgo (IDER)	IDER	Soto et al. (2012)					
	IDER	Sotelo et al (2012)				X	
Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Ramos-Vera et al. (2023)					
	EDAR	Chávez y Escurra (2002)					
Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Merino-Soto et al. (2022)					
	CDI-S	Ventura-León et al. (2020)					X
	EPDS	Zhong et al. (2014)					
	EPDS	Lam et al. (2010)					

Instrumento	Versión validada	Autor	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural
			Etapas 1: Traducción directa al idioma de interés	Etapas 2: Síntesis de traducciones y solución de discrepancias	Etapas 3: Traducción inversa (idioma destino - idioma original)	Etapas 4: Revisión por comité de expertos	Etapas 5: Prueba piloto
Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Vega-Dienstmaier et al. (2002)	X				
Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Ruiz-Grosso et al. (2012)					
	ZSDS	Del Aguila et al. (2021)					
Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Vega-Dienstmaier (2018)					
	EPD/DPS	Vega-Dienstmaier et al. (2011)					
	EPD/DPS	Lee-Díaz et al. (2024)					
Inventario de Depresión de Beck (BDI)	BDI	Vega-Dienstmaier et al. (2014)					
	BDI-II	Sánchez-Villena y Farfán (2019)					
	BDI-II	Sánchez-Villena et al. (2022)					
	BDI-II	Sánchez-Pachas (2019)					X
	BDI-IIA	Rodríguez y Farfán (2015)					X
	BDI-IIA	Carranza (2013)					
	BDI-PC	Lam et al. (2009)					
Patient Health Questionnaire (PHQ)	PHQ 2	Baños-Chaparro et al. (2021)					
	PHQ 2	Villarreal-Zegarra et al. (2023)					
	PHQ 8	Villarreal-Zegarra et al. (2023)					
	PHQ 9	Cjuno et al. (2023)	X	X	X	X	X
	PHQ 9	Renteria et al. (2022)					

Instrumento	Versión validada	Autor	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural	Adaptación cultural
			Etapas 1: Traducción directa al idioma de interés	Etapas 2: Síntesis de traducciones y solución de discrepancias	Etapas 3: Traducción inversa (idioma destino - idioma original)	Etapas 4: Revisión por comité de expertos	Etapas 5: Prueba piloto
	PHQ 9	Huarcaya-Victoria et al. (2020)					
	PHQ 9	Villarreal-Zegarra et al. (2020)					
	PHQ 9	Villarreal-Zegarra et al. (2019)					
	PHQ 9	Alvarez et al. (2022)					
	PHQ 9	Zhong et al. (2014)					
	PHQ 9	Lingán-Huamán et al. (2023)					
	PHQ 9	Villarreal-Zegarra et al. (2023)					
	PHQ 9	Smith et al. (2022)					

La Tabla 7 presenta los tipos de evidencia de confiabilidad reportados en los estudios, organizados en cuatro categorías. El patrón más destacado es la predominancia de los coeficientes de coherencia interna, que fueron el único tipo de evidencia reportado en la gran mayoría de los estudios (38 validaciones), lo que revela una tendencia a reducir la evaluación de la confiabilidad a un único indicador. En contraste, los coeficientes test-retest fueron reportados de manera muy escasa (3 veces), mientras que los índices de coherencia entre evaluadores se registraron exclusivamente en la GDS-4, y ningún estudio reportó coeficientes de formularios alternativos.

Tabla 7

Fuentes de evidencia de confiabilidad reportados en los estudios de validación de instrumentos de depresión en población peruana

Instrumento	Versión validada	Autor	Confiabilidad/precisión			Índices de coherencia entre evaluadores
			Coefficientes de formularios alternativos	Coefficientes test-retest	Coefficientes de coherencia interna	
Escala de Depresión de Birleson modificada (DSRS)	DSRS	Vivar et al. (2005)		X	X	
Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	De la Torre et al. (2006)				X
Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Ruiz-Grosso et al. (2012)			X	
Escala Europea de Depresión (EURO-D)	EURO-D	Guerra et al. (2015)			X	
Escala de Trastorno Depresivo Persistente (TDP)	TDP	Elejabo-Pacheco et al. (2025)			X	
Inventario de Depresión Estado-Rasgo (IDER)	IDER	Soto et al. (2012)			X	
	IDER	Sotelo et al (2012)			X	
Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Ramos-Vera et al. (2023)			X	
	EDAR	Chávez y Ecurra (2002)		X	X	
Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Merino-Soto el al. (2022)			X	
	CDI-S	Ventura-León et al. (2020)			X	
Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Zhong et al. (2014)			X	
	EPDS	Lam et al. (2010)			X	
	EPDS	Vega-Dienstmaier et al. (2002)			X	
Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Ruiz-Grosso et al. (2012)			X	
	ZSDS	Del Aguila et al. (2021)			X	
Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Vega-Dienstmaier (2018)			X	
	EPD/DPS	Vega-Dienstmaier et al. (2011)			X	

Instrumento	Versión validada	Autor	Confiabilidad/precisión			Índices de coherencia entre evaluadores
			Coefficientes de formularios alternativos	Coefficientes test-retest	Coefficientes de coherencia interna	
Inventario de Depresión de Beck (BDI)	EPD/DPS	Lee-Díaz et al. (2024)			X	
	BDI	Vega-Dienstmaier et al. (2014)			X	
	BDI-II	Sánchez-Villena y Farfán (2019)			X	
	BDI-II	Sánchez-Villena et al. (2022)			X	
	BDI-II	Sánchez-Pachas (2019)			X	
	BDI-IIA	Rodríguez y Farfán (2015)			X	
	BDI-IIA	Carranza (2013)		X	X	
	BDI-PC	Lam et al. (2009)			X	
Patient Health Questionnaire (PHQ)	PHQ 2	Baños-Chaparro et al. (2021)			X	
	PHQ 2	Villarreal-Zegarra et al. (2023)			X	
	PHQ 8	Villarreal-Zegarra et al. (2023)			X	
	PHQ 9	Cjuno et al. (2023)			X	
	PHQ 9	Renteria et al. (2022)			X	
	PHQ 9	Huarcaya-Victoria et al. (2020)			X	
	PHQ 9	Villarreal-Zegarra et al. (2020)			X	
	PHQ 9	Villarreal-Zegarra et al. (2019)			X	
	PHQ 9	Alvarez et al. (2022)			X	
	PHQ 9	Zhong et al. (2014)			X	
	PHQ 9	Lingán-Huamán et al. (2023)			X	
	PHQ 9	Villarreal-Zegarra et al. (2023)			X	
	PHQ 9	Smith et al. (2022)			X	

La Tabla 8 presenta las fuentes de evidencia de validez reportadas en los estudios incluidos, organizadas en cinco categorías. El patrón predominante fue la combinación de evidencia basada en la estructura interna (36 veces) y en las relaciones con otras variables (15 veces), que constituyó el perfil más frecuente en el conjunto de estudios. La evidencia basada en el contenido de la prueba fue reportada solo 5 veces, en

contraste, dos fuentes de evidencia estuvieron ausentes en toda la revisión: la evidencia basada en los procesos de respuesta y la evidencia basada en las consecuencias del uso de las pruebas.

Tabla 8

Fuentes de evidencia de validez reportadas en los estudios de validación de instrumentos de depresión en población peruana

Instrumento	Versión validada	Autor	Validez				
			Evidencia basada en el contenido de la prueba	Evidencia basada en los procesos de respuesta	Evidencia basada en la estructura interna	Evidencia basada en relaciones con otras variables	Evidencia basada en las consecuencias de las pruebas
Escala de Depresión de Birleson modificada (DSRS)	DSRS	Vivar et al. (2005)					
Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	De la Torre et al. (2006)					
Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Ruiz-Grosso et al. (2012)			X	X	
Escala Europea de Depresión (EURO-D)	EURO-D	Guerra et al. (2015)			X	X	
Escala de Trastorno Depresivo Persistente (TDP)	TDP	Elejabo-Pacheco et al. (2025)	X		X	X	
Inventario de Depresión Estado-Rasgo (IDER)	IDER	Soto et al. (2012)			X		
	IDER	Sotelo et al (2012)	X		X		
Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Ramos-Vera et al. (2023)			X		
	EDAR	Chávez y Ecurra (2002)			X		
Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Merino-Soto et al. (2022)			X		
	CDI-S	Ventura-León et al. (2020)			X	X	
Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Zhong et al. (2014)			X		
	EPDS	Lam et al. (2010)			X		

Instrumento	Versión validada	Autor	Validez				
			Evidencia basada en el contenido de la prueba	Evidencia basada en los procesos de respuesta	Evidencia basada en la estructura interna	Evidencia basada en relaciones con otras variables	Evidencia basada en las consecuencias de las pruebas
Escala de Autoevaluación para la Depresión de Zung (ZSDS)	EPDS	Vega-Dienstmaier et al. (2002)				X	
	ZSDS	Ruiz-Grosso et al. (2012)			X	X	
	ZSDS	Del Aguila et al. (2021)	X		X	X	
Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Vega-Dienstmaier (2018)			X	X	
	EPD/DPS	Vega-Dienstmaier et al. (2011)			X	X	
	EPD/DPS	Lee-Díaz et al. (2024)			X	X	
Inventario de Depresión de Beck (BDI)	BDI	Vega-Dienstmaier et al. (2014)			X		
	BDI-II	Sánchez-Villena y Farfán (2019)			X	X	
	BDI-II	Sánchez-Villena et al. (2022)			X		
	BDI-II	Sánchez-Pachas (2019)	X		X		
	BDI-IIA	Rodríguez y Farfán (2015)			X		
	BDI-IIA	Carranza (2013)	X		X		
	BDI-PC	Lam et al. (2009)			X	X	
Patient Health Questionnaire (PHQ)	PHQ 2	Baños-Chaparro et al. (2021)			X	X	
	PHQ 2	Villarreal-Zegarra et al. (2023)			X		
	PHQ 8	Villarreal-Zegarra et al. (2023)			X		
	PHQ 9	Cjuno et al. (2023)			X		
	PHQ 9	Renteria et al. (2022)			X		
	PHQ 9	Huarcaya-Victoria et al. (2020)			X		
	PHQ 9	Villarreal-Zegarra et al. (2020)			X		
	PHQ 9	Villarreal-Zegarra et al. (2019)			X		
	PHQ 9	Alvarez et al. (2022)			X		
	PHQ 9	Zhong et al. (2014)			X	X	
	PHQ 9	Lingán-Huamán et al. (2023)			X	X	
	PHQ 9	Villarreal-Zegarra et al. (2023)			X		

Instrumento	Versión validada	Autor	Validez				
			Evidencia basada en el contenido de la prueba	Evidencia basada en los procesos de respuesta	Evidencia basada en la estructura interna	Evidencia basada en relaciones con otras variables	Evidencia basada en las consecuencias de las pruebas
	PHQ 9	Smith et al. (2022)			X		

La Tabla 9 presenta los indicadores normativos reportados en los estudios incluidos, organizados en dos categorías. El patrón predominante fue el reporte de sensibilidad y especificidad, presente en 14 validaciones; la baremación, en cambio, fue reportada de forma muy escasa, limitándose a 5 validaciones. Destaca además que un número considerable de estudios no reportó ningún tipo de indicador normativo, lo que limita la posibilidad de interpretar las puntuaciones en términos clínicamente significativos.

Tabla 9

Indicadores normativos en los estudios de validación de instrumentos de depresión en población peruana

Instrumento	Versión validada	Autor	Normas	Normas
			Sensibilidad (s) / Especificidad (e)	Baremación
Escala de Depresión de Birleson modificada (DSRS)	DSRS	Vivar et al. (2005)	X	
Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	De la Torre et al. (2006)	X	
Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Ruiz-Grosso et al. (2012)	X	

Instrumento	Versión validada	Autor	Normas	Normas
			Sensibilidad (s) / Especificidad (e)	Baremación
Escala Europea de Depresión (EURO-D)	EURO-D	Guerra et al. (2015)	X	
Escala de Trastorno Depresivo Persistente (TDP)	TDP	Elejabo-Pacheco et al. (2025)		
Inventario de Depresión Estado-Rasgo (IDER)	IDER	Soto et al. (2012)		
	IDER	Sotelo et al (2012)		
Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Ramos-Vera et al. (2023)		
	EDAR	Chávez y Ecurra (2002)		X
Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Merino-Soto et al. (2022)		
	CDI-S	Ventura-León et al. (2020)		
Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Zhong et al. (2014)		
	EPDS	Lam et al. (2010)		
	EPDS	Vega-Dienstmaier et al. (2002)	X	
Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Ruiz-Grosso et al. (2012)	X	
	ZSDS	Del Aguila et al. (2021)		X
Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Vega-Dienstmaier (2018)	X	
	EPD/DPS	Vega-Dienstmaier et al. (2011)	X	
	EPD/DPS	Lee-Díaz et al. (2024)	X	
Inventario de Depresión de Beck (BDI)	BDI	Vega-Dienstmaier et al. (2014)	X	
	BDI-II	Sánchez-Villena y Farfán (2019)		
	BDI-II	Sánchez-Villena et al. (2022)		X
	BDI-II	Sánchez-Pachas (2019)		
	BDI-IIA	Rodríguez y Farfán (2015)		X
	BDI-IIA	Carranza (2013)		X
	BDI-PC	Lam et al. (2009)		
Patient Health Questionnaire (PHQ)	PHQ 2	Baños-Chaparro et al. (2021)		

Instrumento	Versión validada	Autor	Normas	Normas
			Sensibilidad (s) / Especificidad (e)	Baremación
	PHQ 2	Villarreal-Zegarra et al. (2023)	X	
	PHQ 8	Villarreal-Zegarra et al. (2023)	X	
	PHQ 9	Cjuno et al. (2023)		
	PHQ 9	Renteria et al. (2022)		
	PHQ 9	Huarcaya-Victoria et al. (2020)		
	PHQ 9	Villarreal-Zegarra et al. (2020)		
	PHQ 9	Villarreal-Zegarra et al. (2019)		
	PHQ 9	Alvarez et al. (2022)		
	PHQ 9	Zhong et al. (2014)		
	PHQ 9	Lingán-Huamán et al. (2023)		
	PHQ 9	Villarreal-Zegarra et al. (2023)	X	
	PHQ 9	Smith et al. (2022)	X	

A continuación, se presenta a detalle los resultados de la revisión sistemática, ordenados según los instrumentos hallados en los estudios incluidos.

Escala de Depresión de Birleson modificada (DSRS)

La Escala de Birleson modificada (Vivar et al., 2005) es un instrumento de autorreporte utilizado para la detección de sintomatología depresiva en niños y adolescentes peruanos, adaptado de la versión original de Birleson (1981). La versión empleada corresponde a una adaptación dirigida a población infantil y adolescente y quedó conformada por 21 ítems con tres alternativas de respuesta (siempre, a veces y nunca), donde puntajes más altos indican mayor gravedad de síntomas depresivos experimentados durante las últimas dos semanas. Asimismo, fue diseñada para identificar desde sintomatología depresiva hasta cuadros de mayor severidad. En esta versión modificada se incorporaron reactivos sobre ideación suicida, alteraciones de la concentración, irritabilidad, desesperanza, somatización y pérdida del placer por compartir con pares, y se eliminó el ítem *me gusta hablar con mi familia* por considerarse poco específico para el trastorno depresivo.

Según los estudios revisados, Vivar et al. (2005) llevaron a cabo la validación de la Escala de Birleson modificada para trastornos depresivos en 215 niños y adolescentes limeños recolectados por muestreo no probabilístico, divididos en 3 grupos: diagnosticados con depresión, diagnosticados con trastorno de ajuste con síntomas depresivos y sin diagnóstico de depresión (escolares), evaluando su desempeño entre estas agrupaciones.

Según lo reportado, la escala derivó de una adaptación de la versión original, en la que se añadieron y retiraron algunos ítems, además de realizarse cambios de vocabulario; asimismo, el instrumento fue sometido a juicio de expertos. No obstante, el estudio no describe de manera detallada las etapas del proceso de adaptación cultural, ni en lo referente a la modificación del número de ítems, ni en la adaptación léxica, ni en el procedimiento seguido para el juicio de expertos.

En cuanto a la confiabilidad, la escala mostró adecuada estabilidad temporal en los distintos grupos evaluados. En el grupo con trastorno depresivo, se observaron correlaciones test-retest altas tanto con el coeficiente de Pearson ($r = 0,893$) como con el de Spearman ($\rho = 0,858$), ambas estadísticamente significativas; en el grupo con trastorno de ajuste con síntomas depresivos, la estabilidad fue moderada, con coeficientes de Pearson = 0,542 y Spearman = 0,547, también significativos; por su parte, en el grupo de niños y adolescentes sin enfermedad mental se reportó igualmente una asociación alta entre test y retest ($r = 0,78$).

Respecto a la coherencia interna, la Escala de Birleson modificada presentó un alfa de Cronbach de 0,73 en el grupo con trastorno depresivo, lo que indica una consistencia interna aceptable. En contraste, en el grupo con trastorno de ajuste con síntomas depresivos, el alfa de Cronbach fue 0,46, lo que sugiere menor homogeneidad entre los ítems en este grupo; no se reportó este coeficiente para el grupo sin trastorno mental.

En relación con la sensibilidad y especificidad, el análisis exploratorio del punto de corte mostró que los valores 13 y 14 ofrecieron el mejor equilibrio entre ambos indicadores. En ambos puntos de corte, la sensibilidad fue de 98,8% y la especificidad de 98,7%, lo que respalda su utilidad diagnóstica y sugiere que dichos valores constituyen los puntos de corte más adecuados para la identificación de casos depresivos en la muestra estudiada.

Escala de Depresión Geriátrica - 4 ítems (GDS-4)

La Escala de Depresión Geriátrica (GDS) es un instrumento de tamizaje desarrollado por Yesavage et al. (1982) para la detección de depresión en adultos mayores, diseñado específicamente para evitar síntomas somáticos que pueden confundirse con comorbilidades propias de la edad. La versión de 15 ítems (GDS-15), abreviada por Sheikh y Yesavage (1986), está compuesta por preguntas dicotómicas (sí/no) que evalúan síntomas afectivos y cognitivos depresivos experimentados recientemente, con puntajes de 0 a 15 donde valores ≥ 5 sugieren depresión. Por su parte, la GDS-4 es una versión reducida de 4 ítems seleccionados de la GDS-

15, utilizada para tamizaje rápido en contextos clínicos con limitaciones de tiempo, manteniendo formato dicotómico y considerando positivo 2 o más respuestas afirmativas.

Según los estudios revisados, De la Torre et al. (2006) llevaron a cabo la validación de la GDS-4 en 400 pacientes mayores de 60 años atendidos en el consultorio externo de geriatría del Hospital Nacional Cayetano Heredia de Lima, comparándola con la Escala de Depresión Geriátrica de 15 ítems GDS-15 y utilizando el DSM-IV como estándar de referencia, elegidos por muestreo no probabilístico.

Según lo reportado, la GDS-4 derivó de una traducción directa del inglés al castellano, realizada por una traductora y un psiquiatra, y fue revisada por un grupo de geriatras expertos. No obstante, el estudio no describe de manera detallada las etapas del proceso de traducción, adaptación cultural, revisión formal o evaluación de equivalencia semántica.

En cuanto a la confiabilidad, la concordancia entre la GDS-15 y la GDS-4 fue moderada ($Kappa = 0,52$), lo que indica una moderada concordancia entre evaluadores al comparar la versión reducida con la original.

En relación con la sensibilidad y especificidad, la GDS-4 mostró una sensibilidad del 49% y una especificidad del 97% frente al diagnóstico DSM-IV de depresión, con un valor predictivo positivo del 86% y un valor predictivo negativo del 91%. Por su parte, la GDS-15 presentó una sensibilidad del 49% y una especificidad del 95%, con un valor predictivo positivo del 68% y un valor predictivo negativo del 90%.

Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)

La Escala de Depresión del Centro de Estudios Epidemiológicos (*Center for Epidemiological Studies Depression Scale*, CES-D) es un instrumento de autorreporte desarrollado por Radloff (1977) para la detección de sintomatología depresiva en población general. Está compuesta por 20 ítems, en los que cada reactivo se puntúa en una escala de 0 a 3 según la frecuencia con que la persona experimentó cada síntoma durante la última semana,

de modo que el puntaje total varía entre 0 y 60, donde valores más altos indican mayor presencia de síntomas depresivos. Asimismo, su propósito principal es funcionar como instrumento de tamizaje o detección de casos de depresión, tanto en investigación epidemiológica como en contextos clínicos.

Según los estudios revisados, Ruiz-Grosso et al. (2012) llevaron a cabo una validación comparativa de las versiones en español de la CES-D y de la Zung Self-Rating Depression Scale en 194 adultos pacientes atendidos de manera ambulatoria en el Hospital Nacional Cayetano Heredia de Lima, seleccionados por métodos no probabilísticos.

En cuanto a la confiabilidad, la CES-D mostró adecuada consistencia interna, con un alfa de Cronbach de 0,93, así como un coeficiente omega de McDonald jerárquico de 0,90.

Respecto a la evidencia basada en la estructura interna, la CES-D mostró un funcionamiento interno adecuado. Las correlaciones ítem-test oscilaron entre 0,36 y 0,83, mientras que las correlaciones ítem-resto variaron entre 0,29 y 0,81, lo que sugiere una adecuada asociación entre los reactivos y el puntaje global de la escala. En particular, el ítem 18 presentó el desempeño más favorable, mientras que el ítem 4 mostró el menor aporte relativo a la estructura interna del instrumento.

Respecto a la evidencia basada en relaciones con otras variables, Ruiz-Grosso et al. (2012) reportaron una correlación positiva alta entre la CES-D y la ZSDS ($\rho = 0,86$; $p < 0,001$), así como una correlación positiva moderada entre la CES-D y la Clinical Global Impression (CGI) ($\rho = 0,51$; $p < 0,001$), lo que respalda la convergencia de la escala con otras medidas de sintomatología y severidad depresiva.

En relación con la sensibilidad y especificidad, la CES-D mostró buen poder discriminativo en los análisis ROC realizados. Cuando se comparó el grupo con episodio depresivo mayor frente al grupo con otros diagnósticos psiquiátricos, el AUROC fue de 0,83 (IC 95%: 0,76-0,90), y el punto de corte óptimo fue ≥ 29 , con sensibilidad de 77,1%,

especificidad de 79,4% y 78,2% de casos correctamente clasificados. Por su parte, al comparar el grupo con episodio depresivo mayor frente al grupo sin evidencia de trastorno psiquiátrico, el AUROC fue de 0,98 (IC 95%: 0,97-0,99), y el punto de corte óptimo fue ≥ 24 , con sensibilidad de 91,4%, especificidad de 96,7% y 93,9% de casos correctamente clasificados.

Escala Europea de Depresión (EURO-D)

La Escala Europea de Depresión (EURO-D) es un instrumento breve compuesto por 12 ítems, concebido para medir la sintomatología depresiva en adultos mayores de 65 años, su desarrollo se enmarca en el Programa de Acción Concertada EURODEP (European Depression in Elderly Persons) (Prince et al., 1999), cuya finalidad fue disponer de una medida estandarizada y común que permitiera comparar la prevalencia y los factores de riesgo asociados a la depresión tardía en poblaciones de distintos países europeos, los ítems que la integran fueron extraídos del Geriatric Mental State (GMS) (Copeland et al., 2002), instrumento de referencia en la evaluación psicopatológica de la vejez. El sistema de puntuación es dicotómico y sumativo: cada síntoma se califica con 0 si está ausente y con 1 si está presente, obteniéndose una puntuación total que oscila entre 0 y 12, donde los valores más altos reflejan mayor carga sintomática. Desde el punto de vista estructural, los análisis factoriales han identificado de manera consistente dos dimensiones subyacentes en prácticamente todos los contextos estudiados: un factor de sufrimiento afectivo y otro de motivación (Guerra et al., 2015). La solidez psicométrica de esta estructura se consolidó con los datos de la Encuesta sobre Salud, Envejecimiento y Jubilación en Europa (SHARE) (Castro-Costa et al., 2007), realizada en diez naciones, cuyos resultados confirmaron que la EURO-D constituye una escala jerárquica con un ordenamiento de ítems comparable entre países, cargas factoriales similares en todos los contextos evaluados y adecuadas evidencias de consistencia interna y validez de constructo.

Según la literatura revisada, Guerra et al. (2015) realizaron una validación multinacional del instrumento en 1933 adultos mayores peruanos, con dos muestras, una urbana ($n = 1381$) y otra rural ($n = 552$), seleccionados a través de un proceso probabilístico.

Para el análisis de coherencia interna, Guerra et al. (2015) evaluaron la consistencia interna de la EURO-D mediante alfa de Cronbach en cada sitio del estudio, en Perú, el alfa fue de 0.71 en la muestra urbana y de 0.64 en la muestra rural, estos resultados indican una consistencia interna aceptable en la muestra peruana urbana y más baja en la rural.

Para la estructura interna, los autores reportaron que el KMO fue 0.74 y la prueba de esfericidad de Bartlett fue significativa ($p < 0.001$) para ambas muestras (urbana y rural) en conjunto, lo que respaldó la factorizabilidad de la matriz. El análisis de componentes principales identificó tres factores con *eigenvalues* mayores a 1 en Perú, aunque los dos primeros concentraron la mayor parte de la varianza: el primero explicó 24.3% y el segundo 12.1%, sumando 36.4% de varianza acumulada. En esta solución, el primer factor estuvo dominado por los ítems depresión y llanto, mientras que el segundo estuvo definido principalmente por interés; adicionalmente, aparecieron cargas secundarias de culpa, suicidabilidad, pesimismo, apetito y disfrute.

En el CFA para Perú, las cargas factoriales del factor de sufrimiento afectivo fueron: depresión = 1.00, llanto = 0.87, suicidabilidad = 0.25, sueño = 0.38, culpa = 0.22, irritabilidad = 0.22 y fatiga = 0.36; para el factor de motivación fueron: interés = 1.00, disfrute = 0.76, pesimismo = 0.38 y concentración = 0.52. La correlación entre ambos factores en Perú fue de 0.30, lo que sugiere relación moderada entre las dimensiones.

Para la evidencia basada en relaciones con otras variables, los autores correlacionaron los puntajes de la EURO-D con salud global autopercebida, discapacidad medida con WHODAS 2.0 (World Health Organization Disability Assessment Schedule 2.0) y felicidad. En Perú urbano, la EURO-D correlacionó con salud autopercebida en -0.45, con discapacidad

en 0.46 y con felicidad en -0.41; en rural, las correlaciones fueron -0.20, 0.37 y -0.17, respectivamente. En la dirección esperada, mayores síntomas depresivos se asociaron con peor salud percibida, mayor discapacidad y menor felicidad.

En el análisis de sensibilidad y especificidad, calibraron la EURO-D frente al diagnóstico de episodio depresivo CIE-10 mediante análisis ROC. En Perú urbano, el AUROC fue 0.94 (IC 95%: 0.93–0.96) y en Perú rural fue 0.96 (IC 95%: 0.94–0.98), lo que indica excelente capacidad discriminativa. El punto de corte óptimo fue 4/5 en ambas muestras peruanas; con ese punto de corte, la sensibilidad fue de 92.0% y la especificidad de 84.5% en Perú urbano, mientras que en Perú rural la sensibilidad fue de 100.0% y la especificidad de 87.0%.

Escala de Trastorno Depresivo Persistente (TDP)

La Escala de Trastorno Depresivo Persistente (TDP) fue elaborada por Elejabo-Pacheco et al. (2025) como una medida específica para evaluar la presencia de rasgos de trastorno depresivo persistente, de acuerdo con los criterios diagnósticos del DSM-5, en estudiantes universitarios peruanos mayores de 18 años. Se trata de un instrumento de nueva construcción, diseñado para valorar de manera breve y estructurada la sintomatología asociada al trastorno depresivo persistente o distimia en población universitaria. La escala presenta una propuesta unidimensional y está conformada por 12 ítems, elaborados para representar los criterios diagnósticos del trastorno. El ítem 1 evalúa el criterio temporal del estado de ánimo depresivo en los dos últimos años, mientras que los ítems 2 al 12 exploran síntomas asociados al trastorno depresivo persistente en las últimas semanas. En cuanto al formato de respuesta, el ítem 1 se responde en una escala ordinal de frecuencia que va de 0 = ningún día a 4 = todos los días, mientras que los ítems restantes presentan opciones tipo Likert desde 0 = nunca hasta 4 = siempre.

Según los estudios revisados, Elejabo-Pacheco et al. (2025) desarrollaron y validaron el instrumento en una muestra de 868 estudiantes universitarios peruanos de varias regiones del país reclutados de manera no probabilística.

Sobre la confiabilidad, la escala mostró excelente consistencia interna, con un alfa de Cronbach de 0,88 y un coeficiente omega de McDonald de 0,88.

Respecto a la evidencia basada en el contenido de la prueba, evaluaron la validez de contenido mediante el coeficiente V de Aiken, obteniendo valores superiores a 0,80 en relevancia, coherencia, claridad y contenido para todos los ítems. Los promedios fueron cercanos a 1,00; todos los ítems alcanzaron valores $\geq 0,89$ en relevancia y coherencia, $\geq 0,78$ en claridad y 1,00 en contenido, superando el criterio mínimo esperado de $V \geq 0,70$ y evidenciando una adecuada representación del constructo.

Sobre la evidencia basada en la estructura interna, el análisis factorial exploratorio identificó una estructura unifactorial compuesta por 12 ítems, con cargas factoriales entre 0,65 y 0,88, que explicó el 63% de la varianza total. Previamente, se verificó la adecuación de los datos para el análisis factorial mediante el índice KMO = 0,96 y la prueba de esfericidad de Bartlett, $\chi^2(66) = 2221.93$, $p < .001$. Posteriormente, el análisis factorial confirmatorio respaldó un modelo de un único factor latente, con $\chi^2 = 318.72$, $gl = 54$, $p < .001$, CFI = .98, TLI = .97, RMSEA = .09, IC 90% (.08, .10), y SRMR = .06; además, todas las saturaciones factoriales fueron mayores a .50.

En cuanto a la invarianza factorial según sexo, se evaluaron modelos configural, métrico, escalar y estricto. El modelo configural mostró $\chi^2 = 415.91$, $gl = 108$, CFI = .880, TLI = .854, RMSEA = .100 y SRMR = .058. El modelo métrico presentó $\chi^2 = 428.97$, $gl = 119$, CFI = .879, TLI = .866, RMSEA = .096 y SRMR = .066, con $\Delta\chi^2 = 13.06$, $\Delta gl = 11$, $p = .289$ y $\Delta CFI = -.001$. El modelo escalar obtuvo $\chi^2 = 448.26$, $gl = 130$, CFI = .876, TLI = .874, RMSEA = .093 y SRMR = .067, con $\Delta\chi^2 = 19.29$, $\Delta gl = 11$, $p = .056$ y $\Delta CFI = -.003$. Finalmente, el

modelo estricto alcanzó $\chi^2 = 464.76$, $gl = 142$, $CFI = .874$, $TLI = .883$, $RMSEA = .090$ y $SRMR = .067$, con $\Delta\chi^2 = 16.50$, $\Delta gl = 12$, $p = .169$ y $\Delta CFI = -.002$. En conjunto, las diferencias de CFI entre modelos consecutivos fueron menores a .01, lo que respalda la invarianza configural, métrica y estricta, y aporta evidencia parcial a favor de la invarianza escalar entre hombres y mujeres.

En relación con la evidencia basada en relaciones con otras variables, la validez convergente se evaluó mediante modelamiento de ecuaciones estructurales, encontrándose que la TDP covarió significativamente y en la dirección esperada con depresión medida por el PHQ-9, ansiedad evaluada con el GAD-7 y satisfacción con la vida valorada con la SWLS. En particular, la TDP mostró relaciones positivas con ansiedad (0,73) y depresión (0,84), y una relación negativa con satisfacción con la vida (-0,34). Además, el modelo presentó índices de ajuste adecuados: $\chi^2 = 1005,75$, $gl = 489$, $\chi^2/gl = 2,05$, $p < 0,001$, $CFI = 0,91$, $TLI = 0,91$, $SRMR = 0,050$ y $RMSEA = 0,043$ (IC 90%: 0,040-0,047), lo que respalda la validez convergente de la escala.

Inventario de Depresión Estado-Rasgo (IDER)

El Inventario de Depresión Estado-Rasgo (IDER) es un autoinforme breve que evalúa el componente afectivo de la depresión mediante dos constructos diferenciados: depresión como estado (situación emocional actual) y depresión como rasgo (tendencia estable a experimentar síntomas depresivos) (Spielberger et al., 2002a; Spielberger et al., 2002b; Spielberger et al., 2008). Se trata de una adaptación hispana del modelo de depresión estado-rasgo, con 20 ítems organizados en dos escalas de 10 reactivos cada una, Depresión Estado (IDER-E) y Depresión Rasgo (IDER-R), subdivididas en subescalas de distimia (afecto negativo) y eutimia (afecto positivo). Los ítems se responden en escala tipo Likert de 4 opciones que gradúan la intensidad (Estado) y la frecuencia (Rasgo) de los afectos, con puntajes por escala de 10-40 y total de 20-80, donde valores más altos indican mayor presencia de

afectos negativos y menor afecto positivo. El instrumento, dirigido a adolescentes, jóvenes y adultos, se centra exclusivamente en síntomas emocionales-afectivos, excluyendo manifestaciones somáticas, y requiere 7-10 minutos para su aplicación individual o colectiva.

Según los estudios revisados, Soto et al. (2012) realizaron una validación del instrumento en 314 adolescentes escolares limeños utilizando la adaptación colombiana de Agudelo et al. (2005) y Ocampo (2007), mientras que Sotelo et al. (2012) realizaron una adaptación y validación en 251 adultos limeños con categoría clínica (consultantes de psicología) y no clínica (estudiantes universitarios), ambos estudios por muestreo de tipo no probabilístico.

Sotelo et al. (2012) realizaron una adecuación lingüística de algunos ítems mediante jueces expertos con experiencia en psicología clínica, quienes revisaron y modificaron determinados reactivos conservando su propósito original, pero ajustándolos a la realidad lingüística de la población evaluada. No obstante, con la información disponible no se detallan de manera exhaustiva las etapas del proceso de adaptación ni se reportan indicadores cuantitativos de acuerdo entre jueces.

Para el análisis de confiabilidad/precisión se utilizó el coeficiente alfa de Cronbach en ambos estudios. En Soto et al. (2012), los valores fueron de 0.76 para Estado y 0.68 para Rasgo en la muestra total, con subescalas de 0.71 para Distimia-Estado, 0.80 para Eutimia-Estado, 0.67 para Distimia-Rasgo y 0.65 para Eutimia-Rasgo. Las correlaciones interítem promedio fueron de 0.232 para Estado y 0.189 para Rasgo, y las estimaciones de alfa resultaron menores en subgrupos por sexo y edad, particularmente en Rasgo. En Sotelo et al. (2012), los coeficientes alfa fueron de 0.864 para Eutimia-Estado, 0.861 para Distimia-Estado, 0.793 para Eutimia-Rasgo y 0.862 para Distimia-Rasgo, lo que indica niveles adecuados de consistencia interna.

Sotelo et al. (2012) evaluaron la evidencia de validez basada en el contenido de la prueba mediante juicio de expertos, obteniendo coeficientes V de Aiken entre 0,67 y 1,00 para los reactivos, con predominio de valores altos y acuerdo moderado a alto sobre su relevancia para medir el componente afectivo de la depresión. Algunos ítems presentaron índices relativamente menores, indicando menor uniformidad en la valoración de su pertinencia.

Respecto a la evidencia basada en la estructura interna, ambos estudios aplicaron análisis factorial exploratorio. En Soto et al. (2012), los índices de factorizabilidad fueron adecuados, con KMO de 0.80 para Estado y 0.76 para Rasgo, y pruebas de esfericidad de Bartlett significativas. Se identificó una solución bidimensional que explicó 52.53% de la varianza en Estado y 44.38% en Rasgo, con correlaciones interfactoriales negativas y cargas factoriales generalmente altas, aunque algunos ítems presentaron saturaciones menores a 0.45. En Sotelo et al. (2012), también se extrajeron dos factores por escala, explicando 65.38% de la varianza en Depresión-Estado y 60.23% en Depresión-Rasgo, con índices KMO de 0.854 y 0.866, respectivamente. De manera complementaria, las correlaciones ítem-test oscilaron entre 0.504 y 0.768 en las distintas subescalas, lo que respalda la homogeneidad interna de los reactivos.

Escala de Depresión para Adolescentes de Reynolds (EDAR)

La Escala de Depresión para Adolescentes de Reynolds (EDAR) en inglés *Reynolds Adolescent Depression Scale* (RADS) diseñada por Reynolds (1987) es un instrumento de autoreporte de 30 ítems, diseñado para evaluar la presencia e intensidad de síntomas depresivos en adolescentes de 13 a 18 años, cada ítem se responde en una escala Likert de 4 puntos (1 = casi nunca, 4 = casi siempre), con 7 ítems inversos (ítems 1, 5, 10, 12, 23, 25 y 29), y se obtiene una puntuación total comprendida entre 30 y 120, donde las puntuaciones más altas indican mayor sintomatología depresiva. La escala explora cuatro dimensiones principales: anhedonia/afecto negativo, disforia, quejas somáticas y autoevaluación negativa/baja

autoestima, y se aplica en 10 a 15 minutos de forma individual o colectiva, principalmente en contextos escolares y preventivos. La segunda edición del instrumento, denominada RADS-2, mantuvo el contenido de los ítems de la versión original, aunque reorganizó su fundamento teórico en cuatro dimensiones: anhedonia/afecto negativo, disforia, quejas somáticas y autoevaluación negativa. Además, extendió el rango etario de evaluación de 11 a 20 años y se alineó con criterios diagnósticos más recientes que la versión original (Ramos-Vera et al., 2023).

Según la literatura revisada, la primera validación en nuestro país sobre el EDAR fue realizada por Chávez y Ecurra (2002) en una muestra de 1963 estudiantes limeños de secundaria adolescentes, esta versión es posteriormente utilizada por Ramos-Vera et al. (2023) para realizar una validación en 917 adolescentes escolares limeños, ambos estudios con muestreo no probabilístico, sin embargo, la identificación exacta del instrumento en este segundo estudio resulta parcialmente confusa, debido a que los autores mencionan tanto la versión peruana previamente adaptada como la RADS-2. En conjunto, esto sugiere que se tomó como base la adaptación previa, aunque el estudio fue presentado metodológicamente como una evaluación de la RADS-2.

Para el análisis de confiabilidad/precisión, Chávez y Ecurra (2002) reportaron un alfa de Cronbach de 0.87 para el puntaje total de la EDAR, con valores entre 0.86 y 0.90 en colegios particulares y 0.76-0.79 en varones de primeros grados de secundaria; la confiabilidad por división en mitades corregida mediante la fórmula de Spearman-Brown fue de 0.85 en la muestra total y 0.86 en mujeres/colegios particulares, y test-retest de 0.71 (1°-2° secundaria) y 0.89 (3°-5° secundaria) a dos semanas. Por su parte, Ramos-Vera et al. (2023) calcularon coeficientes omega para el RADS-2, una medida de 25 ítems, obteniendo $\omega = 0.93$ para el puntaje total y valores entre 0.70 y 0.86 para las subescalas: disforia ($\omega = 0.83$), anhedonia ($\omega = 0.70$), autoevaluación negativa ($\omega = 0.86$) y quejas somáticas ($\omega = 0.74$).

Para la evidencia basada en la estructura interna, Chávez y Ecurra (2002) verificaron la adecuación muestral de la EDAR ($KMO = 0.927$, $\chi^2 = 13100.207$, $p < 0.001$) y aplicaron análisis de componentes principales con rotación Varimax, identificando seis factores que explicaron el 45.786% de la varianza. Un análisis factorial de segundo orden extrajo dos componentes que explicaron el 66.231% de la varianza, confirmado mediante CFA con $\chi^2 = 3.52$, $p = 0.318$; de manera adicional, las correlaciones ítem-total fueron todas significativas ($p < 0.001$), con 22 de 30 ítems superando 0.40 en la muestra total. Por otro lado, Ramos-Vera et al. (2023) confirmaron la estructura de cuatro dimensiones del RADS-2 mediante CFA, ajustando el modelo original de 30 ítems a 25 ítems tras eliminar cinco reactivos con cargas < 0.40 (ítems 2, 10, 23, 25, 29). El modelo final presentó cargas entre 0.55 y 0.82, con índices de ajuste $\chi^2/df = 1.49$, CFI = 0.996, TLI = 0.995, RMSEA = 0.023 (IC 95%: 0.018–0.028) y SRMR = 0.038.

Para la invariancia factorial del modelo RADS-2 de 25 ítems (M2), Ramos-Vera et al. (2023) evaluaron la equivalencia de medición por género mediante cuatro modelos progresivos. El modelo configural sin restricciones presentó $\chi^2/df = 0.96$, CFI = 1.00, RMSEA = 0.00 y SRMR = 0.044, sirviendo como base de comparación. El modelo de invariancia métrica mostró índices de ajuste adecuados ($\chi^2/df = 4.01$, CFI = 0.998, RMSEA = 0.014, SRMR = 0.048), con diferencias aceptables frente al configural. Tanto la invariancia escalar ($\chi^2/df = 2.003$, CFI = 0.999, RMSEA = 0.012, SRMR = 0.048; $\Delta CFI \leq 0.001$, $\Delta RMSEA \leq 0.003$, $\Delta SRMR = 0.000$) como la invariancia residual ($\chi^2/df = 1.46$, CFI = 0.999, RMSEA = 0.009, SRMR = 0.048; $\Delta CFI = 0.000$, $\Delta RMSEA \leq 0.003$, $\Delta SRMR = 0.000$) presentaron ajustes satisfactorios y diferencias dentro de los criterios establecidos.

Chávez y Ecurra (2002) elaboraron normas percentilares para la EDAR en estudiantes secundarios limeños, pero no establecieron categorías interpretativas de severidad clínica, por

lo que no son directamente convertibles ni utilizables en la práctica clínica o la investigación y requerirían interpretación adicional.

Inventario de Depresión para Niños - versión corta (CDI-S)

El CDI-S, versión corta del Children's Depression Inventory (CDI) original desarrollado por Kovacs (1992) y adaptado al español por Del Barrio et al. (2002) para población española, es una escala autoaplicada diseñada para evaluar síntomas nucleares de depresión en niños y adolescentes durante las últimas dos semanas. Consta de 10 ítems seleccionados del CDI original que cubren principalmente autoestima negativa (ítems 10, 4, 7, 8), anhedonia (ítems 5, 1, 6) y desesperanza (ítems 2, 3, 9); cada ítem presenta tres opciones de respuesta en formato Likert con intensidades crecientes (0=ausencia del síntoma, 1=baja intensidad/frecuencia, 2=alta intensidad/frecuencia), donde algunos requieren inversión de puntuación para reflejar adecuadamente la sintomatología depresiva. Se administra de forma individual en 10-15 minutos, y su puntuación total se obtiene sumando los valores de los ítems (rango 0-20), de modo que puntajes más altos indican mayor severidad depresiva.

Según la literatura revisada, Ventura-León et al. (2020) realizan la primera validación de la versión española del CDI-S de Del Barrio et al. (2002) sobre una muestra de 1059 niños escolares de Lima y Callao, posteriormente Merino-Soto et al. (2022) también validan esta misma versión, pero en 189 adolescentes limeños de secundaria, comparando grupos anónimo (n=100) vs identificado (n=89), a fin de analizar las variaciones que el anonimato produce en la psicometría del test; ambos estudios con muestreo no probabilístico.

Como parte de un proceso de adaptación cultural, Ventura-León et al. (2020) reportaron que la adaptación del CDI-S siguió las directrices de la *International Test Commission* a través de una prueba piloto, incluyendo entrevistas semiestructuradas con 10 niños para evaluar la comprensión de los ítems y reducir sesgos lingüísticos, a partir de este proceso solo se realizó

una modificación menor en el ítem 7, añadiendo el término “físico” a la frase “tengo buen aspecto” para mejorar su equivalencia semántica en población peruana.

Como parte del análisis de coherencia interna, Ventura-León et al. (2020) estimaron el coeficiente omega en dos momentos del análisis. En la fase exploratoria obtuvieron $\omega = 0.85$ para disforia y $\omega = 0.82$ para autoestima negativa, mientras que, en la fase confirmatoria, con el modelo bidimensional final, los coeficientes omega fueron 0.89 para disforia y 0.81 para autoestima negativa, lo que indica niveles adecuados de coherencia interna en ambas dimensiones del CDI-S. Por su parte, Merino-Soto et al. (2022) informaron coeficientes alfa de Cronbach de 0.741 en el grupo anónimo y 0.633 en el grupo identificado, sin diferencias significativas entre ellos, así como coeficientes omega de 0.898 y 0.835 respectivamente, también similares entre grupos.

Para la evidencia basada en la estructura interna, Ventura-León et al. (2020) reportaron un índice KMO de 0.72 y una prueba de esfericidad de Bartlett significativa ($p < 0.001$), lo que respaldó la factorizabilidad de la matriz de correlaciones. El análisis factorial exploratorio, guiado por análisis paralelo, identificó una solución de dos factores que explicó el 59.03% de la varianza total, con un factor de disforia (ítems 2, 5, 6 y 7; cargas 0.43–0.90, eigenvalue 3.90, 39.05% de varianza) y un factor de autoestima negativa (ítems 1, 3, 4, 8, 9 y 10; cargas 0.42–0.84, eigenvalue 2.00, 19.98% de varianza), correlacionados moderadamente ($\phi = 0.39$) y con comunalidades entre 0.17 y 0.85.

Por su parte, Merino-Soto et al. (2022) verificaron una estructura unifactorial del CDI-S mediante análisis factorial confirmatorio robusto con matriz policórica y corrección de Satorra-Bentler, obteniendo en el grupo anónimo $\chi^2 = 34.63$, CFI = 1.00 y RMSEA = 0.00 (IC 90%: 0.00–0.07), y en el grupo identificado $\chi^2 = 36.11$, CFI = 0.99 y RMSEA = 0.01. Las cargas factoriales variaron entre 0.377 y 0.876 (todas superiores a 0.30) y la congruencia métrica entre grupos fue muy alta ($\phi = 0.989$), lo que indica que la estructura de medición se

mantiene prácticamente igual bajo condiciones anónimas y con nombre. De manera complementaria, evaluaron la invarianza de medición entre las aplicaciones anónima e identificada mediante análisis de funcionamiento diferencial de los ítems. Los modelos de regresión logística ordinal no evidenciaron DIF no uniforme ($\Delta-2LL$ con $p > 0.005$ tras corrección de Bonferroni) ni DIF uniforme (diferencias de betas $|\Delta\beta| < 10\%$) en ninguno de los 10 ítems, respaldando la ausencia de sesgo por condición de aplicación. A nivel de puntajes totales, se observó equivalencia descriptiva entre grupos ($r = 0.23$, $p = 0.185$; $d = 0.10$), lo que sugiere diferencias triviales en las medias del CDI-S entre aplicaciones anónimas y con nombre.

Finalmente, respecto a la relación con otras variables, Ventura-León et al. (2020) correlacionaron las dimensiones del CDI-S con las dimensiones correspondientes de la versión extensa de 27 ítems. Se encontraron correlaciones positivas fuertes entre autoestima negativa de la versión breve y la extensa ($r = 0.74$) y entre disforia de ambas versiones ($r = 0.62$), lo que respalda la coherencia conceptual y empírica entre la forma abreviada y la versión completa del CDI.

Escala de Depresión Posnatal de Edimburgo (EPDS)

La Escala de Depresión Postnatal de Edimburgo (Edinburgh Postnatal Depression Scale, EPDS) es un instrumento de tamizaje desarrollado por Cox et al. (1987), originalmente en idioma inglés, para la detección de sintomatología depresiva en el posparto (puerperio), y que también se utiliza en la práctica clínica e investigación durante el embarazo (anteparto). Es una escala unifactorial de autoinforme de 10 ítems que indaga síntomas afectivos y cognitivos relevantes en depresión perinatal (p. ej., anhedonia, tristeza, culpa, ansiedad/pánico, sensación de sobrecarga, dificultades del sueño, llanto e ideación de autolesión), poniendo menos énfasis en síntomas somáticos que pueden confundirse con cambios propios del embarazo o del posparto. Evalúa la experiencia de los últimos 7 días, lo que permite su repetición para

seguimiento; cada ítem se puntúa de 0 a 3 y el puntaje total varía de 0 a 30, donde valores más altos indican mayor carga de síntomas. En la puntuación, los ítems 1, 2 y 4 se califican en sentido directo, mientras que los ítems 3 y 5–10 se califican en sentido inverso.

Según los estudios revisados, Vega-Dienstmaier et al. (2002) es el primer estudio peruano identificado que emplea una traducción en español de esta escala y la valida en 321 madres limeñas en posparto temprano (<1 año) por selección no probabilística; esta traducción es utilizada posteriormente por Lam et al. (2010) en 593 gestantes limeñas por muestreo probabilístico y Zhong et al. (2014) en 1517 mujeres limeñas de primer control de gestación por muestreo no probabilístico; sin embargo, el artículo de Vega-Dienstmaier et al. (2002) no detalla exhaustivamente los procedimientos de traducción y adaptación.

Para el análisis de confiabilidad/precisión se utilizó el coeficiente alfa de Cronbach, con valores entre 0,704 y 0,836 en las investigaciones revisadas (Vega-Dienstmaier et al., 2002; Lam et al., 2010; Zhong et al., 2014).

Respecto a la estructura interna, la EPDS fue propuesta originalmente para interpretarse mediante un puntaje unidimensional. En Lam et al. (2010) se aplicó un análisis factorial exploratorio (AFE) que identificó una estructura de dos factores (componente ansioso y componente depresivo) que explicó 54,67% de la varianza; además, reportaron alfa de 0,756 para el factor ansioso y 0,792 para el factor depresivo. En Zhong et al. (2014), el AFE mostró una solución de 2 factores (anhedonia y ansiedad/depresión) que explicó 50% de la varianza; sin embargo, el análisis factorial confirmatorio (CFA) indicó que un modelo de 3 factores (anhedonia: ítems 1–2; ansiedad: ítems 3–6; ánimo deprimido: ítems 7–10) presentaba mejor ajuste (SRMR = 0,023; RMSEA = 0,034; CFI = 0,985). En conjunto, estos hallazgos apoyan una estructura multidimensional de la EPDS durante el embarazo.

De manera complementaria, Lam et al. (2010) reportaron correlaciones corregidas ítem–total entre 0,46 y 0,77, y para Zhong et al. (2014) las correlaciones ítem–total oscilaron

entre 0,50 y 0,75 ($p<0,0001$), lo que respalda la dirección de la puntuación de los ítems respecto a la escala global y la estructura interna de la EPDS en gestantes.

Respecto a la evidencia basada en relaciones con otras variables, en el estudio de Vega-Dienstmaier et al. (2002) se observó una asociación positiva de magnitud moderada entre el puntaje total de la EPDS y el número de síntomas del criterio A para depresión mayor, con $r=0,4478$ ($p<0,001$), lo que sugiere que mayores puntajes en EPDS se relacionan con mayor carga sintomática depresiva. Así mismo, en Zhong et al. (2014), la relación entre EPDS y PHQ-9 fue también moderada (Spearman = 0,52; $p<0,0001$), aportando evidencia de validez convergente al mostrar consistencia con otra medida de sintomatología depresiva durante el inicio del embarazo.

Finalmente, en Vega-Dienstmaier et al. (2002) se evaluó el desempeño de clasificación de la EPDS frente a un criterio clínico externo (SCID para DSM-IV para episodio depresivo mayor), identificándose un punto de corte óptimo de 13,5 (usualmente interpretado como $EPDS \geq 14$) con sensibilidad 84,21% y especificidad 79,47%. Dado que este estudio reporta el rendimiento para múltiples puntos de corte, sus resultados pueden orientar la selección de umbrales en poblaciones comparables según el objetivo del tamizaje; en contraste, Lam et al. (2010) y Zhong et al. (2014), al no incorporar un criterio externo de referencia, no permiten estimar sensibilidad/especificidad para episodio depresivo mayor.

Escala de Autoevaluación para la Depresión de Zung (ZSDS)

La Zung Self-Rating Depression Scale (ZSDS), desarrollada por Zung (1965), es un instrumento de autoinforme de 20 ítems diseñado para el tamizaje y cuantificación de síntomas depresivos, especialmente útil en contextos clínicos y de atención primaria. Evalúa manifestaciones afectivas, cognitivas y somáticas/psicofisiológicas de la depresión, incluyendo también componentes psicomotores, y solicita al evaluado responder según la frecuencia de los síntomas en los últimos días o la última semana mediante una escala tipo Likert de 4 puntos;

incluye ítems redactados en sentido positivo y negativo (ítems 2, 5, 6, 11, 12, 14, 16, 17, 18, 20 requieren inversión al puntuar). La puntuación total suele interpretarse como puntaje bruto (rango 20–80), por lo que, a mayor puntaje, mayor severidad sintomática. Originalmente, la ZSDS se concibió como una escala de puntaje total, y la evidencia factorial posterior ha mostrado estructuras dimensionales variables según la muestra y el método de análisis (Romera et al., 2008).

Según los estudios revisados, Ruiz-Grosso et al. (2012) constituyen la primera validación publicada en contexto peruano de la ZSDS, en una muestra de 194 adultos pacientes ambulatorios de un hospital de Lima, empleando una versión en español reportada como estandarizada y utilizada por el Instituto Nacional de Salud Mental (INSM) en Lima; sin embargo, no se detallan los procedimientos ni resultados de dicha estandarización realizada por el INSM ni se brinda una referencia específica para su acceso. Posteriormente, Del Águila et al. (2021) evaluaron las propiedades psicométricas de la escala en una muestra clínica de 203 pacientes adultos con VIH, aplicando la adaptación utilizada por Rivera et al. (2007); en ambos estudios se utilizó muestreo no probabilístico.

En confiabilidad, Ruiz-Grosso et al. (2012) reportaron consistencia interna adecuada ($\alpha = 0,89$; ω jerárquico = 0,75). Del Águila et al. (2021) informaron consistencia interna alta ($\alpha = 0,855$; $\Omega = 0,869$) en pacientes con VIH.

Respecto a la evidencia basada en el contenido, Del Águila et al. (2021) reportaron un coeficiente de validez de contenido (CVC) = 0,98, señalando participación de cinco especialistas que evaluaron criterios de pertinencia, relevancia y claridad; no obstante, no se describen los formatos, el procedimiento de calificación ni el método de cálculo del CVC.

En evidencia basada en la estructura interna, Ruiz-Grosso et al. (2012) reportaron indicadores de funcionamiento de ítems (correlaciones ítem-test 0,18–0,77; ítem-resto 0,10–0,73; covarianza inter-ítems promedio 0,26–0,30), destacando el ítem 2 como el de peor

desempeño (α aumentaría a 0,90 si se retira) y el ítem 18 como el de mejor desempeño, lo que indica una contribución variable pero generalmente adecuada de los ítems al constructo depresión. Del Águila et al. (2021) informaron índices de homogeneidad corregida $>0,30$ y comunalidades $>0,40$, y evaluaron tres modelos confirmatorios iniciales (4 factores oblicuos, 3 factores y unifactorial) sin ajuste adecuado; posteriormente, el AFE ($KMO = 0,873$; Bartlett $p < 0,001$) sugirió un modelo de 2 factores, el cual presentó buen ajuste ($\chi^2 = 273,2$; $gl = 169$; $CFI = 0,917$; $TLI = 0,907$; $RMSEA = 0,055$; $RMR = 0,056$), aunque no se presenta una descripción detallada y sistemática de la asignación de ítems/cargas factoriales por factor en el modelo final.

En evidencia basada en relaciones con otras variables, Ruiz-Grosso et al. (2012) reportaron asociación fuerte entre ZSDS y Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D) ($\rho = 0,86$; $p < 0,001$) y correlación moderada con la severidad clínica evaluada mediante la Clinical Global Impression-Severity (CGI-S) ($\rho = 0,50$; $p < 0,001$), sugiriendo que mayores puntajes en Zung se relacionan con mayor gravedad clínica. Del Águila et al. (2021) reportaron convergencia con una medida de ansiedad ($\rho = 0,718$; $p < 0,001$) y divergencia con una medida de bienestar psicológico ($\rho = -0,623$; $p < 0,001$), aunque sin especificar los instrumentos utilizados para ansiedad y bienestar, lo que limita la interpretación de esta evidencia.

Finalmente, en capacidad discriminativa, Ruiz-Grosso et al. (2012) informaron un auROC de 0,96 (IC95%: 0,93–0,99) para discriminar entre participantes con episodio depresivo mayor (MDE) y participantes sin evidencia de trastornos psiquiátricos (NEP), y propusieron punto de corte ≥ 45 (sensibilidad 91,4%; especificidad 91,8%), indicando un desempeño elevado para la detección de depresión en este contraste. En cuanto a normas, Del Águila et al. (2021) establecieron categorías por percentiles (P25, P50, P75) en pacientes con VIH: leve (20–32), moderada (33–48) y severa (≥ 49).

Escala de Psicopatología Depresiva (EPD/DPS)

La Escala de Psicopatología Depresiva (EPD) o *Depressive Psychopathology Scale* (DPS) es un instrumento desarrollado por Vega-Dienstmaier et al. (2011) para evaluar sintomatología depresiva, y fue sometida a una validación inicial en una muestra clínica peruana de pacientes psiquiátricos ambulatorios no psicóticos (18–60 años) atendidos en Lima, Perú (1999–2001). La construcción de la escala se basó en criterios estandarizados del DSM-IV (incluyendo características melancólicas y atípicas) y en la sintomatología del CIE-10 para episodio depresivo. Compuesta por 20 ítems de autorreporte referidos a síntomas experimentados en las últimas dos semanas, la escala califica sus reactivos generalmente en un formato tipo Likert de 0 a 3, con excepciones en los ítems 1 y 16 (0 a 2) y el ítem 20 (0 a 4); además, mientras los ítems 1 y 20 evalúan principalmente severidad, el resto se puntúa en función de la frecuencia. En el estudio de validación, el tiempo medio de aplicación fue 7,22 minutos (DE = 3,99). Su estructura interna se organizó en cinco factores principales: depresión (ítems 15, 18, 11, 10, 19 y 20), anergia (4, 5, 6 y 9), inquietud/*uneasiness* (8, 14 y 12), insomnio (2 y 3) y ausencia de afectos positivos (13 y 7). Finalmente, en el análisis de componentes principales se reportó que tres ítems – 1 (alteraciones del apetito), 16 (depresión peor por la mañana) y 17 (reactividad del ánimo) – presentaron cargas factoriales bajas ($< 0,40$) en los cinco factores.

La versión abreviada de 4 ítems (EPD-4), seleccionada de la EPD original, está compuesta por fatiga, reactividad del ánimo, anhedonia y tristeza. Mantiene el formato Likert de 0-3 por ítem y fue diseñada para el tamizaje eficiente de depresión mayor, preservando la capacidad de evaluar los componentes centrales del constructo depresivo en un formato ultra-breve.

Por su parte, la versión de 6 ítems (EPD-6) incorpora a la EPD-4 los ítems de alteración del sueño y pesadez de extremidades, conservando el formato Likert de 0-3. Su estructura

interna se organiza en dos factores: uno corporal (fatiga, pesadez de extremidades) y otro afectivo-sueño (tristeza, anhedonia, reactividad del ánimo, alteración del sueño). Fue desarrollada específicamente para optimizar la detección de depresión mayor utilizando el diagnóstico clínico como criterio de referencia.

Según los estudios revisados, Vega-Dienstmaier et al. (2011) llevaron a cabo la elaboración y primera validación de la Escala de Psicopatología Depresiva (EPD/DPS) de 20 ítems en contexto clínico limeño con una muestra de 226 pacientes psiquiátricos ambulatorios no psicóticos por muestreo no probabilístico. A partir de esa base de datos y, por ende, en la misma muestra, Vega-Dienstmaier (2018) desarrolló dos versiones abreviadas (la EPD-4 y la EPD-6), posteriormente, Lee-Díaz et al. (2024) validaron la EPD-4 mediante aplicación online en 256 participantes de población general adulta peruana general por reclutamiento no probabilístico.

En cuanto a la confiabilidad, Vega-Dienstmaier et al. (2011) reportaron consistencia interna adecuada para la EPD/DPS de 20 ítems, con α de Cronbach = 0,86 a nivel global y por dimensiones: depresión ($\alpha = 0,76$), anergia ($\alpha = 0,702$), inquietud ($\alpha = 0,63$), insomnio ($\alpha = 0,681$) y ausencia de afectos positivos ($\alpha = 0,538$). Por su parte, Vega-Dienstmaier (2018) obtuvo $\alpha = 0,6603$ para la EPD-4 y $\alpha = 0,7359$ para la EPD-6. Finalmente, Lee-Díaz et al. (2024) informaron para la EPD-4 valores apropiados de ω de McDonald = 0,8229 y $\alpha = 0,8126$; además, el análisis de alfa si el ítem es eliminado mostró rangos entre 0,7226 y 0,7812, destacando que la exclusión del ítem de anhedonia produjo la mayor disminución, lo que evidencia su relevante contribución a la fiabilidad global del instrumento.

Respecto a la evidencia basada en la estructura interna, la EPD/DPS fue propuesta originalmente como multidimensional. En Vega-Dienstmaier et al. (2011), mediante análisis de componentes principales con rotación Varimax se identificó una solución de cinco factores ($KMO = 0,857$; Bartlett $p < 0,001$): depresión (ítems 10, 11, 15, 18, 19, 20), anergia (ítems 4,

5, 6, 9), inquietud (ítems 8, 12, 14), insomnio (ítems 2, 3) y ausencia de afectos positivos (ítems 7, 13); ítems 1, 16 y 17 mostraron cargas bajas ($<0,40$), y el factor depresión correlacionó significativamente con los demás ($\rho = 0,294-0,541$; $p < 0,01$). En Vega-Dienstmaier (2018), la EPD-4 presentó estructura unifactorial, mientras que la EPD-6 identificó dos factores (corporal: fatiga, pesadez; afectivo-sueño: reactividad del ánimo, anhedonia, tristeza, alteración del sueño). De manera concordante, Lee-Díaz et al. (2024) confirmaron la unidimensionalidad de la EPD-4 mediante AFE (eigenvalue = 2,64 para factor 1; cargas factoriales entre 0,7497-0,8897, máxima en anhedonia) y AFC (RMSEA = 0,055 [IC 90%: 0,000-0,148]; CFI = 0,999; TLI = 0,996; SRMR = 0,019), con correlaciones ítem-resto entre 0,6004 y 0,7182. En conjunto, estos hallazgos apoyan flexibilidad estructural según longitud de la escala y población.

Respecto a la evidencia basada en relaciones con otras variables, Vega-Dienstmaier et al. (2011) respaldaron la validez convergente de la EPD/DPS mediante correlaciones de Spearman del puntaje total con ZSDS ($\rho = 0,804$; $p < 0,001$), CGI-S ($\rho = 0,621$; $p < 0,001$) y severidad psiquiátrica ($\rho = 0,589$; $p < 0,001$), indicando convergencia moderada-alta con medidas externas. Asimismo, Vega-Dienstmaier (2018) reportó para EPD-4 correlaciones de Pearson con EPD-20 ($r = 0,8623$; $p < 0,0001$) y EDZ (Escala de Depresión de Zung) ($r = 0,7186$; $p < 0,0001$), y para EPD-6 con EPD-20 ($r = 0,8890$; $p < 0,0001$) y EDZ ($r = 0,7641$; $p < 0,0001$). En Lee-Díaz et al. (2024), la EPD-4 correlacionó positivamente con CES-D ($\rho = 0,8255$; $p < 0,001$), lo que respalda su consistencia con otra escala de síntomas depresivos en población general.

Vega-Dienstmaier et al. (2011) evaluaron el poder discriminativo de la EPD/DPS frente a criterios diagnósticos para trastorno depresivo mayor (TDM); tomando el módulo SCID como estándar de referencia, se obtuvo $\text{auROC} = 0,872$ ($\text{EE} = 0,024$), con punto de corte 26/27 (sensibilidad 81,32%; especificidad 80,0%); con diagnóstico psiquiátrico, el auROC fue 0,832

(EE = 0,028), con punto de corte 25/26 (sensibilidad 77,67%; especificidad 72,32%). De manera complementaria, Vega-Dienstmaier (2018) seleccionó ítems para versiones cortas maximizando auROC; para EPD-4 (fatiga, reactividad del ánimo, anhedonia, tristeza) frente a SCID se alcanzó auROC = 0,9024 (IC 95%: 0,8647–0,9419), con punto de corte >6 (Youden = 0,6264; sensibilidad 82,8%; especificidad 79,8%). Para EPD-6 (EPD-4 + alteración del sueño, pesadez de extremidades) frente al psiquiatra, el auROC = 0,8611 (IC 95%: 0,8185–0,9119), con punto de corte >8 (Youden = 0,5387; sensibilidad 80,2%; especificidad 73,7%). Finalmente, Lee-Díaz et al. (2024) reportaron para EPD-4 frente a presencia de depresión según CES-D un auROC de 0,9243 (IC 95%: 0,8899–0,9587), con corte óptimo 4 (Youden = 0,70; sensibilidad 82,19%; especificidad 87,43%). En conjunto, la EPD/DPS y sus versiones cortas muestran excelente capacidad clasificatoria (auROC >0,80) en muestras clínicas y comunitarias peruanas.

Inventario de Depresión de Beck (BDI)

El Inventario de Depresión de Beck, o Beck Depression Inventory (BDI o BDI-IA), publicado originalmente por Beck et al. (1961), constituyó un instrumento fundamental en la evaluación de la depresión, ya que puso énfasis en las manifestaciones cognitivas y conductuales observadas clínicamente. Está conformado por 21 ítems de opción múltiple, en los que cada reactivo representa un síntoma o actitud característica de la depresión e incluye tanto componentes cognitivo-afectivos como somático-conductuales. Las respuestas se puntúan de 0 a 3, por lo que la puntuación total oscila entre 0 y 63 puntos, y su propósito es evaluar la presencia e intensidad de síntomas depresivos durante los últimos siete días.

El BDI-II, publicado por Beck et al. (1996) y adaptado al español por Brenlla y Rodríguez (2006), constituye la revisión más importante de esta escala. Su desarrollo respondió principalmente a la necesidad de adecuar el instrumento a los criterios diagnósticos del DSM-IV, por lo que, a diferencia de la versión original de 1961, evalúa la depresión conforme a una

formulación clínica más actual. En esta versión se eliminaron algunos ítems antiguos, como la distorsión de la imagen corporal, la hipocondría, la pérdida de peso y la disminución de la actividad, y se incorporaron otros nuevos, entre ellos agitación, desvalorización o inutilidad, dificultad de concentración y pérdida de energía. Además, el periodo de referencia se amplió a las dos últimas semanas, incluido el día de la evaluación, en concordancia con el tiempo mínimo requerido para identificar un episodio depresivo mayor. Otra mejora relevante fue que permitió registrar tanto incrementos como disminuciones en funciones como el sueño y el apetito, manteniendo los 21 ítems con una escala de respuesta de 0 a 3; se trata de un cuestionario de autorreporte que suele completarse en aproximadamente 5 a 10 minutos. El BDI-IIA es una adaptación lingüística peruana realizada por Carranza (2013) en base al BDI-II con ajustes lingüísticos/culturales no especificados.

Por su parte, el BDI-PC (Beck Depression Inventory for Primary Care), también denominado BDI-FS (Fast Screen), fue desarrollado por Beck et al. (2000) como una versión breve orientada específicamente a contextos de medicina general y atención primaria. Su objetivo principal es detectar depresión en pacientes que consultan por problemas físicos, reduciendo la posibilidad de que síntomas médicos, como el cansancio o la pérdida de peso, incrementen artificialmente la puntuación total. Consta de solo 7 ítems seleccionados del BDI-II y se centra exclusivamente en síntomas cognitivos y afectivos, como tristeza, pesimismo, errores del pasado, disgusto por sí mismo, autocrítica, deseos o pensamientos suicidas y pérdida de interés. Cada reactivo se puntúa de 0 a 3, de modo que el rango total va de 0 a 21 puntos, y su aplicación es muy breve, ya que por lo general puede completarse en menos de 2 o 3 minutos.

Según la literatura revisada, Vega-Dienstmaier et al. (2014) validaron una versión en español del BDI original en 136 pacientes hospitalizados de medicina general en Lima por muestreo no probabilístico. Para ello, emplearon la versión estandarizada por el Instituto

Nacional de Salud Mental Honorio Delgado-Hideyo Noguchi, elaborada por Novara et al. (1985); no obstante, dicha versión no se encuentra actualmente accesible en formato digital.

En cuanto al BDI-II, se identificaron tres estudios peruanos en poblaciones distintas. Sánchez-Pachas (2019) analizó sus propiedades psicométricas en 400 pacientes atendidos en consulta externa de un centro especializado en salud mental de Lima. Por su parte, Sánchez-Villena y Farfán (2019) lo evaluaron en 326 estudiantes universitarios de Cajamarca, mientras que Sánchez-Villena et al. (2022) examinaron su estructura factorial y datos normativos en 1665 participantes de población general procedentes de diversas regiones del país. En conjunto, estos estudios aportan evidencia del BDI-II en contextos clínicos, universitarios y comunitarios, todos realizados por muestreo no probabilístico.

Respecto al BDI-IIA, Carranza (2013) desarrolló esta forma en 2005 estudiantes universitarios de Lima, conservando los 21 ítems del inventario, aunque sin precisar con detalle las modificaciones introducidas. Posteriormente, Rodríguez y Farfán (2015) validaron el BDI-IIA en 300 adolescentes escolares de Huancayo, ambos estudios con muestreo no probabilístico.

Finalmente, Lam et al. (2009) validaron el BDI-PC en 307 mujeres gestantes atendidas en Lima Metropolitana, también por selección no probabilística.

Para el BDI, Vega-Dienstmaier et al. (2014), reportaron una adecuada consistencia interna del BDI en la muestra total, con un alfa de Cronbach de 0.889. Así mismo, informaron correlaciones ítem-total corregidas para los 21 ítems del BDI, en general positivas y estadísticamente significativas. Los coeficientes oscilaron entre 0.2556 en el ítem 19 (pérdida de peso) y 0.6656 en el ítem 5 (sentimientos de culpa), observándose también valores relativamente bajos en irritabilidad (0.3438) y expectativas futuras (0.3655). En conjunto, estos resultados sugieren una adecuada homogeneidad de los ítems, aunque algunos reactivos mostraron menor asociación con el puntaje total.

Tomando como criterio el diagnóstico de depresión mayor según DSM-IV mediante la SCID-IV, propusieron un punto de corte de 18.5/19.5 para el BDI en pacientes hospitalizados de medicina general. Con este punto de corte, la sensibilidad fue 87.5% (IC 95%: 67.6-97.3) y la especificidad 98.21% (IC 95%: 93.7-99.8), con un valor predictivo positivo de 91.3% y un valor predictivo negativo de 97.35%. Además, ningún paciente con puntuación menor de 7.5 presentó depresión mayor y todos los sujetos con puntuaciones superiores a 24.5 fueron casos confirmados, lo que respalda la utilidad del instrumento para el tamizaje de depresión mayor en pacientes médicos hospitalizados peruanos.

Para el BDI-II, Sánchez-Pachas (2019) realizó una prueba piloto en 35 personas atendidas en el servicio de psicología de dos centros de salud, uno especializado y otro de atención primaria. Aunque no se detalla el procedimiento, se reportó que no hubo objeciones al responder el cuestionario, lo que sugiere adecuada comprensión de los ítems en esta fase preliminar.

Sánchez-Pachas (2019) reportó un alfa de Cronbach de 0.93, lo que indica alta consistencia interna, sin eliminación de ninguno de los 21 ítems. Sánchez-Villena y Farfán (2019) hallaron coeficientes alfa de 0.945 para un modelo unifactorial y, en el modelo de dos factores, valores de 0.843 para el componente somático y 0.935 para el componente cognitivo-afectivo. Por su parte, Sánchez-Villena et al. (2022) estimaron un coeficiente omega de 0.929 para el modelo unidimensional.

Sánchez-Pachas (2019) evaluó la validez de contenido mediante juicio de cinco psicólogos especialistas en depresión, obteniendo un coeficiente V de Aiken de 0.99 y considerando válidos el 100% de los ítems. No obstante, no se detallaron los procedimientos específicos empleados para la evaluación de los reactivos.

Sánchez-Pachas (2019) reportó correlaciones ítem-total corregidas significativas para todos los ítems, con valores entre 0.41 en el ítem 21 (pérdida de interés sexual) y 0.70 en el

ítem 1 (tristeza). Asimismo, el análisis factorial exploratorio mostró adecuados indicadores de factorizabilidad ($KMO = 0.954$; Bartlett $\chi^2 = 3603.2$; $p < 0.001$) e identificó inicialmente dos factores que explicaban el 48.46% de la varianza total; sin embargo, dado que todos los ítems saturaron principalmente en el primer factor con cargas superiores a 0.40, los autores concluyeron que la estructura del BDI-II era esencialmente unidimensional.

Sánchez-Villena y Farfán (2019) también reportaron adecuados índices previos al análisis factorial ($KMO = 0.958$ y Bartlett significativa), hallando primero un modelo unifactorial que explicaba el 48.16% de la varianza y luego un modelo de dos factores que explicaba el 54.48%. La alta correlación entre ambos factores ($r = 0.842$) llevó a los autores a considerar que, pese al mejor ajuste del modelo bidimensional, existía también sustento para una interpretación unidimensional del inventario.

Sánchez-Villena et al. (2022) evaluaron mediante análisis factorial confirmatorio catorce modelos previos del BDI-II con estimador WLSMV para datos ordinales. Los modelos bidimensionales de Rodríguez-Farfán ($CFI = 0.976$, $TLI = 0.973$, $RMSEA = 0.043$; r interfactorial = 0.883) y de Rosas-Santiago ($CFI = 0.981$, $TLI = 0.979$, $RMSEA = 0.036$; $r = 0.790$) mostraron los mejores índices de ajuste, mientras que el modelo unidimensional de Barreda también presentó ajuste aceptable ($CFI = 0.955$, $TLI = 0.950$, $RMSEA = 0.055$), con todas las cargas factoriales superiores a 0.53. Los modelos bifactor y de segundo orden no convergieron, por lo que fueron descartados. Finalmente, los autores optaron por el modelo unidimensional por parsimonia, coherencia teórica con la propuesta original de Beck, elevadas correlaciones interfactoriales y baja varianza explicada por un eventual segundo factor.

En cuanto a la invarianza factorial por sexo, Sánchez-Villena et al. (2022) probaron secuencialmente el modelo unidimensional retenido y hallaron evidencia de invarianza configuracional, de umbrales, de cargas e incluso de interceptos, con cambios mínimos en los índices de ajuste (ΔCFI entre 0.001 y 0.004; $\Delta RMSEA$ entre 0.001 y 0.004), lo que respalda la

equivalencia de la medición entre hombres y mujeres. Asimismo, las medias latentes fueron mayores en mujeres ($M = 6.42$) que en hombres ($M = 4.66$), aunque con un tamaño del efecto pequeño ($\xi = 0.173$).

Sánchez-Villena y Farfán (2019) reportaron que el BDI-II correlacionó positivamente con la Escala de Depresión de Reynolds ($r = 0.708$), lo que respalda la validez convergente. Además, mostró correlaciones inversas con la Escala de Felicidad de Alarcón ($r = -0.585$), interpretadas como evidencia de validez divergente, y correlaciones positivas con la Escala de Soledad de Jong Gierveld ($r = 0.392$), lo que aporta evidencia adicional de relación con variables externas teóricamente asociadas.

Sánchez-Villena et al. (2022) desarrollaron baremos percentilares del BDI-II en población general peruana ($n = 1665$; 18 a 54 años), considerando la no normalidad de la distribución (Shapiro-Wilk = 0.820, $p < 0.001$) y utilizando el coeficiente K^2 de Livingston para evaluar la confiabilidad de los puntos de corte. El percentil 25 correspondió a un puntaje directo de 1 ($K^2 = 0.880$), el percentil 75 a puntajes entre 10 y 11 ($K^2 = 0.821-0.834$), el percentil 90 a puntajes entre 16 y 19 ($K^2 = 0.905-0.933$), el percentil 95 a puntajes entre 20 y 24 ($K^2 = 0.940-0.961$) y el percentil 99 a puntajes entre 25 y 37 ($K^2 = 0.964-0.986$). Los autores señalan que puntajes por encima de 11 podrían indicar depresión clínicamente relevante y que la precisión de los puntos de corte aumenta en los extremos de la distribución.

Para el BDI-IIA, Rodríguez y Farfán (2015) mencionaron una muestra piloto de 150 participantes; sin embargo, no describieron los procedimientos aplicados, los análisis previos realizados, los resultados obtenidos ni las modificaciones derivadas de dicha fase piloto. Por ello, aunque se reporta la existencia de una prueba piloto, la información disponible no permite describir con precisión su alcance metodológico.

Carranza (2013) informó un coeficiente test-retest de 0.996 ($p < 0.01$), lo que sugiere una muy alta estabilidad temporal del BDI-IIA. No obstante, el estudio no detalla con claridad

el intervalo entre aplicaciones ni el procedimiento de reaplicación, por lo que esta evidencia debe interpretarse con cautela.

Así mismo, Carranza (2013) reportó una adecuada consistencia interna del BDI-IIA, con un alfa de Cronbach de 0.878 para la escala total de 21 ítems, los valores de alfa si se elimina el ítem oscilaron entre 0.869 y 0.878, lo que sugiere un aporte homogéneo de los reactivos al constructo medido. Como evidencias complementarias de homogeneidad interna, también se informaron una correlación entre mitades pares e impares de 0.803 ($p < 0.01$), un coeficiente de Guttman split-half de 0.799 y un coeficiente de Spearman-Brown de 0.888. Rodríguez y Farfán (2015) hallaron una alta consistencia interna, con un alfa total de 0.891; en sus dimensiones, reportaron un alfa de 0.837 para el componente cognitivo-afectivo y de 0.808 para el componente somático-motivacional.

Carranza (2013) reportó validez de contenido mediante criterio de jueces, con la participación de 10 evaluadores y cuantificación a través del coeficiente V de Aiken. En la escala global y en los 21 ítems se obtuvieron valores de $V = 1$ con $p = 0.001$ en criterios como claridad, ausencia de dificultad léxica, suficiencia del detalle y dominio del constructo, lo que respalda la adecuada representatividad del contenido.

Carranza (2013) reportó correlaciones ítem-test significativas para todos los ítems, generalmente superiores a 0.30, con un valor mínimo de 0.427 en el ítem 21. Además, el análisis factorial mostró adecuados indicadores de factorizabilidad ($KMO = 0.929$; Bartlett $\chi^2 = 10844.609$, $gl = 210$, $p = 0.000$). La extracción factorial identificó dos factores que explicaron el 38.099% de la varianza, correspondientes a componentes somático-motivacional y cognitivo-afectivo, con cargas superiores a 0.35, finalmente, aunque los autores discuten posteriormente una posible unidimensionalidad, sus propios resultados empíricos se orientan más claramente hacia una estructura bifactorial.

Rodríguez y Farfán (2015) evaluaron la homogeneidad de los ítems mediante correlaciones de Pearson, encontrando en general valores superiores a 0.50, salvo en los ítems 6 (0.42) y 21 (0.47), que fueron eliminados. Asimismo, justificaron la reubicación de los ítems 7 y 14 al factor cognitivo-afectivo por sus correlaciones con dicho componente. En el análisis factorial exploratorio eliminaron los ítems 6 y 21 y obtuvieron una solución de dos factores. Posteriormente, el análisis factorial confirmatorio corroboró un modelo de dos factores correlacionados, cognitivo-afectivo y somático-motivacional, con 19 ítems y una correlación interfactorial de 0.84. Los índices de ajuste fueron adecuados: $\chi^2 = 210.898$, $gl = 151$, $GFI = 0.931$, $RMSEA = 0.036$, $NFI = 0.875$, $CFI = 0.961$ y $RMR = 0.027$, lo que respalda el buen ajuste global del modelo.

Carranza (2013) reportó percentiles para la interpretación de los puntajes del BDI-IIA, entre ellos $P25 = 6$, $P50 = 11$, $P75 = 16$, $P90 = 22$ y $P95 = 26$. Con base en estos valores, propuso la clasificación de niveles de depresión en mínimo (0-6), leve (7-11), moderado (12-16) y severo (17-63).

Rodríguez y Farfán (2015) elaboraron baremos percentilares y categorías clínicas para la escala total de 19 ítems, con rango de 0 a 57, así como para sus dos dimensiones. En la escala total, reportaron $P25 = 10$, $P50 = 15$ y $P75 = 21$, proponiendo las categorías mínimo (0-10), leve (11-15), moderado (16-21) y severo (22-57). Para la dimensión cognitivo-afectiva establecieron las categorías mínimo (0-3), leve (4-7), moderado (8-11) y severo (12-30), mientras que en la dimensión somático-motivacional las categorías fueron mínimo (0-6), leve (7-9), moderado (10-11) y severo (12-27).

Para el BDI-PC, Lam et al. (2009) reportaron un alfa de Cronbach de 0.683, valor que indica una consistencia interna moderada y por debajo de lo esperado para una adecuada homogeneidad del instrumento. Además, al eliminar cualquiera de los ítems, el coeficiente alfa

osciló entre 0.62 y 0.68, sin superar el umbral de 0.70, lo que sugiere que la debilidad de la consistencia interna no depende de un reactivo problemático aislado.

Lam et al. (2009) aplicaron un análisis factorial exploratorio con rotación Varimax y encontraron que el BDI-PC no presentaba una estructura unidimensional, sino bidimensional, que explicó aproximadamente el 49% de la varianza total. El primer factor estuvo compuesto por los ítems pérdida de interés, disgusto por sí misma, pesimismo y tristeza, mientras que el segundo factor agrupó errores del pasado, autocrítica y deseos o pensamientos suicidas. No obstante, los coeficientes alfa obtenidos para estos factores fueron aún menores que el de la escala total, con valores de 0.60 y 0.48, por lo que la estructura hallada no fortaleció la validez interna del instrumento.

La validez concurrente del BDI-PC fue evaluada frente a la Escala de Depresión Postnatal de Edimburgo (EPDS), encontrándose una correlación de Pearson de 0.60 ($p < 0.001$), lo que representa una asociación moderada entre ambos instrumentos. A partir de este resultado, los autores concluyeron que dicha magnitud no era suficiente para considerar al BDI-PC como sustituto de la EPDS en el tamizaje de depresión durante la gestación.

Patient Health Questionnaire (PHQ)

El PHQ-9 (Patient Health Questionnaire-9) es un instrumento breve de autoinforme diseñado para evaluar la presencia y gravedad de síntomas depresivos durante las últimas dos semanas. Está compuesto por 9 ítems basados en los criterios diagnósticos del trastorno depresivo mayor, y cada uno se responde en una escala tipo Likert de 0 a 3, según la frecuencia de los síntomas. Su puntuación total oscila entre 0 y 27, donde los puntajes más altos indican mayor severidad depresiva; se utiliza ampliamente tanto para tamizaje como para seguimiento clínico (Kroenke y Spitzer, 2002; Kroenke et al., 2001; Spitzer et al., 1999).

El PHQ-8 es una versión derivada del PHQ-9 que conserva la misma estructura, formato de respuesta y periodo de referencia de las últimas dos semanas, pero excluye el ítem

relacionado con pensamientos de muerte o autolesión. Consta de 8 ítems, cada uno puntuado de 0 a 3, por lo que su rango total va de 0 a 24 (Villarreal-Zegarra et al., 2023).

El PHQ-2 es la versión más breve de esta familia de instrumentos y funciona como una prueba de cribado inicial para síntomas depresivos. Está formado por 2 ítems que exploran los síntomas centrales de la depresión: estado de ánimo deprimido y pérdida de interés o placer en las actividades, también con referencia a las últimas dos semanas. Cada ítem se puntúa de 0 a 3, con una puntuación total entre 0 y 6, y su utilidad principal radica en identificar de manera rápida a personas que requieren una evaluación más completa con instrumentos como el PHQ-9 (Levis et al., 2020; Villarreal-Zegarra et al., 2023).

Según la literatura revisada en Perú, se han validado las versiones PHQ-9 (versión general), PHQ-8 (versión reducida del PHQ-9 sin el último ítem sobre ideación suicida) y el PHQ-2 (reducción del PHQ-9 que incluye únicamente los 2 primeros ítems, relacionados con los síntomas nucleares de la depresión, que son anhedonia y estado de ánimo deprimido).

En el contexto peruano, la evidencia sobre el PHQ-2 es aun relativamente limitada, pero muestra aplicaciones en poblaciones adultas y hospitalarias. Baños-Chaparro et al. (2021) validaron el PHQ-2 en 283 adultos residentes en Lima, tomando como referencia la propuesta de Levis et al. (2020). Posteriormente, Villarreal-Zegarra et al. (2023) evaluaron esta versión en 1,347 pacientes adultos hospitalizados de un hospital peruano, a partir de la versión original del instrumento (Kroenke y Spitzer, 2002; Kroenke et al., 2001; Spitzer et al., 1999). Ambos estudios emplearon métodos muestrales no probabilísticos.

La evidencia identificada para el PHQ-8 en Perú es más escasa y procede, principalmente, del estudio de Villarreal-Zegarra et al. (2023). En esa investigación, el PHQ-8 fue analizado en una muestra de 1,347 pacientes adultos hospitalizados de Lima, utilizando como base la versión original, como ya se ha mencionado, por muestreo no probabilístico (Kroenke y Spitzer, 2002; Kroenke et al., 2001; Spitzer et al., 1999).

La evidencia peruana sobre el PHQ-9 es más amplia y heterogénea que la encontrada para el PHQ-2 y el PHQ-8, ya que incluye estudios en gestantes, población general, universitarios, internos de medicina, adolescentes, pacientes hospitalizados y población quechuahablante. Además, esta evidencia puede organizarse según la procedencia de la versión utilizada: estudios basados en la versión original, estudios que emplearon una versión original modificada para el recuerdo de síntomas durante el último año, investigaciones que usaron la versión difundida por Calderón et al. (2012), y trabajos que partieron de la adaptación peruana de Villarreal-Zegarra et al. (2019).

Dentro de los estudios que utilizaron la versión original del PHQ-9, Zhong et al. (2014) validaron el instrumento en 1,517 gestantes de Lima mediante muestreo no probabilístico. Más adelante, Villarreal-Zegarra et al. (2019) analizaron el PHQ-9 en una muestra nacional de 30449 participantes peruanos de ambos sexos a partir de la ENDES 2016 (la validación de este estudio sería retomada por otros autores como se verá más adelante), mientras que Álvarez et al. (2022) lo validaron en 343 internos de medicina humana de una universidad de Lima, estos 2 últimos estudios mencionados emplearon muestreo probabilístico. Adicionalmente, Smith et al. (2022) evaluaron el PHQ-9 en 5,399 mujeres gestantes atendidas en consulta ambulatoria en Lima, y Villarreal-Zegarra et al. (2023) volvieron a examinar esta versión en 1,347 pacientes adultos hospitalizados, ambos por selección no probabilística.

Varios estudios partieron de la validación peruana reportada por Villarreal-Zegarra et al. (2019), mencionada anteriormente, aunque Villarreal y colaboradores no realizaron ningún cambio significativo en la estructura, sintaxis o contenido del instrumento, por lo que a términos prácticos sigue siendo el instrumento original. Entre ellos, Renteria et al. (2022) validaron el PHQ-9 en 308 adolescentes de secundaria de Lima. Del mismo modo, Lingán-Huamán et al. (2023) evaluaron el instrumento en 409 estudiantes universitarios limeños, y Cjuno et al. (2023) realizaron la adaptación cultural y validación del PHQ-9 a tres variedades

del quechua en 970 adultos bilingües de Áncash, Ayacucho, Cusco y Puno. Los 3 estudios utilizaron muestreo no probabilístico.

Otra línea de evidencia corresponde a una versión modificada del PHQ-9 empleada en la ENDES para evaluar síntomas depresivos durante el último año. Villarreal-Zegarra et al. (2020), en un análisis secundario de las ENDES 2014–2018, estudiaron esta versión en 161,061 participantes peruanos de ambos sexos a nivel nacional mediante métodos muestrales probabilísticos. En esta adaptación, se pidió a los participantes recordar un evento problemático ocurrido en los últimos 12 meses y responder el cuestionario considerando las dos semanas cercanas a dicho episodio.

También se identificó el uso de una versión del PHQ-9 difundida por Calderón et al. (2012). Huarcaya-Victoria et al. (2020) validaron esta versión en 200 estudiantes universitarios de Lima, seleccionados mediante muestreo no probabilístico, aunque dicha publicación de Calderón y colaboradores no es una validación psicométrica propiamente dicha.

Para el PHQ-2, Baños-Chaparro et al. (2021) reportaron coeficientes de coherencia interna aceptables en una muestra de adultos peruanos residentes en Lima, con $\alpha = 0.75$ (IC 95%: 0.69–0.81) y $\omega = 0.76$ (IC 95%: 0.68–0.82). En la misma línea, Villarreal-Zegarra et al. (2023) informaron valores adecuados de consistencia interna en pacientes hospitalizados, con $\alpha = 0.83$ y $\omega = 0.80$.

En cuanto a la evidencia basada en relaciones con otras variables, Baños-Chaparro et al. (2021) evaluaron la asociación del PHQ-2 con ideación suicida y ansiedad mediante modelamiento de ecuaciones estructurales. El modelo presentó un buen ajuste global, con CFI = 0.97, TLI = 0.96 y RMSEA = 0.05 (IC 90%: 0.04–0.07). Además, las puntuaciones del PHQ-2 se correlacionaron de manera positiva y estadísticamente significativa con ideación suicida ($r = 0.61$, $p = 0.001$) y ansiedad ($r = 0.72$, $p = 0.001$), con tamaño del efecto fuerte en ambos casos.

Respecto a la sensibilidad y especificidad, Villarreal-Zegarra et al. (2023) encontraron que el punto de corte ≥ 2 fue el que mostró el mejor desempeño global en la muestra evaluada. Con ese punto de corte, el PHQ-2 alcanzó una sensibilidad de 84.7%, especificidad de 55.9%, valor predictivo positivo de 38.8%, valor predictivo negativo de 91.7% e índice de Youden de 40.6. En cambio, con el punto de corte ≥ 3 , la sensibilidad descendió a 55.4% y la especificidad aumentó a 82.9%, con valor predictivo positivo de 51.7%, valor predictivo negativo de 84.9% e índice de Youden de 38.3.

De manera complementaria, los datos suplementarios mostraron que para el punto de corte ≥ 2 el cociente de probabilidad positivo fue 1.92 y el cociente de probabilidad negativo fue 0.27. Para el punto de corte ≥ 3 , el cociente de probabilidad positivo aumentó a 3.24, mientras que el cociente de probabilidad negativo se incrementó a 0.54. En conjunto, estos resultados evidencian el intercambio habitual entre una mayor capacidad de detección con puntos de corte más bajos y una mayor especificidad con puntos de corte más altos.

Para el PHQ-8, Villarreal-Zegarra et al. (2023) reportaron evidencias favorables de fiabilidad y estructura interna en población hospitalaria peruana. En particular, el instrumento mostró adecuados coeficientes de coherencia interna ($\alpha = 0.88$; $\omega = 0.85$) y un modelo unidimensional con índices de ajuste $CFI = 0.977$, $TLI = 0.977$, $SRMR = 0.050$ y $RMSEA = 0.082$. Asimismo, en el análisis de rendimiento diagnóstico, el punto de corte ≥ 7 fue identificado como el óptimo frente al criterio clínico CIE-10, al presentar el mejor balance entre sensibilidad (79.9%) y especificidad (66.0%).

Respecto al PHQ-9, Cjuno et al. (2023) realizaron la traducción directa del original en inglés a tres variedades del quechua: Central, Chanca y Cuzco-Collao. Para cada variedad participaron dos traductores independientes, nativos quechuahablantes y con conocimientos avanzados de inglés.

Una vez finalizadas las traducciones directas, los dos traductores de cada variedad y dos investigadores quechuahablantes nativos se reunieron para discutir las diferencias encontradas entre versiones. A partir de esa revisión se resolvieron discrepancias y se unificó una versión consensuada para cada variedad del quechua.

Después de unificar las versiones en quechua, los autores realizaron la traducción inversa del quechua al inglés para las tres variantes del instrumento. Esta retrotraducción fue efectuada por dos traductores con inglés como lengua nativa y con conocimiento avanzado del quechua correspondiente, y luego fue revisada junto con la traducción inicial por traductores e investigadores para afinar detalles y aprobar la versión final.

La adaptación cultural incluyó una evaluación por expertos mediante método Delphi, usando una ficha específica de adaptación cultural con preguntas sobre términos difíciles, equivalencia conceptual y pertinencia cultural. Participaron dos psicólogos para quechua Central, dos para Chanca y cinco para Cuzco-Collao, todos quechuahablantes y con experiencia clínica en depresión, y las versiones fueron revisadas en dos rondas para Central y cuatro rondas para Chanca y Cuzco-Collao, evaluando relevancia, representatividad, claridad y equivalencia cultural en una escala de 0 a 3; cuando se alcanzó la máxima calificación, se realizó una reunión virtual final de consenso.

Si bien los autores no emplean de forma literal el término prueba piloto, sí desarrollaron una fase equivalente mediante grupos focales para cada variedad del quechua. En esta participaron cinco personas de Cuzco-Collao, cuatro de Chanca y cuatro de Central, todos bilingües y mayores de 18 años; primero respondieron el PHQ-9 en línea y luego comentaron la claridad y comprensión de los ítems, por lo que esta fase puede clasificarse como piloto o pretest.

Respecto a la confiabilidad por coherencia interna, Zhong et al. (2014) reportaron un alfa de Cronbach de 0.81 en mujeres embarazadas, lo que indica una consistencia interna

adecuada. De manera similar, Villarreal-Zegarra et al. (2019), en población general peruana, hallaron un $\alpha = 0.870$ y $\omega = 0.873$. En estudiantes y personal en formación médica también se observan resultados favorables. Álvarez et al. (2022) reportaron un alfa de Cronbach de 0.882 en internos de medicina humana. En el estudio de Smith et al. (2022) el PHQ-9 completo mostró un alfa de Cronbach de 0.80, considerado alto para el instrumento total. Además, los autores reportaron que la confiabilidad no mejoraba al eliminar ningún ítem, mientras que en el modelo de dos factores la subescala somática alcanzó un alfa de 0.75 y la subescala afectiva/estado de ánimo, un alfa de 0.67, lo que sugiere mayor solidez del puntaje global que de algunas subdimensiones por separado.

Villarreal-Zegarra et al. (2023) informaron una consistencia interna adecuada-alta para el PHQ-9, con $\alpha = 0.90$ y $\omega = 0.87$. En adolescentes peruanos, Renteria et al. (2022) reportaron un alfa ordinal de 0.848 para el factor somático, 0.9 para el factor cognitivo-afectivo y 0.931 para la puntuación global, lo que indica una consistencia interna alta de la escala total. De forma semejante, Lingán-Huamán et al. (2023) hallaron coeficientes elevados, con alfa de Cronbach = 0.91 y $\omega = 0.92$, ambos con intervalos de confianza altos. Tras corregir correlaciones entre residuos, el omega corregido se mantuvo en 0.89, mientras que el alfa corregido descendió a 0.70, lo que sugiere que la estimación de confiabilidad puede variar según el modelo especificado, aunque en general el instrumento mantiene una precisión aceptable.

En población quechuahablante, Cjuno et al. (2023) también reportaron adecuada confiabilidad por consistencia interna tanto en la muestra total como en cada variante lingüística evaluada. En la muestra global obtuvieron $\alpha = 0.895$ y $\omega = 0.861$; por variedades, los coeficientes fueron 0.865 y 0.834 para quechua Central, 0.877 y 0.833 para Chanca, y 0.915 y 0.881 para Cuzco-Collao. Estos hallazgos sugieren que la consistencia

interna del PHQ-9 se mantiene favorable incluso en contextos de adaptación lingüística y cultural.

Por su parte, Villarreal-Zegarra et al. (2020), al analizar la versión modificada del PHQ-9 en series anuales, reportaron coeficientes alfa entre 0.85 y 0.87 y coeficientes omega entre 0.85 y 0.87 entre 2014 y 2018, lo que respalda la estabilidad de una consistencia interna adecuada-alta a lo largo del tiempo. Por su parte, en la versión de Huarcaya-Victoria et al. (2020), en estudiantes de medicina, encontraron un alfa de Cronbach de 0.903, y señalaron que este valor no aumentaba al eliminar ningún ítem, lo que respalda la contribución pertinente de todos los reactivos al funcionamiento global de la escala.

Por parte de la evidencia basada en la estructura interna, se hallaron evidencias relacionadas a uno y dos factores. Respecto a la estructura bidimensional, Zhong et al. (2014) identificaron en gestantes una estructura de dos factores del PHQ-9 mediante AFE. Los autovalores fueron 3.64 y 1.13, y ambos factores explicaron en conjunto el 53% de la varianza total; los autores los interpretaron como somático y no somático. En el análisis factorial confirmatorio, esta solución mostró un ajuste adecuado, con SRMR = 0.040 y RMSEA = 0.069, mientras que las correlaciones ítem-total oscilaron entre 0.41 y 0.74.

Smith et al. (2022) también encontraron mejor desempeño para una solución de dos factores que para una de un solo factor. El modelo bidimensional presentó CFI robusto = 0.93, TLI robusto = 0.91, RMSEA robusto = 0.075 y SRMR = 0.042; además, superó significativamente al modelo unidimensional en la prueba de diferencia de Satorra-Bentler, $\chi^2 = 145$, $gl = 1$, $p < .001$. Los factores fueron somático y afectivo/estado de ánimo, con correlación interfactorial de 0.82, y las correlaciones ítem-total variaron entre 0.34 para ideación suicida y 0.61 para ánimo deprimido.

Renteria et al. (2022) reportaron asimismo una estructura de dos factores, somático y cognitivo-afectivo, con índices de ajuste muy favorables: SB- $\chi^2/gl = 0.492$, CFI = 0.995, TLI =

0.994, SRMR = 0.032 y RMSEA = 0.041. De manera similar, Álvarez et al. (2022) compararon un modelo unidimensional y uno bidimensional, y encontraron mejor ajuste para este último, con CFI = 0.954, TLI = 0.939, GFI = 0.948, SRMR = 0.042 y RMSEA = 0.078, frente al modelo unidimensional con CFI = 0.947, TLI = 0.921, GFI = 0.947, SRMR = 0.045 y RMSEA = 0.090. Además, el AIC (*Akaike Information Criterion*) fue menor en el modelo bidimensional, 6289.617 frente a 6361.104, y las correlaciones ítem-test fueron todas superiores a 0.50, con un rango entre 0.537 en el ítem 8 y 0.707 en el ítem 7.

Respecto a evidencias sobre estructura bifactorial y predominio del factor general, Huarcaya-Victoria et al. (2020) hallaron inicialmente condiciones adecuadas para factorizar la matriz, con KMO = 0.904 y prueba de Bartlett $\chi^2 = 917.791$, $p < 0.001$. En el análisis factorial exploratorio se extrajo un solo factor, con autovalor de 5.079 y 56.43% de varianza explicada; sin embargo, en el análisis factorial confirmatorio, el mejor ajuste correspondió al modelo bifactorial, con $\chi^2 (17) = 26.451$, $p = 0.067$, CFI = 0.991, GFI = 0.969 y RMSEA = 0.056, superando al modelo unidimensional y al bidimensional. En ese modelo, los ítems 3, 4 y 5 cargaron en el factor somático, mientras que los ítems 1, 2, 6, 7, 8 y 9 cargaron en el factor cognitivo-afectivo.

Así mismo, Lingán-Huamán et al. (2023) compararon modelos unidimensionales, de dos factores correlacionados y bifactor mediante AFC con estimador WLSMV, y encontraron en todos ellos índices de ajuste favorables, con CFI > 0.97, TLI > 0.96, RMSEA < 0.09, WRMR < 1.06 y cargas factoriales > 0.50. No obstante, la correlación interfactorial de 0.89 en el modelo de dos factores sugirió un solapamiento importante entre los componentes cognitivo-afectivo y somático, mientras que en el modelo bifactor el factor general explicó la mayor parte de la varianza ECV = 0.86, PUC = 0.50 y $\omega_h = 0.89$; los factores específicos aportaron poco ω_{hs} somático = 0.16 y ω_{hs} cognitivo-afectivo = 0.039. Tras modelar correlaciones residuales entre los ítems 6-9, 1-2 y 2-6, el modelo unidimensional alcanzó CFI = 0.98,

RMSEA = 0.07 y WRMR = 0.81, por lo que los autores concluyeron que el instrumento podía considerarse esencialmente unidimensional.

Sobre la evidencia a favor de unidimensionalidad en el instrumento, Villarreal-Zegarra et al. (2019) compararon modelos de una y dos dimensiones y concluyeron que la solución unidimensional era la más apropiada en población peruana. Aunque los modelos bidimensionales mostraron ajuste aceptable, las correlaciones entre factores fueron muy elevadas, entre 0.97 y 0.99, lo que indicó un solapamiento excesivo; además, el modelo unidimensional presentó CFI = 0.936, RMSEA = 0.089 y SRMR = 0.039, con correlaciones ítem-test entre 0.62 y 0.77. El estudio también aportó evidencia de invarianza configural, métrica, fuerte y estricta según sexo, edad, nivel educativo, nivel socioeconómico, estado civil y área de residencia, con $\Delta\text{CFI} < 0.01$ entre comparaciones sucesivas.

Villarreal-Zegarra et al. (2020) confirmaron igualmente un modelo unidimensional para la versión modificada del PHQ-9 entre 2014 y 2018. Los índices de ajuste por año fueron CFI = 0.918, 0.925, 0.912, 0.920 y 0.921; TLI = 0.900, 0.901, 0.898, 0.901 y 0.901; y RMSEA = 0.051, 0.053, 0.058, 0.059 y 0.050, con cargas factoriales entre 0.51 y 0.71. Villarreal-Zegarra et al. (2023) aportaron nuevamente evidencia favorable para una estructura unidimensional, con CFI = 0.978, TLI = 0.972, SRMR = 0.048 y RMSEA = 0.099. Aunque el RMSEA fue relativamente elevado, los autores consideraron que la estructura interna resultó adecuada en su muestra.

Cjuno et al. (2023) también hallaron buen ajuste para un modelo de un solo factor en 970 participantes, con CFI = 0.990, TLI = 0.987, SRMR = 0.048 y RMSEA = 0.071. Las cargas factoriales oscilaron entre 0.57 y 0.79, y además se observó invarianza según variante del quechua, sexo, residencia, edad, estado civil y nivel educativo; sin embargo, la versión en quechua central mostró indicadores menos favorables, con RMSEA = 0.112 y SRMR = 0.082, en comparación con las variantes Chanca y Cuzco-Collao.

Respecto a la evidencia basada en relaciones con otras variables, Zhong et al. (2014) reportaron que la correlación entre el puntaje total del PHQ-9 y el de la Edinburgh Postnatal Depression Scale (EPDS) fue de $\rho = 0.52$, lo que indica una relación positiva de magnitud moderada entre ambas escalas. Lingán-Huamán et al. (2023) encontraron evidencia de validez al observar que el PHQ-9 correlacionó de manera positiva y moderada con la Escala de Agotamiento Emocional Académico (EES) ($r=0.62$).

Con relación a la sensibilidad y especificidad, Smith et al. (2022) evaluaron el desempeño del PHQ-9 utilizando como estándar de referencia el *Composite International Diagnostic Interview* (CIDI) para trastorno depresivo mayor en los últimos 12 meses. El punto de corte óptimo fue ≥ 8 , ya que maximizó conjuntamente la sensibilidad 61% y la especificidad 62%. En ese punto de corte, el valor predictivo positivo fue 15%, el valor predictivo negativo fue 94% y el área bajo la curva AUC fue 0.67. Los autores señalaron que, aunque este punto de corte permite identificar a la mayoría de las mujeres que podrían requerir seguimiento, también genera una cantidad considerable de falsos positivos, por lo que el PHQ-9 resulta más útil como instrumento de tamizaje que como herramienta diagnóstica.

Villarreal-Zegarra et al. (2023) informaron que el PHQ-9 mostró un buen rendimiento diagnóstico frente al criterio clínico. El punto de corte óptimo fue ≥ 7 , al presentar el mayor índice de Youden entre los puntos evaluados. Con este punto de corte se obtuvo una sensibilidad de 76.0%, especificidad de 72.1%, valor predictivo positivo de 47.3%, valor predictivo negativo de 90.1% e índice de Youden de 48.1. Además, el análisis complementario mostró una razón de verosimilitud positiva de 2.72, una razón de verosimilitud negativa de 0.33 y un AUC de 0.805 con IC 95%: 0.779–0.831, lo que sugiere una capacidad discriminativa global buena. Cuando el punto de corte se elevó a ≥ 10 , la sensibilidad disminuyó a 54.2% y la especificidad aumentó a 87.4%, lo que evidencia el equilibrio esperado entre una mayor detección de casos y una menor proporción de falsos positivos.

V. DISCUSIÓN DE RESULTADOS

Los resultados de la presente revisión sistemática permiten trazar un panorama actualizado de la medición psicométrica de la depresión en población peruana. A partir de una búsqueda en seis bases de datos internacionales, se identificaron 35 artículos que reportaban un total de 39 validaciones correspondientes a 13 instrumentos distintos. Si bien este cuerpo de evidencia ha experimentado un crecimiento sostenido —con el año 2022 concentrando el mayor número de publicaciones (14,29%)—, sigue siendo reducido en comparación con la magnitud del problema de salud mental que la depresión representa en el país (MINSA, 2024; OMS, 2022, 2023). Este hallazgo es, no obstante, consistente con lo reportado en estudios que han documentado un aumento progresivo de publicaciones instrumentales sobre depresión en el contexto latinoamericano en los últimos años (Sarmiento, 2020).

Con relación al primer objetivo específico, la distribución de validaciones es marcadamente desigual. El PHQ (en sus versiones PHQ-9, PHQ-8 y PHQ-2) encabeza el campo con 13 validaciones y más de 204 000 participantes evaluados, lo que se explica por su inclusión en encuestas poblacionales de gran escala como la ENDES (Villarreal-Zegarra et al., 2020); este volumen muestral supera en casi 200 000 casos al segundo instrumento más estudiado. El BDI ocupa ese segundo lugar con 7 validaciones (versiones BDI, BDI-II, BDI-IIA y BDI-PC) en muestras universitarias, clínicas y de población general, con algo más de 5 000 participantes en total. Ambos instrumentos se han convertido, de facto, en los principales referentes de la medición de la depresión en el Perú, tanto en contextos epidemiológicos como clínicos (Cjuno et al., 2022).

En un segundo nivel se ubican la EPD/DPS y la EPDS con tres validaciones cada una, que articulan adaptaciones de instrumentos internacionales con desarrollos vinculados al contexto local. Un tercer grupo de importancia intermedia reúne a la ZSDS, el CDI-S, la EDAR y el IDER con dos validaciones cada uno, mientras que la CES-D, la EURO-D, la GDS-4, la

TDP y la DSRS aparecen con una sola validación, aportando evidencia puntual en nichos poblacionales específicos.

El predominio del PHQ refleja su adopción por grandes estudios nacionales y replica una tendencia internacional: a nivel global, el PHQ-9 se ha consolidado como el tamizaje breve de referencia en atención primaria y epidemiología, con múltiples metaanálisis que respaldan su sensibilidad, especificidad y versatilidad en diferentes países e idiomas (Carroll et al., 2020). Que este instrumento concentre la mayor parte de la muestra total sugiere que la medición de la depresión en el Perú está alineándose con estándares globales de salud pública, aunque al costo de relegar a un segundo plano otros instrumentos con potencial clínico relevante

Estos hallazgos adquieren mayor relevancia al ponerlos en relación con las recomendaciones oficiales y la práctica clínica, la cual identifica una alta carga de depresión en el país —con más de 280 000 casos atendidos en el sistema público tras la pandemia (MINSA, 2025)— lo que contrasta con el uso histórico de instrumentos escasamente validados a nivel local, como el SRQ-18 y la escala de Hamilton, que continúan siendo recomendados por el propio MINSA (2008) a pesar de que Cjuno et al. (2022) ya advirtieron sobre la insuficiencia de su evidencia psicométrica en población peruana.

La presente revisión muestra que existe hoy un conjunto de escalas con propiedades psicométricas documentadas —sobre todo el PHQ-9 y el BDI-II, pero también la EPDS, la EPD/EPD-4, la EURO-D, la TDP, el IDER y la ZSDS— que cuestionan esa inercia institucional. Puede inferirse que la brecha no reside tanto en la ausencia de instrumentos válidos, sino en su incorporación tardía a las normas y protocolos oficiales y en las dificultades logísticas y formativas para su implementación masiva en atención primaria y en contextos con recursos limitados (Lipskaya-Velikovsky y Krupa, 2023).

Finalmente, la síntesis global sugiere que la medición de la depresión en el Perú se encuentra en un punto de inflexión. Se ha pasado de un escenario caracterizado por el uso de

escalas no validadas y reportes psicométricos escasos —descrito por Cjuno et al. (2021, 2022) y Navarro-Loli y Moscoso (2017)— a otro donde existe un repertorio de instrumentos con evidencias más sólidas y sistematizadas. Con todo, persisten vacíos importantes en poblaciones indígenas, rurales, amazónicas y en grupos etarios extremos (niñez temprana y vejez avanzada), así como en el uso de métodos psicométricos más sofisticados que garanticen la comparabilidad entre grupos.

En relación con el segundo objetivo específico, la adaptación cultural garantiza que un instrumento desarrollado en un contexto lingüístico y sociocultural determinado pueda utilizarse de manera válida en una población diferente. Según las directrices de la ITC (2013), este proceso exige abordar de forma explícita aspectos culturales, conceptuales, lingüísticos y métricos a través de etapas diferenciadas: traducción directa, síntesis, retrotraducción, revisión por comité de expertos y prueba piloto (Muñiz et al., 2013; Luján-Tangarife y Cardona-Arias, 2015). Esta exigencia es especialmente relevante en Latinoamérica, donde la mayoría de los instrumentos psicológicos fueron desarrollados en culturas anglosajonas y la fidelidad semántica y conceptual no puede asumirse a priori (Chahín-Pinzón, 2014). En esta línea, los lineamientos de adaptación transcultural subrayan que no basta con traducir literalmente los ítems, sino que es necesario garantizar equivalencias semánticas, idiomáticas, experienciales y conceptuales entre la versión original y la adaptada (Guillemin et al., 1993).

Bajo este estándar, el PHQ-9 en sus versiones quechuas constituyen el caso más sólido de adaptación documentada: Cjuno et al. (2023) realizaron traducción directa a tres variedades (Central, Chanca y Cuzco-Collao) por traductores nativos bilingües, síntesis de versiones, retrotraducción y revisión comparada, un procedimiento Delphi con psicólogos quechuahablantes para evaluar relevancia, claridad y equivalencia cultural, y una fase piloto mediante grupos focales, cumpliendo todas las etapas de la rúbrica. El CDI-S también se aproxima a este estándar: Ventura-León et al. (2020) siguieron las directrices de la ITC,

realizaron entrevistas semiestructuradas a 10 niños como prueba piloto y modificaron términos para mejorar la equivalencia semántica en población peruana.

En comparación, muchos estudios internacionales sobre el PHQ-9 en español se han centrado en su estructura factorial, confiabilidad y precisión diagnóstica en distintos países, pero no siempre describen con el mismo detalle los pasos de adaptación cultural, pese a mostrar buen desempeño psicométrico en migrantes y diversos grupos latinoamericanos (Martínez et al., 2023). Esto sugiere que las versiones quechuas del PHQ-9 no solo cumplen con los estándares internacionales, sino que representan un ejemplo avanzado de adaptación en lenguas originarias dentro de la región, alineado con la recomendación de asegurar invariancia factorial y ausencia de sesgos de medición en diferentes subgrupos (Cjuno et al., 2021).

La EPD/DPS (Vega-Dienstmaier et al., 2011) y la TDP (Elejabo-Pacheco et al., 2025) representan un caso distinto: contruidos directamente en el Perú en castellano y desde la práctica clínica, su pertinencia cultural está incorporada desde el diseño. La EPD dio lugar a versiones abreviadas EPD-4 y EPD-6 sobre la misma base (Vega-Dienstmaier, 2018), y la TDP operacionaliza criterios DSM-5 en universitarios peruanos; aunque esta pertinencia se limita al castellano urbano y a los grupos estudiados, sin extensiones documentadas a lenguas originarias o contextos rurales.

Esta estrategia coincide con una de las alternativas consideradas en la literatura internacional, donde algunos autores defienden la construcción de instrumentos propios cuando se considera que los constructos se expresan de modo diferencial en determinados contextos socioculturales, aunque esto complique la comparabilidad con grandes estudios multicéntricos (Hilaire, 2023). En este sentido, el panorama peruano refleja una combinación de adaptación de medidas globales (como el PHQ-9) y desarrollo de escalas locales (EPD/DPS, TDP), lo que puede verse como fortaleza en términos de sensibilidad cultural, pero también como un desafío para la armonización de resultados a nivel internacional (Black et al., 2024).

En una posición intermedia se ubican el IDER, la DSRS y la GDS-4, con ajustes parcialmente documentados. Sotelo et al. (2012) reportan una adecuación lingüística del IDER mediante jueces expertos sin detallar el protocolo ni indicadores de acuerdo; Vivar et al. (2005) modificaron la DSRS añadiendo y retirando ítems con juicio de expertos, pero sin describir etapas ni criterios; y De la Torre et al. (2006) tradujeron la GDS-4 del inglés con revisión de geriatras, pero sin detallar equivalencia semántica ni pilotaje.

En los tres casos, la documentación no permite verificar la sistematicidad del proceso, algo que también ha sido observado en revisiones de la aplicación de los estándares en otros países, donde numerosos estudios mencionan traducción o juicio de expertos, pero omiten describir con precisión cómo se ejecutaron las etapas propuestas (Dos Santos et al., 2025). Esa falta de transparencia metodológica dificulta evaluar si la equivalencia cultural es suficiente como para sostener comparaciones válidas entre contextos, y muestra que el problema identificado en esta investigación no es exclusivo del Perú, sino parte de una tendencia más amplia en la región y en otros entornos de habla portuguesa y española (Ramada-Rodilla et al., 2013).

La mayoría de los instrumentos—BDI (Rodríguez y Farfán, 2015; Sánchez-Pachas, 2019), EPDS (Vega-Dienstmaier et al., 2002), ZSDS (Del Águila et al., 2021; Ruiz-Grosso et al., 2012), CES-D (Ruiz-Grosso et al., 2012), EURO-D (Guerra et al., 2015) y EDAR (Chávez y Ecurra, 2002; Ramos-Vera et al., 2023)—emplean versiones en español ya estandarizadas sin documentar adaptación cultural formal, siendo la más destacada la BDI-II/IIA por 2 pruebas piloto. Esta brecha es especialmente problemática en el Perú, donde no puede asumirse que una versión válida para Lima sea equivalente para poblaciones de la sierra, la selva o grupos indígenas (Bada et al., 2025). El estudio de Cjuno et al. (2023) lo ilustra con claridad: adaptar el PHQ-9 al quechua requirió múltiples rondas de revisión y ajustes dialectales, evidenciando que las versiones en español son insuficientes para garantizar equivalencia conceptual y

cultural en poblaciones indígenas o bilingües (Cjuno et al., 2021), en consonancia con estudios internacionales que han documentado variaciones en puntos de corte y funcionamiento de los ítems del PHQ-9 entre distintos países y grupos migrantes (Martínez et al., 2023).

En este marco, los hallazgos de la presente tesis coinciden con la preocupación global por la calidad de las adaptaciones de instrumentos de depresión y sitúan al contexto peruano como un ejemplo donde se hace visible la urgencia de alinear la práctica local con los estándares internacionales de adaptación y uso de pruebas psicológicas (AERA, APA y NCME, 2014; ITC, 2013; Muñiz et al., 2013; Luján-Tangarife y Cardona-Arias, 2015).

Con relación al tercer objetivo específico, la evidencia de los instrumentos revisados muestra una confiabilidad sólida en términos de coherencia interna, pero muy limitada respecto a estabilidad temporal, formularios alternativos y acuerdo entre evaluadores. Aunque la tesis considera cuatro fuentes de confiabilidad (formularios alternativos, test-retest, coherencia interna e índices entre evaluadores), en la práctica casi todos los estudios reportan solo alfa de Cronbach y, en menor medida, omega de McDonald, reflejando una preferencia por la homogeneidad de los ítems frente a otras formas de precisión.

Esta asimetría implica que un test puede ser internamente consistente y, aun así, producir puntuaciones inestables entre aplicaciones o dependientes del evaluador, lo que es problemático para el seguimiento clínico y la toma de decisiones diagnósticas (McCrae et al., 2011). Por ello, se plantea que las investigaciones futuras deben ir más allá del reporte rutinario de alfa y omega e incorporar diseños que estimen directamente la estabilidad de las puntuaciones en el tiempo, sobre todo en los instrumentos más utilizados en la política de salud mental peruana (Dunn et al., 2014).

El PHQ acumula la mayor evidencia de confiabilidad del conjunto. Sus tres versiones —PHQ-2, PHQ-8 y PHQ-9— muestran coherencia interna alta y consistente en todos los grupos evaluados: gestantes, población general, universitarios, internos de medicina, pacientes

hospitalizados y hablantes de quechua (Baños-Chaparro et al., 2021; Zhong et al., 2014; Villarreal-Zegarra et al., 2019, 2020, 2023; Álvarez et al., 2022; Smith et al., 2022; Lingán-Huamán et al., 2023; Cjuno et al., 2023; Huarcaya-Victoria et al., 2020; Renteria et al., 2022). Lo que resulta especialmente destacable es que esta coherencia se mantiene estable incluso cuando el instrumento se adapta a tres variedades del quechua y cuando se analiza en series anuales de cinco años, lo que sugiere que los ítems del PHQ funcionan de manera homogénea con independencia del contexto de aplicación. Sin embargo, la ausencia total de test-retest en todos los estudios peruanos del PHQ impide saber si sus puntuaciones son también estables en el tiempo, una limitación importante para su uso en evaluaciones longitudinales o de seguimiento clínico (Carroll et al., 2020).

El BDI, en sus versiones BDI original, BDI-II y BDI-IIA, también presenta coherencia interna alta y homogénea en contextos clínicos, universitarios y escolares (Vega-Dienstmaier et al., 2014; Sánchez-Pachas, 2019; Sánchez-Villena y Farfán, 2019; Sánchez-Villena et al., 2022; Carranza, 2013; Rodríguez y Farfán, 2015). El BDI-IIA es la única versión de la familia que aporta evidencia de estabilidad temporal, con un coeficiente test-retest muy alto reportado por Carranza (2013); no obstante, la falta de detalle sobre el intervalo entre aplicaciones y el procedimiento seguido obliga a interpretar este dato con reserva. La excepción dentro de la familia es el BDI-PC, cuya coherencia interna en gestantes peruanas se sitúa por debajo del umbral generalmente aceptado (Lam et al., 2009), lo que cuestiona su adecuación como herramienta de tamizaje en este grupo y advierte sobre el riesgo de asumir que una versión abreviada conserva automáticamente la fiabilidad de la escala completa (Kleka y Paluchowski, 2017).

Entre las escalas para adultos, CES-D (Ruiz-Grosso et al., 2012) y ZSDS (Del Águila et al., 2021) destacan por su coherencia interna elevada en las muestras peruanas disponibles, lo que respalda su uso en investigación y tamizaje. La EURO-D, en cambio, muestra una

coherencia interna aceptable en la muestra urbana pero más baja en la rural (Guerra et al., 2015), diferencia que podría reflejar que los ítems de la escala no capturan el constructo de depresión de manera igualmente homogénea en contextos con distintos niveles educativos y exposición lingüística.

La EPD/DPS y su versión abreviada EPD-4 presentan una trayectoria interesante: la escala completa tiene buena consistencia global, pero con dimensiones menos homogéneas, mientras que la EPD-4, aunque más breve, alcanza niveles adecuados en población general, siendo el ítem de anhedonia el de mayor aporte a la fiabilidad total (Vega-Dienstmaier et al., 2011; Vega-Dienstmaier, 2018; Lee-Díaz et al., 2024). La TDP destaca como el instrumento de nueva construcción con mayor coherencia interna del conjunto, lo que es esperable dado que fue diseñado específicamente para la muestra con que se valida (Elejabo-Pacheco et al., 2025). La EPDS muestra niveles aceptables en gestantes y púrperras (Vega-Dienstmaier et al., 2002; Lam et al., 2010; Zhong et al., 2014), y la GDS-4 es el único instrumento sin alfa reportado, con un índice de concordancia moderado con la GDS-15 (De la Torre et al., 2006), lo que indica que ambas formas coinciden solo parcialmente en sus clasificaciones.

En población infantojuvenil, la EDAR es el instrumento con evidencia de confiabilidad más completa, al combinar una buena coherencia interna con coeficientes de test-retest que varían según el grado escolar (Chávez y Ecurra, 2002; Ramos-Vera et al., 2023), lo que puede sugerir que la estabilidad de las puntuaciones depende en parte del nivel de desarrollo de los adolescentes evaluados. El CDI-S presenta coherencia interna adecuada y, llamativamente, sin diferencias entre grupos anónimo e identificado, lo que indica que el anonimato no afecta la precisión de la medida (Ventura-León et al., 2020; Merino-Soto et al., 2022).

La DSRS combina estabilidad temporal alta en el grupo con trastorno depresivo con una coherencia interna que se deteriora considerablemente en el grupo con trastorno de ajuste, lo que revela que la homogeneidad de sus ítems depende del diagnóstico del grupo y que su

fiabilidad no puede generalizarse entre subgrupos clínicos (Vivar et al., 2005). El IDER muestra mayor coherencia interna en adultos que en adolescentes, lo que puede interpretarse como una señal de que el constructo de depresión estado–rasgo resulta más difícil de capturar de manera homogénea en etapas más tempranas del desarrollo (Soto et al., 2012; Sotelo et al., 2012).

Respecto al cuarto objetivo específico, la validez de los instrumentos revisados se sustenta principalmente en evidencia basada en la estructura interna y en relaciones con otras variables, mientras que la validez de contenido, los procesos de respuesta y las consecuencias del uso de las pruebas han sido explorados de manera más limitada (e.g., Ruiz-Grosso et al., 2012; Guerra et al., 2015; Vega-Dienstmaier et al., 2011; Villarreal-Zegarra et al., 2019). En términos generales, los estudios se han centrado en demostrar que las escalas funcionan psicométricamente mediante análisis factoriales y correlaciones con otras medidas, más que en documentar con detalle la construcción de los ítems, la forma en que responden los evaluados o los efectos prácticos de su aplicación. Esto implica que, aunque muchos instrumentos cuentan con respaldo suficiente para su uso en investigación, la evidencia disponible no siempre es suficiente para justificar su aplicación clínica sin precauciones adicionales (Kelly et al., 2005).

Dentro de este panorama, el conjunto de versiones del PHQ (PHQ-2, PHQ-8 y PHQ-9) reúne la evidencia más amplia y diversa de validez. En PHQ-9, los estudios con gestantes, población general, internos de medicina, adolescentes, pacientes hospitalizados y población quechuahablante muestran estructuras principalmente unidimensionales o bidimensionales, con buenos índices de ajuste y, en varios casos, evidencia de invarianza entre subgrupos, lo que significa que el instrumento permite comparaciones válidas entre grupos distintos sin que la forma de medir cambie según el grupo (Zhong et al., 2014; Villarreal-Zegarra et al., 2019, 2020, 2023; Álvarez et al., 2022; Renteria et al., 2022; Lingán-Huamán et al., 2023; Cjuno et al., 2023). Además, el PHQ-2 y el PHQ-8 han mostrado modelos estructurales aceptables,

asociaciones fuertes con ideación suicida, ansiedad y otros criterios clínicos, así como puntos de corte óptimos frente a diagnósticos CIE-10, lo que refuerza su validez convergente y respalda con mayor solidez su uso en tamizaje rápido en contextos asistenciales (Baños-Chaparro et al., 2021; Villarreal-Zegarra et al., 2023).

En cuanto al BDI y sus variantes, la evidencia respalda de manera consistente la representación del constructo depresivo, aunque con cierta heterogeneidad según la versión y la muestra. El BDI-II presenta correlaciones ítem-total adecuadas, estructuras factoriales esencialmente unidimensionales o de dos factores altamente correlacionados, buenos índices de ajuste y evidencias de invarianza por sexo, lo que indica que la escala capta un factor general de severidad depresiva que, en algunas muestras, puede diferenciar entre componentes cognitivo-afectivos y somático-motivacionales (Sánchez-Pachas, 2019; Sánchez-Villena y Farfán, 2019; Sánchez-Villena et al., 2022).

Esta ambigüedad estructural no invalida el instrumento, pero sí sugiere que la interpretación por dimensiones debe hacerse con cautela, priorizando el puntaje global. Por su parte, el BDI-IIA muestra alta validez de contenido según jueces, estructura bifactorial coherente y relaciones sólidas con otras medidas (Carranza, 2013; Rodríguez y Farfán, 2015), mientras que el BDI-PC presenta una estructura interna menos clara y una validez concurrente moderada frente a la EPDS, lo que limita su uso como reemplazo en el tamizaje perinatal (Lam et al., 2009).

Entre las escalas aplicadas en adultos y contextos clínicos, la EPDS, la ZSDS, la CES-D y la EPD/DPS aportan evidencias relevantes, aunque con distinto nivel de detalle y sistematicidad. En la EPDS, los estudios con madres posparto y gestantes encuentran soluciones de dos y tres factores con buenos índices de ajuste, así como correlaciones moderadas con criterios diagnósticos y con el PHQ-9, lo que respalda una estructura multidimensional y sugiere que interpretar la EPDS solo como puntaje total puede ocultar

diferencias clínicamente relevantes entre ansiedad y estado de ánimo deprimido (Vega-Dienstmaier et al., 2002; Lam et al., 2010; Zhong et al., 2014).

La ZSDS (Ruiz-Grosso et al., 2012; Del Águila et al., 2021) y la CES-D (Ruiz-Grosso et al., 2012) muestran estructuras internas aceptables y correlaciones fuertes con otras escalas depresivas y con indicadores de severidad clínica, mientras que la EPD/DPS y sus versiones abreviadas combinan soluciones factoriales claras, correlaciones altas con escalas relacionadas y consistencia entre forma larga y corta, lo que sugiere que la reducción de ítems no compromete la validez si el modelo subyacente está bien definido (Vega-Dienstmaier et al., 2011; Vega-Dienstmaier. 2018; Lee-Díaz et al., 2024).

En niños y adolescentes, el CDI-S y la EDAR concentran las mejores evidencias. El CDI-S se apoya en una estructura interna consistente, ausencia de funcionamiento diferencial de ítems entre aplicaciones anónimas e identificadas y correlaciones fuertes con la versión extensa, lo que indica que la escala breve retiene la esencia del constructo original y que su formato de aplicación no sesga los resultados (Ventura-León et al., 2020; Merino-Soto et al., 2022). En la EDAR, los análisis factoriales muestran estructuras de múltiples factores con buenos índices de ajuste, correlaciones ítem-total satisfactorias e invarianza por género, lo que respalda su uso para comparar grupos de adolescentes sin que el sexo distorsione la medición (Chávez y Escurra, 2002; Ramos-Vera et al., 2023). En contraste, instrumentos como la DSRS (Vivar et al., 2005) y la GDS-4 (De la Torre et al., 2006) cuentan con estudios en población peruana, pero sin análisis específicos de validez en el sentido amplio considerado aquí, lo que impide hacer afirmaciones sobre la calidad de su medición más allá del contexto clínico originalmente descrito.

Otros instrumentos como la TDP y el IDER ilustran tanto los avances como las brechas en la evaluación de validez. La TDP combina alta validez de contenido, estructura unifactorial sólida, invarianza por sexo y fuertes asociaciones con depresión, ansiedad y satisfacción con

la vida, lo que la convierte en uno de los ejemplos más completos y recientes de validación integral en población universitaria peruana (Elejabo-Pacheco et al., 2025). El IDER, en cambio, aporta evidencia de contenido y de estructura interna aceptable en adolescentes y adultos, pero con una base empírica más reducida y dependiente de muestras no probabilísticas, lo cual sugiere que aún requiere mayor acumulación de estudios para que sus resultados puedan generalizarse con confianza (Soto et al., 2012; Sotelo et al., 2012).

En conjunto, estos hallazgos refuerzan la idea de que la investigación psicométrica peruana en depresión ha avanzado notablemente en estructura interna y relaciones externas, pero que consolidar la validez de estos instrumentos implica también atender con mayor sistematicidad la representatividad del contenido, los procesos cognitivos de respuesta y las consecuencias concretas de su uso en la práctica clínica y educativa.

Respecto al quinto objetivo específico, la evidencia disponible muestra que solo una parte de los instrumentos revisados cuenta con análisis sistemáticos de puntos de corte, sensibilidad, especificidad y normas, y que estos procedimientos se concentran sobre todo en escalas ampliamente utilizadas como la DSRS, GDS-4, CES-D, EURO-D, EPDS, ZSDS, EPD/DPS, BDI, BDI-II, BDI-IIA y PHQ. Esto implica que, mientras algunos instrumentos ofrecen parámetros claros para el tamizaje y la interpretación clínica de los puntajes, otros se estudian en el país casi exclusivamente como medidas continuas, sin umbrales bien establecidos ni baremos adaptados a la población peruana.

En niños y adolescentes, la DSRS y la EDAR son los pocos casos con información explícita sobre desempeño diagnóstico o normas. La DSRS presenta puntos de corte en 13 y 14 con sensibilidades de 98,8% y especificidades de 98,7%, lo que sugiere una capacidad casi óptima para discriminar entre casos con y sin depresión en la muestra estudiada; sin embargo, al tratarse de una muestra no probabilística y específica, estos resultados deben interpretarse como evidencia preliminar más que como umbrales generalizables (Vivar et al., 2005). En la

EDAR, Chávez y Ecurra (2002) elaboran normas percentilares para estudiantes secundarios limeños, pero sin categorías de severidad clínica, lo que limita su uso directo para la toma de decisiones diagnósticas y exige al profesional una interpretación más cualitativa o la construcción posterior de puntos de corte clínicamente significativos.

En adultos mayores, la GDS-4 y la EURO-D ilustran dos enfoques distintos de rendimiento diagnóstico. La primera muestra una sensibilidad baja (49%) y una especificidad muy alta (97%), lo que la convierte en una herramienta que confirma más que detecta, reduciendo su utilidad como test de tamizaje inicial en geriatría si se utiliza de forma aislada (De la Torre et al., 2006); por el contrario, la segunda presenta áreas bajo la curva muy elevadas (AUC 0.94–0.96) y un punto de corte 4/5 con sensibilidades entre 92% y 100% y especificidades entre 84,5% y 87%, configurándola como un instrumento con excelente capacidad discriminativa para episodio depresivo CIE-10 en adultos mayores urbanos y rurales (Guerra et al., 2015), esta diferencia sugiere que, en el contexto peruano, la EURO-D ofrece una base más sólida para el tamizaje poblacional en vejez que la GDS-4.

En población adulta general y clínica, varias escalas cuentan con análisis ROC y puntos de corte específicos. La CES-D muestra AUC de 0,83 y 0,98, con puntos de corte óptimos ≥ 29 y ≥ 24 según el grupo de comparación, alcanzando sensibilidades de 77,1–91,4% y especificidades de 79,4–96,7%, lo que evidencia un muy buen desempeño para discriminar entre episodios depresivos mayores y grupos sin trastorno o con otros diagnósticos (Ruiz-Grosso et al., 2012). La ZSDS, por su parte, presenta un AUC de 0,96 frente a controles sin psicopatología, con un punto de corte ≥ 45 que combina sensibilidad (91,4%) y especificidad (91,8%) altas; además, en pacientes con VIH se han propuesto rangos percentilares para interpretar niveles leve, moderado y severo, ofreciendo una guía práctica para graduar la sintomatología en contextos clínicos específicos (Ruiz-Grosso et al., 2012; Del Águila et al., 2021).

La EPD/DPS y sus versiones abreviadas (EPD-4 y EPD-6) destacan por una sistemática optimización de puntos de corte basada en auROC y en el índice de Youden. En la versión de 20 ítems, se reportan AUC superiores a 0,83 frente a criterios diagnósticos para trastorno depresivo mayor y diagnósticos psiquiátricos, con puntos de corte 26/27 y 25/26 que logran sensibilidades alrededor de 80% y especificidades entre 72% y 80% (Vega-Dienstmaier et al., 2011). Las versiones cortas alcanzan incluso auROC cercanas o superiores a 0,90, con cortes como >6 para la EPD-4 o >8 para la EPD-6, combinando buenas sensibilidades ($\approx 80\text{--}83\%$) y especificidades ($\approx 74\text{--}87\%$); en población general, la EPD-4 logra un AUC de 0,92 frente a la CES-D, con corte óptimo en 4 puntos (Vega-Dienstmaier, 2018; Lee-Díaz et al., 2024); esto indica que es posible disponer de instrumentos brevísimos con un rendimiento diagnóstico comparable o incluso superior al de escalas más largas, lo que tiene implicancias directas para su uso en tamizajes breves en servicios de salud y estudios poblacionales.

El BDI y sus variantes constituyen uno de los conjuntos mejor documentados en términos de normas y puntos de corte. Para el BDI original en pacientes hospitalizados de medicina general, se propone un punto de corte 18,5/19,5 frente a la SCID para depresión mayor, con sensibilidad de 87,5% y especificidad de 98,21%; además, se identifican rangos de puntuación prácticamente excluyentes (por debajo de 7,5 sin casos de depresión mayor y por encima de 24,5 con todos los casos confirmados), lo que le confiere un alto valor para descartar o confirmar la presencia de episodios depresivos en ese contexto (Vega-Dienstmaier et al., 2014).

El BDI-II cuenta con baremos percentilares en población general, con percentiles extremos (90, 95 y 99) asociados a puntajes moderados y altos de sintomatología, y con la sugerencia de considerar puntajes por encima de 11 como indicativos de depresión clínicamente relevante, lo que facilita una interpretación graduada en contextos no necesariamente clínicos (Sánchez-Villena et al., 2022). En el BDI-IIA, los percentiles permiten

agrupar los puntajes en categorías mínimo, leve, moderado y severo, y Rodríguez y Farfán (2015) refinan esta propuesta en adolescentes con baremos y rangos diferenciados por dimensión (cognitivo-afectiva y somático-motivacional), lo que abre la puerta a interpretaciones más finas de los perfiles sintomáticos en este grupo etario (Carranza, 2013).

En el ámbito de la salud perinatal, la EPDS también dispone de información diagnóstica específica. En madres en posparto temprano, un punto de corte de 13,5 ($EPDS \geq 14$) frente a la SCID para episodio depresivo mayor ofrece sensibilidad de 84,21% y especificidad de 79,47%, configurándola como una buena herramienta de tamizaje para depresión posnatal en Lima (Vega-Dienstmaier et al., 2002). Dado que este estudio evalúa múltiples umbrales, sus resultados permiten ajustar el punto de corte según el propósito (mayor sensibilidad si se busca no perder casos, o mayor especificidad si se pretende limitar falsos positivos), algo especialmente relevante en servicios con recursos diagnósticos limitados.

El PHQ es el instrumento con la evidencia más extensa sobre rendimiento diagnóstico y, en menor medida, sobre normas. Para el PHQ-2, se ha propuesto un punto de corte ≥ 2 con buena sensibilidad (84,7%) pero especificidad moderada (55,9%), y un punto de corte ≥ 3 que invierte esta relación, aumentando la especificidad (82,9%) a costa de reducir la sensibilidad (55,4%), lo que ilustra claramente el intercambio entre detectar más casos y generar menos falsos positivos (Villarreal-Zegarra et al., 2023). En el PHQ-8, un corte ≥ 7 ofrece un balance aceptable (sensibilidad 79,9%, especificidad 66,0%), mientras que, en el PHQ-9, los estudios peruanos muestran puntos de corte más bien bajos para tamizaje (≥ 7 u ≥ 8) con sensibilidades alrededor de 60–76% y especificidades en torno a 62–72%, AUC cercanas a 0,80 y valores predictivos negativos altos, lo que confirma su utilidad como herramienta de cribado, pero no como prueba diagnóstica definitiva (Smith et al., 2022; Villarreal-Zegarra et al., 2023), lo que puede indicar que, en términos prácticos, el PHQ debería utilizarse como primera etapa en un

proceso de evaluación en dos pasos, donde los casos positivos sean posteriormente evaluados mediante entrevista clínica o instrumentos diagnósticos más específicos.

En contraste con estos ejemplos, varios instrumentos relevantes, como la TDP, el IDER, el CDI-S o el propio PHQ-9 en varias de sus aplicaciones, carecen de baremos específicos o análisis de sensibilidad/especificidad completos en población peruana, y se emplean principalmente a partir de criterios tomados de otros contextos o como medidas continuas (Elejabo-Pacheco et al., 2025; Soto et al., 2012; Sotelo et al., 2012; Ventura-León et al., 2020; Merino-Soto et al., 2022). Esta situación limita la precisión de la interpretación clínica y la comparabilidad entre estudios, especialmente cuando se requiere estimar prevalencias o tomar decisiones diagnósticas. Así, aunque las normas y los puntos de corte disponibles muestran que es posible construir parámetros diagnósticos sólidos en el país, también evidencian la necesidad de extender este tipo de análisis a instrumentos y poblaciones menos estudiadas, idealmente con muestreos probabilísticos y criterios clínicos bien definidos (Carroll et al., 2020).

Tomando en cuenta sus implicancias, los resultados de esta revisión sugieren que existen instrumentos con respaldo psicométrico suficiente para ser incorporados como herramientas de rutina en la práctica clínica peruana. Los instrumentos que reúnen las mejores capacidades psicométricas globales son el PHQ en sus distintas versiones (especialmente el PHQ-9), el BDI/BDI-II y sus variantes, la EPDS, la familia EPD (EPD-20 y EPD-4) y la ZSDS, con un papel más específico de la EURO-D en adultos mayores y, en segundo plano, el CDI-S y la EDAR para población infanto-juvenil. En conjunto, combinan buena confiabilidad, evidencia de validez estructural y convergente, y en varios casos puntos de corte o baremos desarrollados localmente; al mismo tiempo, revelan vacíos claros, especialmente en adaptación cultural sistemática, normas clínicas y estudios diagnósticos en ciertos subgrupos.

En términos de uso clínico versus uso en investigación, el PHQ y el BDI-II ilustran dos perfiles complementarios. El PHQ-9 muestra estructura interna consistente y notable

adaptabilidad intercultural, con validaciones en población general, pacientes hospitalizados, universitarios, gestantes, adolescentes y personas quechuahablantes, lo que lo hace especialmente atractivo para encuestas y monitoreo poblacional (Villarreal-Zegarra et al., 2019). Sin embargo, sus puntos de corte privilegian la sensibilidad sobre la especificidad, con valores predictivos positivos modestos y un número importante de falsos positivos, por lo que resulta más seguro como herramienta de tamizaje e investigación que como base única para el diagnóstico clínico, y se recomienda complementarlo con entrevista clínica o instrumentos más finos (Smith et al., 2022; Villarreal-Zegarra et al., 2023). El BDI-II, en cambio, es particularmente adecuado para la práctica clínica especializada, gracias a su alta consistencia interna, baremos percentilares para población general peruana y categorías de severidad bien definidas, aunque carece de una descripción detallada y reciente de adaptación cultural, la solidez de sus normas locales y su sensibilidad para graduar la severidad lo hacen especialmente útil para el seguimiento clínico y la monitorización de cambios terapéuticos (Carranza, 2013; Rodríguez y Farfán, 2015; Sánchez-Villena et al., 2022).

La familia EPD (EPD-20 y EPD-4) ocupa un lugar estratégico como opción breve y eficiente en contextos clínicos y comunitarios. La EPD-4 alcanza áreas bajo la curva superiores a 0,90 frente a criterios diagnósticos y frente a la CES-D, con puntos de corte definidos, demostrando que es posible un tamizaje confiable con muy pocos ítems; no obstante, la evidencia se concentra en pocas muestras limeñas y no probabilísticas, por lo que aún se requiere replicación en otras regiones y niveles de atención (Lee-Díaz et al., 2024; Vega-Dienstmaier, 2018). La ZSDS, por su parte, presenta un AUC cercano a 0,96 en adultos ambulatorios y normas clínicas en pacientes con VIH, pero su documentación de adaptación cultural es limitada y las validaciones se han realizado en un número restringido de contextos, lo que la ubica por detrás del PHQ y del BDI-II en términos de versatilidad (Del Águila et al., 2021; Rivera et al., 2007; Ruiz-Grosso et al., 2012).

En población infanto-juvenil, los instrumentos con mejor perfil psicométrico son el CDI-S y la EDAR. El CDI-S se destaca por un proceso explícito de adaptación cultural, buena consistencia interna, estructura factorial clara y ausencia de sesgos entre aplicaciones anónimas e identificadas; sin embargo, no dispone de normas clínicas ni puntos de corte validados frente a un criterio diagnóstico externo, lo que limita su uso a medida continua para investigación o tamizaje exploratorio (Merino-Soto et al., 2022; Ventura-León et al., 2020). La EDAR, por su parte, muestra estructuras factoriales robustas, invarianza por género y normas percentilares escolares, aunque dichas normas carecen de categorías clínicas directamente aplicables y la evidencia proviene de muestras urbanas no probabilísticas (Chávez y Escurra, 2002; Ramos-Vera et al., 2023). La DSRS presenta resultados muy favorables en una sola muestra pero sin análisis de validez complementarios (Vivar et al., 2005), mientras que el IDER ofrece evidencia aceptable, aunque con escasa información normativa, por lo que su uso debería acompañarse de evaluación clínica adicional (Sotelo et al., 2012; Soto et al., 2012).

En áreas específicas, la EPDS es la herramienta prioritaria para el tamizaje perinatal, respaldada por un estudio peruano con criterio clínico externo, aunque los procedimientos de traducción y adaptación no han sido suficientemente documentados y faltan estudios de invarianza entre gestantes y púerperas de distintas regiones (Vega-Dienstmaier et al., 2002; Lam et al., 2010). La EURO-D, validada en más de 1900 adultos mayores peruanos urbanos y rurales, se posiciona como el mejor candidato para el tamizaje en ese grupo etario, con excelente capacidad discriminativa, aunque deja un vacío para adultos de mediana edad con sintomatología subclínica (Guerra et al., 2015).

Desde el punto de vista teórico y metodológico, el conjunto de hallazgos refuerza la idea de la depresión como un constructo predominantemente continuo, captado por factores generales de severidad a los que se superponen, en algunos instrumentos, dimensiones más específicas (cognitivo-afectiva, somático-motivacional, ansiedad, anhedonia) (Sarmiento,

2020). Asimismo, el hecho de que instrumentos como el PHQ hayan demostrado una buena adaptación a distintos idiomas y regiones del país, mientras que el BDI-II mantiene pendientes aspectos de adaptación cultural, pone de relieve la necesidad de que las futuras validaciones peruanas integren, de manera más sistemática, procedimientos de traducción, equivalencia semántica y análisis de funcionamiento diferencial de ítems (Cardona et al., 2023).

El perfil psicométrico de estos instrumentos también orienta la metodología de evaluación en el país, por ejemplo, para la investigación epidemiológica y los estudios poblacionales, ya que los datos sugieren priorizar escalas como el PHQ-9 (y sus formas abreviadas) y la EPD-4, que combinan economía de tiempo, adaptabilidad intercultural y buenos indicadores de validez y rendimiento diagnóstico. En la práctica clínica, en cambio, parece más razonable apoyarse en instrumentos con normas locales sólidas y sensibilidad al cambio, como el BDI-II y la EPDS en sus respectivos ámbitos, complementados por la ZSDS o la EPD cuando se busque una rápida clasificación de casos. En conjunto, estas decisiones instrumentales no solo tienen implicancias en la calidad del diagnóstico individual, sino también en la forma en que se conceptualiza y se investiga la depresión en el contexto peruano, consolidando un marco en el que la elección de la prueba responde tanto a criterios psicométricos como a las necesidades concretas de cada nivel de atención.

Desde la perspectiva de salud pública, los hallazgos de la tesis refuerzan la urgencia de alinear las políticas de tamizaje con la evidencia psicométrica disponible. Mientras el MINSA ha recomendado históricamente el uso del SRQ-18 y la HDRS en el primer nivel de atención, ambos instrumentos carecen de validaciones actuales y extensivas en población peruana, o incluso de estudios formales en el caso de la HDRS (Cjuno et al., 2022). En contraste, la familia PHQ acumula más de 200 000 participantes evaluados en diversas investigaciones nacionales –incluyendo la ENDES 2014–2018– y ha mostrado propiedades psicométricas adecuadas para el monitoreo poblacional y el tamizaje clínico, lo que ha llevado a proponer su incorporación

explícita en políticas de salud mental y guías de práctica clínica (Villarreal-Zegarra et al., 2020). El MINSA (2023) aprobó una Guía de Práctica Clínica para el tamizaje, diagnóstico y tratamiento de la depresión en mayores de 18 años desde el primer nivel de atención, con el objetivo de estandarizar procedimientos y mejorar la detección temprana, lo que abre una ventana de oportunidad para sustituir o complementar instrumentos poco estudiados por escalas con evidencia local robusta.

Tanto el BDI-II como el PHQ-9 incorporan los criterios nucleares del DSM para depresión, pero el BDI-II puede destacarse teóricamente por cómo aborda de manera explícita las anomalías diagnósticas descritas en tu sección (cambios paradójicos en sueño, apetito y actividad), en este sentido, la segunda edición del BDI se reformuló precisamente para incluir, en un mismo ítem, tanto el aumento como la disminución del apetito y del sueño, y para añadir ítems específicos de agitación psicomotriz, respondiendo a críticas al BDI original que señalaban la ausencia de estos polos atípicos y de la agitación como criterio, lo que podía dejar fuera a depresiones atípicas (Brenlla y Rodríguez, 2006). En ese sentido, su estructura de ítems graduados (0–3) no solo cuantifica severidad, sino que incorpora explícitamente la bidireccionalidad de varios síntomas, alineándose bien con la idea de que la depresión puede expresarse tanto por defecto como por exceso y con la necesidad de instrumentos más complejos y sensibles a la fenomenología clínica detallada.

El PHQ-9 también cubre formalmente los polos opuestos de algunos síntomas (por ejemplo, “dormir demasiado o tener dificultades para dormir” y “poco apetito o comer en exceso”), siguiendo los nueve criterios DSM para episodio depresivo mayor y priorizando, al igual que el BDI-II, una alta sensibilidad para fines de tamizaje en atención primaria (Kroenke y Spitzer, 2002; Kroenke et al., 2001; Spitzer et al., 1999). Sin embargo, su diseño es más conciso y orientado a la frecuencia de síntomas que a matices cualitativos (cognitivos y afectivos), lo que lo hace excelente como herramienta breve y gratuita para detección y

monitoreo poblacional, pero menos rico para explorar contenidos depresivos complejos (culpa elaborada, autoevaluación negativa detallada, variaciones atípicas sutiles) en comparación con el BDI-II. Estudios comparativos muestran que ambos instrumentos tienen fiabilidad y sensibilidad similares y pueden usarse de manera intercambiable como medidas continuas (Kung et al., 2013), pero el BDI-II tiende a clasificar a los mismos pacientes en categorías de mayor severidad y ofrece una gradación más fina de la sintomatología, mientras que el PHQ-9 aporta ventajas prácticas (brevedad, foco DSM, acceso libre) que lo hacen especialmente apropiado como primer filtro en contextos clínicos y de salud pública.

Como parte de las limitaciones de los estudios encontrados, la evidencia psicométrica de las 39 validaciones incluidas está afectada por sesgos que limitan la generalización de los hallazgos: la gran mayoría emplea muestreos no probabilísticos concentrados en Lima y contextos urbanos, dejando subrepresentadas las regiones rurales, amazónicas y las poblaciones indígenas, y predominan adultos universitarios, gestantes y pacientes hospitalarios, con escasa atención a niños, adultos mayores comunitarios y personas con enfermedades crónicas, por lo que no es posible garantizar que las propiedades psicométricas observadas se mantengan en otros grupos poblacionales.

Así mismo, la documentación sobre adaptación cultural es con frecuencia incompleta o escasa: pocos trabajos describen en detalle procedimientos de traducción inversa, comités de expertos, pilotajes cualitativos o análisis de funcionamiento diferencial de ítems, pese a que directrices internacionales como las de la ITC (2013) y los estándares de AERA, APA y NCME (2018) recomiendan estos pasos de forma explícita, además, varios estudios no evaluaron la invarianza de medición entre subgrupos relevantes (sexo, edad, nivel educativo), lo que limita la interpretación comparativa de los resultados.

Otra limitación recurrente es el uso exclusivo del alfa de Cronbach como estimador de confiabilidad, cuando las estructuras multidimensionales o bifactoriales de algunos

instrumentos hacen preferible el coeficiente omega, que no depende del supuesto de tau-equivalencia. Finalmente, la mayoría de los estudios evaluaron las propiedades psicométricas en un único momento sin incluir análisis de sensibilidad al cambio ni de mínima diferencia clínicamente importante, lo que impide establecer su utilidad para el monitoreo de intervenciones terapéuticas.

En sus alcances, esta revisión sistemática presenta un conjunto de fortalezas que la distinguen de trabajos previos sobre el tema. En primer lugar, constituye uno de los esfuerzos más comprehensivos realizados hasta la fecha para sintetizar la evidencia psicométrica sobre instrumentos de evaluación de la depresión en el Perú, reuniendo 39 validaciones que cubren 13 escalas en poblaciones diversas —adultos, universitarios, gestantes, adolescentes, niños y adultos mayores—, lo que ofrece una panorámica actualizada y comparativa que no estaba disponible en revisiones anteriores (Cjuno et al., 2022).

Así mismo, en vez de evaluar los instrumentos de forma aislada, este estudio organizó sistemáticamente la evidencia en torno a cuatro grandes dimensiones psicométricas —adaptación cultural, confiabilidad, validez y normas/rendimiento diagnóstico—, lo que permite identificar con precisión qué aspectos han sido bien estudiados y cuáles permanecen pendientes para cada instrumento (AERA, APA y NCME, 2018; ITC, 2013). Esta estructura multinivel no solo facilita la síntesis, sino que también hace más transparente y replicable el proceso de evaluación de la calidad de la evidencia.

En este plano, el uso de una rúbrica explícita basada en las recomendaciones de la AERA, APA y NCME (2018) e ITC (2013) representa un avance importante para la evaluación sistemática de instrumentos en salud mental. Al organizar la revisión en torno a cuatro grandes criterios, el estudio no solo sintetiza resultados, sino que también ofrece un modelo operativo para valorar futuras validaciones de escalas de depresión u otras variables psicológicas en el Perú, siguiendo la línea de los avances más recientes de la investigación sistemática moderna

(Miranda-Castillo et al., 2026; Page et al., 2021). Esta aproximación permite identificar de manera más clara las áreas descuidadas en muchos trabajos —por ejemplo, ausencia de estudios de procesos de respuesta, de invarianza entre grupos o de consecuencias del uso de las pruebas—, en línea con la llamada de los estándares y las normas, a considerar múltiples fuentes de evidencia de validez más allá de la estructura y la consistencia internas.

El estudio tiene relevancia directa para la práctica clínica y la salud pública peruana, ya que sus resultados pueden orientar decisiones instrumentales en distintos niveles de atención, desde el primer nivel hasta los servicios especializados de salud mental, y pueden servir de insumo para actualizar guías de práctica clínica y protocolos de tamizaje institucionales, incluyendo aquellos que se implementen en el marco de la reciente Guía de Práctica Clínica del MINSA para la depresión en mayores de 18 años (MINSA, 2023).

Finalmente, es también pertinente reconocer las limitaciones propias de esta revisión sistemática, ya que se incluyeron únicamente artículos publicados hasta enero de 2026 en bases de datos específicas y en español o inglés, lo que puede haber dejado fuera tesis, literatura gris, reportes técnicos institucionales o trabajos recientes aún en prensa, con el consiguiente riesgo de sesgo de publicación. Así mismo, la exclusión de entrevistas clínicas estructuradas y otros procedimientos de evaluación no autoinformados significa que los resultados deben entenderse como un mapa actualizado de las escalas psicométricas de autoinforme disponibles, más que como un inventario completo de todas las herramientas diagnósticas utilizadas en el país.

Asimismo, la revisión se limitó a evaluar las propiedades psicométricas formalmente reportadas en los estudios, sin incluir análisis primarios propios de los datos, lo que implica que la calidad de las conclusiones depende directamente de la calidad metodológica de los estudios incluidos. Por último, la heterogeneidad en las poblaciones, versiones del instrumento, criterios de referencia y procedimientos estadísticos empleados entre estudios dificultó la

realización de síntesis cuantitativas precisas, obligando a una narrativa comparativa que, si bien es comprehensiva, podría beneficiarse de un metaanálisis formal en el futuro.

VI. CONCLUSIONES

- Se identificaron trece instrumentos validados en el Perú, siendo el PHQ-9 el de mayor evidencia acumulada, con estudios en gestantes, población general, universitarios, adolescentes, pacientes hospitalizados y población quechuahablante, seguido del BDI-II con validaciones en contextos clínicos, universitarios y comunitarios, no obstante, la mayoría de las muestras fueron no probabilísticas y se concentraron en Lima y algunas ciudades urbanas, con escasa representación de regiones rurales, amazónicas y poblaciones indígenas, lo que limita la generalización de los hallazgos al conjunto de la población peruana.
- La adaptación cultural fue el criterio menos desarrollado en la evidencia revisada: la mayoría de los instrumentos fueron adoptados a partir de versiones ya existentes en español sin describir en detalle procedimientos de traducción, retrotraducción ni pilotajes cualitativos formales, la excepción más notable fue el PHQ-9 en quechua, con traducciones directas e inversas por traductores nativos, evaluación por expertos mediante método Delphi y grupos focales, constituyendo el proceso de adaptación más riguroso de toda la revisión; en contraste, la mayoría de instrumentos fueron aplicados sin documentación detallada, completa o rigurosa de adaptación cultural.
- La mayoría de los instrumentos presentó evidencia de validez basada en la estructura interna y en las relaciones con otras variables, aunque con desigual profundidad metodológica. El PHQ-9 fue el instrumento con mayor riqueza en este plano, con estudios de invarianza factorial entre sexos, grupos etarios y variedades lingüísticas; el BDI-II mostró ajuste adecuado para modelos uni y bidimensionales con invarianza por sexo, en cambio, la evidencia de validez divergente, de procesos de respuesta y de consecuencias del uso de las pruebas fue escasa o ausente en prácticamente todos los instrumentos revisados.

- La consistencia interna fue el indicador de confiabilidad más reportado y en general resultó adecuada o alta en la mayoría de los instrumentos y contextos evaluados, con el PHQ-9 y el BDI-II como los instrumentos con evidencia más amplia y replicada en distintas poblaciones. Sin embargo, la confiabilidad test-retest fue reportada de forma escasa, y la práctica predominante de emplear únicamente el alfa de Cronbach —en lugar de estimadores más apropiados para estructuras complejas como el omega— fue una limitación metodológica recurrente en los estudios analizados.
- El desarrollo de normas fue el criterio con mayor variabilidad entre instrumentos. Algunos, como el BDI-II, contaron con baremos o puntos de corte bien respaldados y aplicables en contextos específicos, mientras que instrumentos como la EPD y la ZSDS mostraron rendimientos diagnósticos sólidos en sus respectivas muestras. En el extremo opuesto, varios instrumentos (en particular los orientados a población infanto-juvenil) carecieron de normas locales suficientes para orientar su uso clínico o investigativo, lo que constituye una de las brechas más urgentes de la psicometría de la depresión en el Perú.

VII. RECOMENDACIONES

- Realizar metaanálisis que sintetizen cuantitativamente los índices de sensibilidad, especificidad, confiabilidad y validez reportados para cada instrumento, orientando con mayor solidez las decisiones sobre qué escalas adoptar en la práctica clínica y en la investigación nacional.
- Ampliar la cobertura geográfica y cultural de las validaciones, adaptando instrumentos en poblaciones rurales, amazónicas y pueblos originarios mediante procedimientos rigurosos que incorporen entrevistas cognitivas, retrotraducción sistemática y análisis de funcionamiento diferencial de ítems, con el fin de reducir las desigualdades regionales en la detección de la depresión.
- Incorporar métodos psicométricos más avanzados, incluyendo análisis factorial confirmatorio, modelos bifactor cuando la teoría lo justifique, estimaciones de confiabilidad basadas en omega, y análisis de sensibilidad al cambio y de diferencia mínima clínicamente importante para los instrumentos orientados al seguimiento terapéutico.
- Ampliar las fuentes de evidencia de validez exploradas, en particular los procesos de respuesta y las consecuencias del uso de las pruebas, complementando los análisis cuantitativos con entrevistas cognitivas y grupos focales que examinen cómo distintos grupos comprenden y responden a los ítems en contextos multiculturales.
- Priorizar la generación de normas locales en grupos con cobertura limitada, como niños y adolescentes en distintas regiones, adultos mayores comunitarios, población indígena y personas con enfermedades crónicas, donde la ausencia de baremos específicos puede producir errores de clasificación con consecuencias directas sobre el diagnóstico y la derivación.

- Evaluar los efectos reales de implementar escalas de depresión en servicios de salud, incluyendo su impacto en tasas de detección temprana, derivación oportuna y acceso a tratamiento, dado que este tipo de evidencia sobre las consecuencias del uso de las pruebas es prácticamente inexistente en el contexto peruano.
- Establecer un protocolo mínimo de validación psicométrica en el Perú, coordinado entre investigadores, instituciones formadoras y el ministerio de salud, que incluya criterios unificados de reporte, criterios diagnósticos estandarizados como referencia externa y registro prospectivo de estudios de validación.

VIII. REFERENCIAS

- Aguado, A., y del Álamo, M. G. (2020). Comorbilidad y síntomas digestivos asociados a depresión en pacientes mayores de 60 años. *Medicina de Familia. SEMERGEN*, 46(1), 27-32. <https://doi.org/10.1016/j.semerg.2019.03.003>
- Agudelo, A., Carretero-Dios, H., Blanco, A., Pitti, C., Spielberger, C. D. y Buela-Casal, G. (2005). Evaluación del componente afectivo de la depresión: *Análisis factorial del ST/DEP revisado. Salud Mental*, 28(3), 32-41. https://www.scielo.org.mx/scielo.php?pid=S0185-33252005000300032&script=sci_arttext
- Alarcón, R. (2020). Panorama histórico de la psicometría en el Perú. *Revista Peruana de Historia de la Psicología*, 6, 7-26. <https://historiapsiperu.org.pe/wp-content/uploads/2021/11/Rev-Peru-hist-psicol-2020-6.pdf#page=7>
- Alonso, M. Y. G., Llamazares, M. D. C. E., Martín, M. Á. M., Gómez, M. B. M., y Val, E. M. (2019). Comorbilidad de los trastornos alimentarios con ansiedad y depresión en estudiantes universitarios: revisión sistemática. *Revista argentina de clínica psicológica*, 28(4), 375-384. <https://psycnet.apa.org/record/2020-35647-006>
- Álvarez, E. A. C., Virú-Flores, H., Alburqueque-Melgarejo, J., Roque-Quezada, J. C., Valencia, J. G., Matta, G. A. G., ... y Quiroz, D. A. L. (2022). Validación del cuestionario sobre la salud del paciente-9 (PHQ-9) en internos de medicina humana de una universidad de referencia del Perú durante la pandemia COVID-19. *Revista de la Facultad de Medicina Humana*, 22(3). <https://doi.org/10.25176/rfmh.v22i3.3179>
- American Educational Research Association, la American Psychological Association y National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.

- American Psychiatric Association. (1994). *Manual diagnóstico y estadístico de los trastornos mentales, cuarta edición*. Masson S. A.
- American Psychiatric Association. (2014). *Manual diagnóstico y estadístico de los trastornos mentales, quinta edición*. Panamericana.
- American Psychological Association. (2023). *Dictionary of Psychology* (versión electrónica).
<https://dictionary.apa.org/depression>
- Aragonès, E., Piñol, J. L., y Labad, A. (2009). Comorbilidad de la depresión mayor con otros trastornos mentales comunes en pacientes de atención primaria. *Atención primaria, 41*(10), 545-551. <https://doi.org/10.1016/j.aprim.2008.11.011>
- Ato, M., López-García, J. J., y Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología/Annals of Psychology, 29*(3), 1038-1059. <http://dx.doi.org/10.6018/analesps.29.3.178511>
- Bada, W., Villacorta, Y. R., Novoa-Pallares, O., Torres, A., Gonzalez, C., Wisum, U., ... y López, N. (2025). Effectiveness of the PHQ-9 for detecting depressive symptoms in Peruvian Amazonian indigenous people. *General hospital psychiatry, 97*, 1-2. [10.1016/j.genhosppsy.2025.08.014](https://doi.org/10.1016/j.genhosppsy.2025.08.014)
- Bademci, V. (2022). Correcting Fallacies about Validity as the Most Fundamental Concept in Educational and Psychological Measurement. *International e-Journal of Educational Studies, 6*(12), 148-154. <https://doi.org/10.31458/iejes.1140672>
- Baños-Chaparro, J., Ynquillay-Lima, P., Delgado, F. L., y Guillen, F. G. F. (2021). Evidencias psicométricas del Patient Health Questionnaire-2 para la evaluación de síntomas depresivos en adultos peruanos. *Revista de Ciencias Médicas de Pinar del Río, 25*(5). http://scielo.sld.cu/scielo.php?pid=S1561-31942021000500021&script=sci_arttext&tlng=en

- Beck, A. T., Steer, R. A. y Gregory, K. B. (2006) BDI-II. *Inventario de Depresión de Beck* (segunda edición). Paidós.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry*, 4(6), 561-571.
10.1001/archpsyc.1961.01710120031004
- Beck, A., Steer, R., & Brown, G. (1996). *BDI-II. Inventario de Depresión de Beck. Segunda Edición*. Manual Paidós.
- Beck, A., Steer, R., & Brown, G. (2000). *BDI-Fast Screen for Medical Patients Manual*. The Psychological Corporation.
- Becker-Haimes, E. M., Tabachnick, A. R., Last, B. S., Stewart, R. E., Hasan-Granier, A., & Beidas, R. S. (2020). Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child & Adolescent Psychology*, 49(1), 1-17.
<https://doi.org/10.1080/15374416.2019.1689824>
- Bibi, A., Lin, M., Zhang, X. C., & Margraf, J. (2020). Psychometric properties and measurement invariance of Depression, Anxiety and Stress Scales (DASS-21) across cultures. *International Journal of Psychology*, 55(6), 916-925.
<https://doi.org/10.1002/ijop.12671>
- Birleson, P. (1981). The validity of depressive disorder in childhood and the development of a self-rating scale: a research report. *Journal of child Psychology and Psychiatry*, 22(1), 73-88. <https://doi.org/10.1111/j.1469-7610.1981.tb00533.x>
- Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of psychosomatic research*, 52(2), 69-77. [https://doi.org/10.1016/S0022-3999\(01\)00296-3](https://doi.org/10.1016/S0022-3999(01)00296-3)
- Black, M. H., Buitelaar, J., Charman, T., Ecker, C., Gallagher, L., Hens, K., ... & Bölte, S. (2024). Conceptual framework for data harmonisation in mental health using the

- International Classification of Functioning, Disability and Health: an example with the R2D2-MH consortium. *BMJ mental health*, 27(1).
<https://mentalhealth.bmj.com/content/27/1/e301283>
- Brenlla, M. E., y Rodríguez, C. M. (2006). Adaptación argentina del Inventario de Depresión de Beck (BDI-II). En A. T. Beck, R. A. Steer y G. K. Brown (Ed.) *BDI-II. Inventario de Depresión de Beck*. (2^a. ed). Manual Paidós.
- Caballo, V. E., Salazar, I. C. & Carroble, J. A. (2014). *Manual de psicopatología y trastornos psicológicos*. Pirámide.
- Cabañero-Martínez, M. J., Richart-Martínez, M., Muñoz-Mendoza, C. L., & Reig-Ferrer, A. (2007). Revisión estructurada de las escalas de depresión en personas mayores. *International Journal of Clinical and Health Psychology*, 7(3), 823-846.
<https://www.redalyc.org/pdf/337/33770317.pdf>
- Cala, M. L. P., & Agudelo, F. J. P. (2018). Depresión en universitarios. Diversas conceptualizaciones y necesidad de intervenir desde una perspectiva compleja. *Archivos de Medicina (Manizales)*, 18(2), 339-351.
<https://doi.org/10.30554/archmed.18.2.2747.2018>
- Calderón, M., Gálvez-Buccollini, J. A., Cueva, G., Ordoñez, C., Bromley, C., & Fiestas, F. (2012). Validación de la versión peruana del PHQ-9 para el diagnóstico de depresión. *Revista peruana de medicina experimental y salud pública*, 29, 578-579.
<https://www.scielosp.org/pdf/rpmesp/2012.v29n4/578-579/es>
- Cardona, L. G., Yang, M., Seon, Q., Karia, M., Velupillai, G., Noel, V., & Linnaranta, O. (2023). The methods of improving cultural sensitivity of depression scales for use among global indigenous populations: a systematic scoping review. *Cambridge Prisms: Global Mental Health*, 10, e83. <https://doi.org/10.1017/gmh.2023.75>

- Carranza, E. F. (2013). Propiedades psicométricas del Inventario de Depresión de Beck universitarios de Lima. *Revista De Psicología (Trujillo)*, 15(2), 170-182.
<https://revistas.ucv.edu.pe/index.php/revpsi/article/view/468>
- Carroll, H. A., Hook, K., Perez, O. F. R., Denckla, C., Vince, C. C., Ghebrehiwet, S., ... & Henderson, D. C. (2020). Establishing reliability and validity for mental health screening instruments in resource-constrained settings: systematic review of the PHQ-9 and key recommendations. *Psychiatry research*, 291, 113236.
<https://doi.org/10.1016/j.psychres.2020.113236>
- Castro-Costa, E., Dewey, M., Stewart, R., Banerjee, S., Huppert, F., Mendonca-Lima, C., ... & Prince, M. (2007). Prevalence of depressive symptoms and syndromes in later life in ten European countries: the SHARE study. *The British journal of psychiatry*, 191(5), 393-401. <https://doi.org/10.1192/bjp.bp.107.036772>
- Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud, la Organización Panamericana de la Salud y la Organización Mundial de la Salud. (23 de marzo de 2023). *DeCS/MeSH – Descriptores en Ciencias de la Salud/Medical Subject Headings*.
<https://decs.bvsalud.org/es/>
- Chahín-Pinzón, N. (2014). Aspectos a tener en cuenta cuando se realiza una adaptación de test entre diferentes culturas. *Psychologia. Avances de la disciplina*, 8(2), 109-112.
http://www.scielo.org.co/scielo.php?pid=S1900-23862014000200009&script=sci_arttext
- Chávez, N. U., & Ecurra, M. (2002). Adaptación psicométrica de la Escala de Depresión para Adolescentes de Reynolds (EDAR) en estudiantes de secundaria de Lima metropolitana. *Persona: Revista de la Facultad de Psicología*, (5), 83-130.
<https://hdl.handle.net/20.500.12724/2104>

- Cjuno, J., Figueroa-Quiñones, J., Marca-Dueñas, G. G., & Carranza-Esteban, R. F. (2021). Producción científica sobre depresión en poblaciones quechua hablantes: una revisión narrativa. *Terapia psicológica*, 39(2), 163-174. <http://dx.doi.org/10.4067/s0718-48082021000200163>
- Cjuno, J., Julca-Guerrero, F., Oruro-Zuloaga, Y., Cruz-Mendoza, F., Auccatoma-Quispe, A., Gómez Hurtado, H., ... & Bazo-Alvarez, J. C. (2023). Adaptación cultural al Quechua y análisis psicométrico del Patient Health Questionnaire (PHQ-9) en población peruana. *Revista Peruana de Medicina Experimental y Salud Pública*, 40, 267-277. <https://www.scielo.org/article/rpmesp/2023.v40n3/267-277/>
- Cjuno, J., Moya, A., Calderón-Pérez, E., Quispe-Ilizarbe, C., Mayon, L., & Livia, J. (2022). Producción científica sobre validación y adaptación de instrumentos de tamizaje de la depresión en población peruana. *Revista Peruana de Medicina Experimental y Salud Pública*, 39, 357-361. <https://doi.org/10.17843/rpmesp.2022.393.11197>
- Copeland, J. R., Prince, M., Wilson, K. C. M., Dewey, M. E., Payne, J., & Gurland, B. (2002). The geriatric mental state examination in the 21st century. *International journal of geriatric psychiatry*, 17(8), 729-732. <https://doi.org/10.1002/gps.667>
- Coto-Lesmes, R., Fernández-Rodríguez, C., & González-Fernández, S. (2020). Activación Conductual en formato grupal para ansiedad y depresión. Una revisión sistemática. *Terapia psicológica*, 38(1), 63-84. <http://dx.doi.org/10.4067/S0718-48082020000100063>
- Cox, J. L., Chapman, G., Murray, D., & Jones, P. (1996). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *Journal of affective disorders*, 39(3), 185-189. [https://doi.org/10.1016/0165-0327\(96\)00008-0](https://doi.org/10.1016/0165-0327(96)00008-0)
- De la Torre, J., Maeki, R. S., Pinedo, L. V., Malpartida, H. K., Falconí, L. H., Zevallos, J. C., & Cano, M. G. (2006). Validación de la versión reducida de la escala de depresión

- geriátrica en el consultorio externo de geriatría del Hospital Nacional Cayetano Heredia. *Acta médica peruana*, 23(3), 144-147.
<https://dialnet.unirioja.es/servlet/articulo?codigo=7759269>
- De Los Reyes & Langer D. A. (2018). Assessment and the Journal of Clinical Child and Adolescent Psychology's evidence base updates series: Evaluating the tools for gathering evidence. *Journal of Clinical Child & Adolescent Psychology*, 47(3), 357–365. <https://doi.org/10.1080/15374416.2018.1458314>
- Deeks, J. J., Bossuyt, P. M., Leeflang, M. M., Takwoingi, Y. (2023). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.
- Del Águila, L. M., Córdova, M. F. P., Sánchez, S. S. S., Dávila, A. T., & Mori, C. T. (2021). Escala de depresión de Zung: propiedades psicométricas en portadores del virus de la inmunodeficiencia humana. *Revista Peruana de Ciencias de la Salud*, 3(4), 261-269.
<https://doi.org/10.37711/rpcs.2021.3.4.358>
- Del Barrio V, Capilla M, Olmedo M, y Colodrón F. (2002). Primera adaptación del CDI-S a población española. *Acción Psicológica*, 1(3), 263-272.
<https://doi.org/10.5944/ap.1.3.558>
- Del Cid, M. T. C. (2021). La depresión y su impacto en la salud pública. *Revista médica hondureña*, 89(1), 46-52. <https://doi.org/10.5377/rmh.v89iSupl.1.12047>
- Dos Santos, M. M., Elias, C. S., & Della, P. C. D. S. (2025). Beaton's guidelines in cross-cultural adaptation studies in Brazil: a systematic review. *O Mundo da Saúde*, 49.
<https://doi.org/10.15343/0104-7809.202549e17202025I>
- Driscoll, D. L. (2011). Introduction to Primary Research: Observations, Surveys, and Interviews. En C. Lowe & P. Zemliansky (eds.), *Writing Spaces: Readings on Writing*, Volume 2 (p. 135-174). Parlor Press.

- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105(3), 399-412. <https://doi.org/10.1111/bjop.12046>
- Elejabo-Pacheco, G., Enriquez-Cuba, M., Conde Rodríguez, I. A., & Adriano-Rengifo, C. E. (2025). Evidencias preliminares de validez de la Escala de Trastorno Depresivo Persistente (TDP) en estudiantes universitarios peruanos. *Ciencias Psicológicas*, 19(2). <https://dialnet.unirioja.es/servlet/articulo?codigo=10357126>
- García-Batista, Z. E., Guerra-Peña, K., Cano-Vindel, A., Cantisano-Guzmán, L. M., Flores-Kanter, P. E., & Medrano, L. A. (2019). Medición de la comorbilidad depresiva en drogodependientes: Validez y eficacia diagnóstica del Inventario de Depresión de Beck (BDI-II) en dominicanos con múltiple consumo de sustancias. *Revista de Psicopatología y Psicología Clínica*, 24(3), 163-171. <https://dialnet.unirioja.es/servlet/articulo?codigo=7254979>
- González, C. E., De Greiff, E., & Avendaño, B. L. (2011). Comorbilidad entre ansiedad y depresión: evaluación empírica del modelo indefensión desesperanza. *Psychologia. Avances de la disciplina*, 5(1), 59-72. <https://www.redalyc.org/articulo.oa?id=297224114006>
- González-Forteza, C., Echeagaray, F. A. W., & Tapia, A. J. (2012). Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D) en México: análisis bibliométrico. *Salud mental*, 35(1), 13-20. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252012000100003
- Guerra, M., Ferri, C., Llibre, J., Prina, A. M., & Prince, M. (2015). Psychometric properties of EURO-D, a geriatric depression scale: a cross-cultural validation study. *BMC psychiatry*, 15(1), 12. <https://link.springer.com/article/10.1186/s12888-015-0390-4>

- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology*, 46(12), 1417-1432. [https://doi.org/10.1016/0895-4356\(93\)90142-N](https://doi.org/10.1016/0895-4356(93)90142-N)
- Guion, R. M. (1980). On Trinitarian doctrines of validity. *Professional Psychology*, 11(3), 385. <https://doi.org/10.1037/0735-7028.11.3.385>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J. (2024). *Cochrane Handbook for Systematic Reviews of Interventions*.
- Hilaire, C. S. (2023). Framing Indigenous perspectives through emic and etic approaches. Indigenous and minority populations. *Perspectives from scholars and writers across the world*, 259. https://books.google.com/books?hl=es&lr=&id=YATIEAAQBAJ&oi=fnd&pg=PA259&dq=Framing+Indigenous+perspectives+through+emic+and+etic+approaches.+Indigenous+and+minority+populations.+Perspectives+from+scholars+and+writers+across+the+world&ots=ssoi38L5wj&sig=ZiNZzkwEB8TXc3w8_8_pr1isBFA
- Huarcaya-Victoria, J., De-Lama-Morán, R., Quiros, M., Bazán, J., López, K., & Lora, D. (2020). Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú. *Revista de Neuro-Psiquiatría*, 83(2), 72-78. <https://doi.org/10.20453/rnp.v83i2.3749>
- Hunsley J, & Mash EJ (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work In Hunsley J & Mash EJ (Eds.), *A guide to assessments that work*. (pp. 3–14). Oxford University Press.
- International Test Commission. (2013). *Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. <https://www.intestcom.org/page/14>

- Kelly, P., O'Malley, K. J., Kallen, M. A., & Ford, M. E. (2005). Integrating validity theory with use of measurement instruments in clinical settings. *Health Services Research, 40*(5p2), 1605-1619. 10.1111/j.1475-6773.2005.00445.x
- Kilbourne, A. M, Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., Pincus, H. A. (2018). Measuring and improving the quality of mental health care: a global perspective. *World Psychiatry, 17*(1), 30-38. <https://doi.org/10.1002/wps.20482>
- Kleka, P., & Paluchowski, W. J. (2017). Shortening of psychological tests—assumptions, methods and doubts. *Polish Psychological Bulletin, 48*(4), 516-522. 10.1515/ppb-2017-0058
- Kovacs, M. (1992). *Children's Depression Inventory*. Multi-Health Systems.
- Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals, 32*(9), 509-515. <https://doi.org/10.3928/0048-5713-20020901-06>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kung, S., Alarcon, R. D., Williams, M. D., Poppe, K. A., Moore, M. J., & Frye, M. A. (2013). Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice. *Journal of affective disorders, 145*(3), 341-343. <https://doi.org/10.1016/j.jad.2012.08.017>
- Lam, N., Contreras, H., Mori, E., Gil, E., Espinoza, E., & Córdova, G. (2009). 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.

- Revista Peruana de Epidemiología*, 13(2), 1-4.
<https://www.redalyc.org/pdf/2031/203120363007.pdf>
- Lam, N., Contreras, H., Mori, E., Hinostroza, W., Hinostroza, R., Torrejón, E., ... & Hinostroza, W. (2010). Factores psicosociales y depresión antenatal en mujeres gestantes. Estudio multicéntrico en tres hospitales de Lima, Perú. Abril a junio de 2008. *Revista peruana de epidemiología*, 14(3), 193-200.
<https://www.redalyc.org/pdf/2031/203119676005.pdf>
- Lee-Diaz, E., Herrera-Rios, G., Espinoza-Quiroga, M., & Vega-Dienstmaier, J. M. (2024). Validación de la Escala de Psicopatología Depresiva de 4 ítems (EPD-4) aplicada en línea en Perú. *Revista de Neuro-Psiquiatría*, 87(4), 381-393.
<https://doi.org/10.20453/rnp.v87i4.5809>
- Levis, B., Sun, Y., He, C., Wu, Y., Krishnan, A., Bhandari, P. M., ... & Thombs, B. D. (2020). Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. *Jama*, 323(22), 2290-2300. <https://jamanetwork.com/journals/jama/fullarticle/2766865>
- Lingán-Huamán, S. K., Dominguez-Lara, S., Alarcón-Almeyda, M. S., de Moura, G. B., & Paiva, T. T. (2023). Psychometric properties of the Patient Health Questionnaire-9 in Peruvian university students. *Psychological Topics*, 32(3), 451-470.
<https://www.pt.ffri.hr/index.php/pt/article/view/827>
- Lipskaya-Velikovsky, L., & Krupa, T. (2023). Closing the gap: From research to practice in mental health interventions. *International Journal of Environmental Research and Public Health*, 20(3), 1-12. <https://doi.org/10.3390/ijerph20032141>
- Luján-Tangarife, J. A., & Cardona-Arias, J. A. (2015). Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. *Archivos de Medicina*, 11(3), 1-10. <https://dialnet.unirioja.es/servlet/articulo?codigo=5178935>

- Martinez, A., Teklu, S. M., Tahir, P., & Garcia, M. E. (2023). Validity of the Spanish-language patient health questionnaires 2 and 9: a systematic review and meta-analysis. *JAMA Network Open*, 6(10), e2336529. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2810774>
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and social psychology review*, 15(1), 28-50. <https://doi.org/10.1177/1088868310366253>
- Merino-Soto, C., Copez-Lonzoy, A., Toledano-Toledano, F., Nabors, L. A., Rodríguez-Castro, J. H., Hernández-Salinas, G., & Núñez-Benítez, M. Á. (2022). Effects of anonymity versus examinee name on a measure of depressive symptoms in adolescents. *Children*, 9(7), 972. <https://doi.org/10.3390/children9070972>
- Ministerio de Salud del Perú. (10 de enero de 2025). Establecimientos del Minsa atendieron más de 250 000 casos de depresión a lo largo del año 2024. <https://www.gob.pe/institucion/minsa/noticias/1088925-establecimientos-del-minsa-atendieron-mas-de-250-000-casos-de-depresion-a>
- Ministerio de Salud del Perú. (2008). *Guía de práctica clínica en depresión*. http://bvs.minsa.gob.pe/local/MINSA/1084_DGSP261.pdf
- Ministerio de Salud del Perú. (2021). *Manual de registro y codificación de actividades en la atención de salud mental. Sistema de información HIS*. DGIESP.
- Ministerio de Salud del Perú. (13 de febrero de 2022). *Minsa: Más de 300 mil casos de depresión fueron atendidos durante el 2021*. <https://www.gob.pe/institucion/minsa/noticias/575899-minsa-mas-de-300-mil-casos-de-depresion-fueron-atendidos-durante-el-2021>
- Ministerio de Salud del Perú. (23 de febrero de 2024). *La depresión es uno de los trastornos de salud mental más frecuentes en el país*.

<https://www.gob.pe/institucion/minsa/noticias/892900-la-depresion-es-uno-de-los-trastornos-de-salud-mental-mas-frecuentes-en-el-pais>

- Miranda-Castillo, C., Paddick, S. M., León-Campos, M. O., Molleda, P., Rosell, J., & Valenzuela, M. (2026). Psychometric Evidence of Instruments for Assessing Mental Health in Older Adults from Latin America and the Caribbean: A Scoping Review. *Healthcare*, 14(2), 265. <https://doi.org/10.3390/healthcare14020265>
- Montes, C. (2004). La depresión y su etiología: Una patología compleja. *Academia Biomédica Digital*, 2(1), 1-15. <https://caibco.ucv.ve/caibco/vitae/VitaeDieciocho/Articulos/Psiquiatria/ArchivosPDF/psiquiatria.pdf>
- Moscoso, M. S. (2014). Avances en la medición psicométrica de la depresión. *Liberabit*, 20(1), 29-39. <https://www.redalyc.org/pdf/686/68631260004.pdf>
- Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151-157. <https://www.redalyc.org/pdf/727/72726347014.pdf>
- Navarro-Loli, J. S., & Moscoso, M. (2017). La investigación de la depresión en adolescentes en el Perú: una revisión sistemática. *Liberabit*, 23(1), 57-74. <https://dialnet.unirioja.es/servlet/articulo?codigo=6246102>
- Novara, J., Sotillo, C., Warthon, D. E. (1985). *Estandarización y correlación de las escalas de Beck, Hamilton y Zung para depresión en población de Lima Metropolitana*. Instituto Nacional de Salud Mental Honorio Delgado – Hideyo Noguchi.
- Ocampo, L. (2007). Análisis correlacional del cuestionario Estado/Rasgo con una muestra de adolescentes y universitarios de Medellín (Colombia). *Psicología desde el Caribe*, 20, 28-49. <https://www.redalyc.org/articulo.oa?id=21302003>

- Organización de las Naciones Unidas para la Alimentación y la Agricultura. (2019). *Objetivos de Desarrollo Sostenible*. Food and Agriculture Organization.
- Organización Mundial de la Salud (2022). *World mental health report: Transforming mental health for all*. OMS. <https://www.who.int/publications/i/item/9789240049338>
- Organización Mundial de la Salud (2023). *Trastorno depresivo (depresión)*. <https://www.who.int/news-room/fact-sheets/detail/depression>
- Organización Mundial de la Salud. (2000). *Clasificación de los Trastornos Mentales y del Comportamiento, décima edición*. Panamericana.
- Organización Mundial de la Salud. (25 de mayo de 2019). *Eleventh revision of the International Classification of Diseases*. https://apps.who.int/gb/ebwha/pdf_files/WHA72/A72_29-en.pdf
- Organización Mundial de la Salud. (14 de febrero de 2023). *Clasificación de los Trastornos Mentales y del Comportamiento, décimo primera edición*. <https://icd.who.int/browse11/l-m/es#/http%3a%2f%2fid.who.int%2fcd%2fentity%2f76398729>
- Organización Mundial de la Salud. (20 de enero de 2023). *Depresión*. [https://www.who.int/es/news-room/fact-sheets/detail/depression#:~:text=Se%20estima%20que%20el%203,personas%20sufren%20depresi%C3%B3n%20\(1\).](https://www.who.int/es/news-room/fact-sheets/detail/depression#:~:text=Se%20estima%20que%20el%203,personas%20sufren%20depresi%C3%B3n%20(1).)
- Organización Panamericana de la Salud. (30 de agosto de 2019). *Vigencia de la CIE -11*. <https://www.paho.org/es/relacsis/foro-dr-roberto-becker/vigencia-cie-11>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Alonso-Fernández, S. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista española de cardiología*, 74(9), 790-799. <https://doi.org/10.1016/j.recesp.2021.06.016>

- Parada, K. R., Guapizaca, J. F., & Bueno, G. A. (2022). Deterioro cognitivo y depresión en adultos mayores: una revisión sistemática de los últimos 5 años. *Revista Científica UISRAEL*, 9(2), 77-93. https://scielo.senescyt.gob.ec/scielo.php?script=sci_arttext&pid=S2631-27862022000200077
- Pérez-Padilla, E. A., Cervantes-Ramírez, V. M., Hijuelos-García, N. A., Pineda-Cortés, J. C., & Salgado-Burgos, H. (2017). Prevalencia, causas y tratamiento de la depresión Mayor. *Revista biomédica*, 28(2), 73-98. <https://doi.org/10.32776/revbiomed.v28i2.557>
- Polaino-Lorente, A., & Ascaso, L. E. (1988). *Las depresiones infantiles*. Ediciones Morata.
- Prince, M. J., Reischies, F., Beekman, A. T., Fuhrer, R., Jonker, C., Kivela, S. L., ... & Copeland, J. R. (1999). Development of the EURO-D scale—a European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry*, 174(4), 330-338. <https://doi.org/10.1192/bjp.174.4.330>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3), 385-401. <https://doi.org/10.1177/014662167700100306>
- Ramada-Rodilla, J. M., Serra-Pujadas, C., & Delclós-Clanchet, G. L. (2013). Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud pública de México*, 55, 57-66. <https://www.scielosp.org/pdf/spm/2013.v55n1/57-66/es>
- Ramos-Vera, C., Quispe Callo, G., Basauri Delgado, M., Vallejos Saldarriaga, J., & Saintila, J. (2023). Factorial and network structure of the Reynolds Adolescent Depression Scale (RADs-2) in Peruvian adolescents. *Plos one*, 18(5), 1-16. <https://doi.org/10.1371/journal.pone.0346101>

- Renteria, E. V. C., Del Pilar, E., Pauta, I. D. C., Diaz, E. G., & Velasquez, O. A. R. (2022). Efectos de estilos de crianza sobre la depresión en adolescentes escolarizados de una muestra peruana. *Revista de Investigación en Psicología*, 25(2), 39-54. <https://dialnet.unirioja.es/servlet/articulo?codigo=8754169>
- Resolución Ministerial N.º 136-2023-MINSA. Aprobar la Guía Técnica: Guía de Práctica Clínica para Tamizaje, Diagnóstico y Tratamiento de la Depresión en Personas Mayores de 18 años desde el Primer Nivel de Atención, que como Anexo forma parte de la presente Resolución Ministerial y que se publica en la sede digital del Ministerio de Salud. (10 de febrero de 2023). <https://www.gob.pe/institucion/minsa/normas-legales/3882024-136-2023-minsa>
- Retamal, P. (1998). *Depresión*. Editorial Universitaria.
- Reynolds, W. (1987). *Reynolds Adolescent Depression Scale. Professional Manual*. Psychological Assessment Resources, Inc. PAR.
- Rivera, B. M., Corrales, A. E., Cáceres, Ó., & Pina, J. A. (2007). Validación de la Escala de Depresión de Zung en Personas con VIH. *Terapia psicológica*, 25(2), 135-140. <http://dx.doi.org/10.4067/S0718-48082007000200004>
- Robles, Y. (2008). Base de Datos de instrumentos de evaluación de Salud Mental y Psiquiatría. *Anales de Salud Mental*, 24(1). <https://openjournal.insm.gob.pe/revistasm/asm/article/view/38>
- Rodríguez, E. A., y Farfán, D. (2015). Propiedades Psicométricas del Inventario de Depresión de Beck en estudiantes de 5to año de secundaria de Huancayo, 2013. *Revista de Investigación Universitaria*, 4(2). <https://revistas.upeu.edu.pe/index.php/riu/article/view/743>
- Romera, I., Delgado-Cohen, H., Perez, T., Caballero, L., & Gilaberte, I. (2008). Factor analysis of the Zung self-rating depression scale in a large sample of patients with major

- depressive disorder in primary care. *BMC psychiatry*, 8(1), 1-8.
<https://link.springer.com/article/10.1186/1471-244X-8-4>
- Ruiz-Grosso, P., Loret de Mola, C., Vega-Dienstmaier, J. M., Arevalo, J. M., Chavez, K., Vilela, A., ... & Huapaya, J. (2012). Validation of the spanish center for epidemiological studies depression and zung self-rating depression scales: a comparative validation study. *PLoS ONE*, 7(10), e45413. doi:10.1371/journal.pone.0045413.
- Saavedra, J. (2018). Estudio Epidemiológico de Salud Mental en Hospitales Generales y Centros de Salud de Lima Metropolitana 2015. *Anales de Salud Mental*, 34(1).
<https://openjournal.insm.gob.pe/revistasm/asm/article/view/15>
- Sánchez-Pachas, D. V. (2019). Propiedades Psicométricas del Inventario de Depresión de Beck-II (IDB-II) en una muestra clínica. *Revista de investigación en psicología*, 22(1), 39-52. <https://dialnet.unirioja.es/servlet/articulo?codigo=8162691>
- Sánchez-Villena, A. R., & Farfán, E. (2019). Análisis factorial exploratorio del Inventario de Depresión de Beck (BDI-II) en universitarios cajamarquinos. *Interacciones. Revista de Avances en Psicología*, 5(3). <https://doi.org/10.24016/2019.v5n3.177>
- Sánchez-Villena, A. R., Farfán, E., & Chávez-Ravines, D. (2022). Estructura factorial y datos normativos del Inventario de Depresión de Beck (BDI-II) en población general peruana. *Acta Colombiana de Psicología*, 25(2), 158-170.
<https://www.redalyc.org/articulo.oa?id=79872419011>
- Sarmiento, B. N. (2020). *Evaluación dimensional de la depresión: una revisión sistemática de estudios instrumentales en adultos hispanohablantes*. [Tesis de pregrado, Corporación Universidad de la Costa]. Repositorio Institucional CUC.
<https://repositorio.cuc.edu.co/entities/publication/96d3f004-55bf-441d-ada3-5a5600eec089>

- Sheikh, J. I., & Yesavage, J. A. (2014). *Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clinical gerontology*. Routledge.
- Siniscalchi, K. A., Broome, M. E., Fish, J., Ventimiglia, J., Thompson, J., Roy, P., ... & Trivedi, M. (2020). Depression screening and measurement-based care in primary care. *Journal of primary care & community health*, 11, (1-11).
<https://doi.org/10.1177/2150132720931261>
- Smith, M. L., Sanchez, S. E., Rondon, M., Gradus, J. L., & Gelaye, B. (2022). Validation of the patient health Questionnaire-9 (PHQ-9) for detecting depression among pregnant women in Lima, Peru. *Current Psychology*, 41(6), 3797-3805.
- Sotelo, L., Sotelo, N., Dominguez, S., Poma, I., Cueto, E., Alarcón, D., ... & Padilla, O. (2012). Propiedades psicométricas del inventario de depresión estado-rasgo (IDER) en adultos de Lima. *Avances en psicología*, 20(2), 59-68.
<https://revistas.unife.edu.pe/index.php/avancesenpsicologia/article/view/319>
- Soto, C. M., Muñoz, D. P., & Riaño-Hernández, D. (2012). Análisis factorial exploratorio del Inventario de Depresión Estado-Rasgo (ST-DEP) en adolescentes. *Diversitas: Perspectivas en Psicología*, 8(2), 319-330.
<https://www.redalyc.org/pdf/679/67925837008.pdf>
- Spielberger, C. D., Agudelo, D. y Buela-Casal, G. (2008). *Inventario de Depresión Estado/Rasgo (IDER)*. TEA Ediciones.
- Spielberger, C.D., Carretero-Dios, H., De los Santos-Roig, M. y Buela-Casal. G. (2002a). Spanish versión of the State Depression Questionnaire (ST-DEP): State subscale (S-DEP). *International Journal of Clinical and Health Psychology*, 2(1), 71-89.
<https://www.redalyc.org/pdf/337/33720104.pdf>
- Spielberger, C.D., Carretero-Dios, H., De los Santos-Roig, M. y Buela-Casal. G. (2002b). Spanish experimental version of the State-Trait Depression Questionnaire (ST-DEP:

- Trait sub-scale (T-DEP). *International Journal of Clinical and Health Psychology*, 2(2), 51-69. <https://www.redalyc.org/pdf/337/33720104.pdf>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Patient Health Questionnaire Primary Care Study Group. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Jama*, 282(18), 1737-1744. <https://jamanetwork.com/journals/jama/fullarticle/192080>
- Tamayo, J. M., Rosales-Barrera, J. I., Villaseñor-Bayardo, S. J., & Rojas-Malpica, C. (2011). Revisión de la literatura médica sobre el manejo de las depresiones resistentes/refractarias al tratamiento. *Salud mental*, 34(3), 257-266. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252011000300009
- Terol-Cantero, M. C., Cabrera-Perona, V., & Martín-Aragón, M. (2015). Revisión de estudios de la Escala de Ansiedad y Depresión Hospitalaria (HAD) en muestras españolas. *Anales de Psicología/Annals of Psychology*, 31(2), 494-503. <https://dx.doi.org/10.6018/analesps.31.2.172701>
- Tornimbeni, S., Pérez, E., Olaz, F., de Kohan, N. C., Fernández, A., & Cupani, M. (2008). *Introducción a la psicometría*. Paidós.
- Urrútia, G., & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina clínica*, 135(11), 507-511. https://bmjopen.bmj.com/content/bmjopen/suppl/2013/06/10/bmjopen-2012-002330.DC1/bmjopen-2012-002330supp_PRISMA-2010.pdf
- Vega-Dienstmaier, J. M. (2018). Construcción de versiones cortas de la Escala de Psicopatología Depresiva para tamizaje de depresión mayor y sus características psicométricas. *Revista de Neuro-Psiquiatría*, 81(3), 154-154. <https://doi.org/10.20453/rnp.v81i3.3383>

- Vega-Dienstmaier, J. M. (2022). Estudios psicométricos de los instrumentos en español para la detección de trastornos mentales comunes: una revisión. *Revista de Neuro-Psiquiatría*, 85(2), 127-138. <https://doi.org/10.20453/rnp.v85i2.4230>
- Vega-Dienstmaier, J. M., Coronado-Molina, O., & Mazzotti, G. (2014). Validez de una versión en español del Inventario de Depresión de Beck en pacientes hospitalizados de medicina general. *Revista de neuro-psiquiatria*, 77(2), 95-103. <https://www.redalyc.org/pdf/3720/372033986006.pdf>
- Vega-Dienstmaier, J. M., Stucchi-Portocarrero, S., Valdez-Huarcaya, N., Cabra-Bravo, M., & Zapata-Vega, M. I. (2011). The Depressive Psychopathology Scale: presentation and initial validation in a sample of Peruvian psychiatric patients. *Revista Panamericana de Salud Pública*, 30, 317-326. <https://www.scielosp.org/pdf/rpsp/2011.v30n4/317-326/en>
- Vega-Dienstmaier, J. M., Suárez, G. M., & Sánchez, M. C. (2002). Validación de una versión en español de la Escala de Depresión Postnatal de Edimburgo. *Actas españolas de psiquiatria*, 30(2), 106-111. <https://actaspsiquiatria.es/index.php/actas/article/view/977>
- Véliz, J. (1998). Validez y confiabilidad del cuestionario autoreporte en personal de una institución militarizada. *Anales de Salud Mental*, 14, 85-102. <https://openjournal.insm.gob.pe/revistas/asm/article/view/226>
- Ventura-León, J., Caycho-Rodríguez, T., Tocco-Muñoz, S., Torres-Maldonado, M., & Curahua-Guillén, K. (2020). Evidencias de validez y fiabilidad de la versión en español del Children's Depression Inventory-Short en población peruana. *Revista de Psicopatología y Psicología Clínica*, 25(2), 111-119. https://www.aepcp.net/wp-content/uploads/2020/10/31018_Psicopatologia_Vol_25_N2_WEB_Parte6.pdf
- Villarreal-Zegarra, D., Barrera-Begazo, J., Otazú-Alfaro, S., Mayo-Puchoc, N., Bazo-Alvarez, J. C., & Huarcaya-Victoria, J. (2023). Sensitivity and specificity of the Patient Health

- Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: a cross-sectional study in a Peruvian hospital population. *BMJ open*, 13(9), e076193. <https://doi.org/10.1136/bmjopen-2023-076193>
- Villarreal-Zegarra, D., Cabrera-Alva, M., Carrillo-Larco, R. M., & Bernabe-Ortiz, A. (2020). Trends in the prevalence and treatment of depressive symptoms in Peru: a population-based study. *BMJ open*, 10(7), e036777. <https://doi.org/10.1136/bmjopen-2020-036777>
- Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabé-Ortiz, A., Melendez-Torres, G. J., & Bazo-Alvarez, J. C. (2019). Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics. *PloS one*, 14(9), e0221717. <https://doi.org/10.1371/journal.pone.0221717>
- Villasís-Keever, M. Á., Rendón-Macías, M. E., García, H., Miranda-Novales, M. G., & Escamilla-Núñez, A. (2020). La revisión sistemática y el metaanálisis como herramientas de apoyo para la clínica y la investigación. *Revista Alergia México*, 67(1), 62-72. <https://doi.org/10.29262/ram.v67i1.733>
- Vivar, R., Pacheco, Z., Adrianzen, C., Macciotta, B., & Marchena, C. (2005). Validación de la Escala de Birleson modificada para trastornos depresivos en niños y adolescentes peruanos. *Revista Peruana de Pediatría*, 58(3), 24-30. <https://doi.org/10.61651/rped.2005v58n3p24-30>
- Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Brazilian Journal of Psychiatry*, 35, 416-431. <https://doi.org/10.1590/1516-4446-2012-1048>
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: a

preliminary report. *Journal of psychiatric research*, 17(1), 37-49.

[https://doi.org/10.1016/0022-3956\(82\)90033-4](https://doi.org/10.1016/0022-3956(82)90033-4)

Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong ML, & Youngstrom JK (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363. <https://doi.org/10.1111/cpsp.12207>

Zhong, Q., Gelaye, B., Rondon, M., Sánchez, S. E., García, P. J., Sánchez, E., ... & Williams, M. A. (2014). Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. *Journal of affective disorders*, 162, 1-7. <https://doi.org/10.1016/j.jad.2014.03.028>

Zung, W. W. (1965). A self-rating depression scale. *Archives of general psychiatry*, 12(1), 63-70. doi:10.1001/archpsyc.1965.01720310065008

IX. ANEXOS

Anexo A. Matriz de consistencia

Objetivo	Variables	Dimensiones	Indicadores	Método
Objetivo general: - Analizar las características psicométricas de los instrumentos para medir la depresión validados en población peruana. Objetivos específicos: - Identificar los instrumentos para medir la depresión validados en población peruana y las características de las muestras en las que fueron aplicados. - Describir los procedimientos de adaptación cultural de los instrumentos para medir depresión validados en población peruana. - Describir la evidencia de confiabilidad de los instrumentos para medir depresión validados en población peruana. - Describir la evidencia de validez de los instrumentos para medir depresión validados en población peruana. - Describir las normas de las puntuaciones de los instrumentos para medir depresión validados en población peruana.	Medición de la depresión Definición conceptual La medición de la depresión se conceptualiza como el proceso de cuantificar y evaluar los síntomas y la gravedad de la depresión en una persona a través del uso de herramientas psicométricas, tales como escalas o cuestionarios específicos diseñados para evaluar el estado emocional, mental y conductual de un individuo en relación con la depresión (Caballo et al., 2014; Moscoso, 2014). Definición operacional Se operacionaliza la medición de la depresión a través de la rúbrica de adaptaciones y validaciones de instrumentos psicométricos que considera los 5 aspectos fundamentales para la adecuada medición psicométrica de dicha variable: adaptación cultural, confiabilidad, validez, y normas (AERA, APA y NCME, 2018; Luján-Tangarife y Cardona-Arias, 2015; Muñiz et al., 2013).	Adaptación cultural	Etapa 1: Traducción directa al idioma de interés Etapa 2: Síntesis de traducciones y solución de discrepancias Etapa 3: Traducción inversa (idioma destino - idioma original) Etapa 4: Revisión por comité de expertos Etapa 5: Prueba piloto	Tipo de investigación: Investigación secundaria, diseño teórico Instrumento: Rúbrica de adaptaciones y validaciones de instrumentos psicométricos
		Confiabilidad/precisión	Coeficientes de formularios alternativos Coeficientes test-retest Coeficientes de coherencia interna Índices de coherencia entre evaluadores	
		Validez	Evidencia basada en el contenido de la prueba Evidencia basada en los procesos de respuesta Evidencia basada en la estructura interna Evidencia basada en relaciones con otras variables Evidencia basada en las consecuencias de las pruebas	
		Normas	Sensibilidad (s) y especificidad (e) Baremación	

Anexo B. Protocolo PRISMA 2020

Sección/tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la publicación
TÍTULO			
Título	1	Identifique la publicación como una revisión sistemática.	
RESUMEN			
Resumen estructurado	2	Vea la lista de verificación para resúmenes estructurados de la declaración PRISMA 2020 (tabla 2).	
INTRODUCCIÓN			
Justificación	3	Describa la justificación de la revisión en el contexto del conocimiento existente.	
Objetivos	4	Proporcione una declaración explícita de los objetivos o las preguntas que aborda la revisión.	
MÉTODOS			
Criterios de elegibilidad	5	Especifique los criterios de inclusión y exclusión de la revisión y cómo se agruparon los estudios para la síntesis.	
Fuentes de información	6	Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otros recursos de búsqueda o consulta para identificar los estudios. Especifique la fecha en la que cada recurso se buscó o consultó por última vez.	
Estrategia de búsqueda	7	Presente las estrategias de búsqueda completas de todas las bases de datos, registros y sitios web, incluyendo cualquier filtro y los límites utilizados.	
Proceso de selección de los estudios	8	Especifique los métodos utilizados para decidir si un estudio cumple con los criterios de inclusión de la revisión, incluyendo cuántos autores de la revisión cribaron cada registro y cada publicación recuperada, si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	
Proceso de extracción de los datos	9	Indique los métodos utilizados para extraer los datos de los informes o publicaciones, incluyendo cuántos revisores recopilaron datos de cada publicación, si trabajaron de manera independiente, los procesos para obtener o confirmar los datos por parte de los investigadores del estudio y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	
Lista de los datos	10a	Enumere y defina todos los desenlaces para los que se buscaron los datos. Especifique si se buscaron todos los resultados compatibles con cada dominio del desenlace (por ejemplo, para todas las escalas de medida, puntos temporales, análisis) y, de no ser así, los métodos utilizados para decidir los resultados que se debían recoger.	
	10b	Enumere y defina todas las demás variables para las que se buscaron datos (por ejemplo, características de los participantes y de la intervención, fuentes de financiación). Describa todos los supuestos formulados sobre cualquier información ausente (missing) o incierta.	
Evaluación del riesgo de sesgo de los estudios individuales	11	Especifique los métodos utilizados para evaluar el riesgo de sesgo de los estudios incluidos, incluyendo detalles de las herramientas utilizadas, cuántos autores de la revisión evaluaron cada estudio y si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	
Medidas del efecto	12	Especifique, para cada desenlace, las medidas del efecto (por ejemplo, razón de riesgos, diferencia de medias) utilizadas en la síntesis o presentación de los resultados.	
Métodos de síntesis	13a	Describa el proceso utilizado para decidir qué estudios eran elegibles para cada síntesis (por ejemplo, tabulando las características de los estudios de intervención y comparándolas con los grupos previstos para cada síntesis (ítem n.º 5)).	
	13b	Describa cualquier método requerido para preparar los datos para su presentación o síntesis, tales como el manejo de los datos perdidos en los estadísticos de resumen o las conversiones de datos.	
	13c	Describa los métodos utilizados para tabular o presentar visualmente los resultados de los estudios individuales y su síntesis.	
	13d	Describa los métodos utilizados para sintetizar los resultados y justifique sus elecciones. Si se ha realizado un metanálisis, describa los modelos, los métodos para identificar la presencia y el alcance de la heterogeneidad estadística, y los programas informáticos utilizados.	
	13e	Describa los métodos utilizados para explorar las posibles causas de heterogeneidad entre los resultados de los estudios (por ejemplo, análisis de subgrupos, metarregresión).	
	13f	Describa los análisis de sensibilidad que se hayan realizado para evaluar la robustez de los resultados de la síntesis.	

Sección/tema	Ítem n.	Ítem de la lista de verificación	Localización del ítem en la publicación
Evaluación del sesgo en la publicación	14	Describa los métodos utilizados para evaluar el riesgo de sesgo debido a resultados faltantes en una síntesis (derivados de los sesgos en las publicaciones).	
Evaluación de la certeza de la evidencia	15	Describa los métodos utilizados para evaluar la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace.	
RESULTADOS			
Selección de los estudios	16a	Describa los resultados de los procesos de búsqueda y selección, desde el número de registros identificados en la búsqueda hasta el número de estudios incluidos en la revisión, idealmente utilizando un diagrama de flujo (ver figura 1).	
	16b	Cite los estudios que aparentemente cumplían con los criterios de inclusión, pero que fueron excluidos, y explique por qué fueron excluidos.	
Características de los estudios	17	Cite cada estudio incluido y presente sus características.	
Riesgo de sesgo de los estudios individuales	18	Presente las evaluaciones del riesgo de sesgo para cada uno de los estudios incluidos.	
Resultados de los estudios individuales	19	Presente, para todos los desenlaces y para cada estudio: a) los estadísticos de resumen para cada grupo (si procede) y b) la estimación del efecto y su precisión (por ejemplo, intervalo de credibilidad o de confianza), idealmente utilizando tablas estructuradas o gráficos.	
Resultados de la síntesis	20a	Para cada síntesis, resuma brevemente las características y el riesgo de sesgo entre los estudios contribuyentes.	
	20b	Presente los resultados de todas las síntesis estadísticas realizadas. Si se ha realizado un metanálisis, presente para cada uno de ellos el estimador de resumen y su precisión (por ejemplo, intervalo de credibilidad o de confianza) y las medidas de heterogeneidad estadística. Si se comparan grupos, describa la dirección del efecto.	
	20c	Presente los resultados de todas las investigaciones sobre las posibles causas de heterogeneidad entre los resultados de los estudios.	
	20d	Presente los resultados de todos los análisis de sensibilidad realizados para evaluar la robustez de los resultados sintetizados.	
Sesgos en la publicación	21	Presente las evaluaciones del riesgo de sesgo debido a resultados faltantes (derivados de los sesgos de en las publicaciones) para cada síntesis evaluada.	
Certeza de la evidencia	22	Presente las evaluaciones de la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace evaluado.	
DISCUSIÓN			
Discusión	23a	Proporcione una interpretación general de los resultados en el contexto de otras evidencias.	
	23b	Argumente las limitaciones de la evidencia incluida en la revisión.	
	23c	Argumente las limitaciones de los procesos de revisión utilizados.	
	23d	Argumente las implicaciones de los resultados para la práctica, las políticas y las futuras investigaciones.	
OTRA INFORMACIÓN			
Registro y protocolo	24a	Proporcione la información del registro de la revisión, incluyendo el nombre y el número de registro, o declare que la revisión no ha sido registrada.	
	24b	Indique dónde se puede acceder al protocolo, o declare que no se ha redactado ningún protocolo.	
	24c	Describa y explique cualquier enmienda a la información proporcionada en el registro o en el protocolo.	
Financiación	25	Describa las fuentes de apoyo financiero o no financiero para la revisión y el papel de los financiadores o patrocinadores en la revisión.	
Conflicto de intereses	26	Declare los conflictos de intereses de los autores de la revisión.	
Disponibilidad de datos, códigos y otros materiales	27	Especifique qué elementos de los que se indican a continuación están disponibles al público y dónde se pueden encontrar: plantillas de formularios de extracción de datos, datos extraídos de los estudios incluidos, datos utilizados para todos los análisis, código de análisis, cualquier otro material utilizado en la revisión.	

Anexo C. Artículos evaluados

Autor	Instrumento	Versión	Título
Vivar et al. (2005)	Escala de Depresión de Birlson modificada (DSRS)	DSRS	Validación de la Escala de Birlson modificada para trastornos depresivos en niños y adolescentes peruanos
De la Torre et al. (2006)	Escala de Depresión Geriátrica - 4 ítems (GDS-4)	GDS-4	Validación de la versión reducida de la escala de depresión geriátrica en el consultorio externo de geriatría del Hospital Nacional Cayetano Heredia
Ruiz-Grosso et al. (2012)	Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D)	CES-D	Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A Comparative Validation Study
Guerra et al. (2015)	Escala Europea de Depresión (EURO-D)	EURO-D	Psychometric properties of EURO-D, a geriatric depression scale: a cross-cultural validation study
Elejabo-Pacheco et al. (2025)	Escala de Trastorno Depresivo Persistente (TDP)	TDP	Evidencias preliminares de validez de la Escala de Trastorno Depresivo Persistente (TDP) en estudiantes universitarios peruanos
Soto et al. (2012)	Inventario de Depresión Estado-Rasgo (IDER)	IDER	Análisis factorial exploratorio del Inventario de Depresión Estado-Rasgo (ST-DEP) en adolescentes
Sotelo et al. (2012)	Inventario de Depresión Estado-Rasgo (IDER)	IDER	Propiedades Psicométricas del Inventario de Depresión Estado-Rasgo (IDER) en Adultos de Lima
Ramos-Vera et al. (2023)	Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Factorial and network structure of the Reynolds Adolescent Depression Scale (RADDS-2) in Peruvian adolescents
Chávez y Ecurra (2002)	Escala de Depresión para Adolescentes de Reynolds (EDAR)	EDAR	Adaptación psicométrica de la Escala de Depresión para Adolescentes de Reynolds (EDAR) en estudiantes de secundaria de Lima metropolitana
Merino-Soto et al. (2022)	Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Effects of anonymity versus examinee name on a measure of depressive symptoms in adolescents
Ventura-León et al. (2020)	Inventario de Depresión para Niños - versión corta (CDI-S)	CDI-S	Evidencias de validez y fiabilidad de la versión en español del Children's Depression Inventory-Short en población peruana
Zhong et al. (2014)	Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression
Lam et al. (2010)	Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Factores psicosociales y depresión antenatal en mujeres gestantes. Estudio multicéntrico en tres hospitales de Lima, Perú. Abril a junio de 2008

Autor	Instrumento	Versión	Título
Vega-Dienstmaier et al. (2002)	Escala de Depresión Posnatal de Edimburgo (EPDS)	EPDS	Validación de una versión en español de la Escala de Depresión Postnatal de Edimburgo
Ruiz-Grosso et al. (2012)	Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A Comparative Validation Study
Del Aguila et al. (2021)	Escala de Autoevaluación para la Depresión de Zung (ZSDS)	ZSDS	Escala de depresión de Zung: propiedades psicométricas en portadores del virus de la inmunodeficiencia humana
Vega-Dienstmaier (2018)	Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Construcción de versiones cortas de la Escala de Psicopatología Depresiva para tamizaje de depresión mayor y sus características psicométricas
Vega-Dienstmaier et al. (2011)	Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	The Depressive Psychopathology Scale: presentation and initial validation in a sample of Peruvian psychiatric patients
Lee-Diaz et al. (2024)	Escala de Psicopatología Depresiva (EPD/DPS)	EPD/DPS	Validación de la Escala de Psicopatología Depresiva de 4 ítems (EPD-4) aplicada en línea en Perú
Vega-Dienstmaier et al. (2014)	Inventario de Depresión de Beck (BDI)	BDI	Validez de una versión en español del Inventario de Depresión de Beck en pacientes hospitalizados de medicina general
Sánchez-Villena y Farfán (2019)	Inventario de Depresión de Beck (BDI)	BDI-II	Análisis Factorial Exploratorio del Inventario de Depresión de Beck (BDI-II) en Universitarios Cajamarquinos
Sánchez-Villena et al. (2022)	Inventario de Depresión de Beck (BDI)	BDI-II	Estructura factorial y datos normativos del Inventario de Depresión de Beck (BDI-II) en población general peruana
Sánchez-Pachas (2019)	Inventario de Depresión de Beck (BDI)	BDI-II	Propiedades Psicométricas del Inventario de Depresión de Beck-II (IDB-II) en una muestra clínica
Rodríguez y Farfán (2015)	Inventario de Depresión de Beck (BDI)	BDI-IIA	Propiedades Psicométricas del Inventario de Depresión de Beck en estudiantes de 5to año de secundaria de Huancayo, 2013
Carranza (2013)	Inventario de Depresión de Beck (BDI)	BDI-IIA	PROPIEDADES PSICOMÉTRICAS DEL INVENTARIO DE DEPRESIÓN DE BECK UNIVERSITARIOS DE LIMA
Lam et al. (2009)	Inventario de Depresión de Beck (BDI)	BDI-PC	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
Baños-Chaparro et al. (2021)	Patient Health Questionnaire (PHQ)	PHQ 2	Evidencias psicométricas del Patient Health Questionnaire-2 para la evaluación de síntomas depresivos en adultos peruanos

Autor	Instrumento	Versión	Título
Villarreal-Zegarra et al. (2023)	Patient Health Questionnaire (PHQ)	PHQ 2	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population
Villarreal-Zegarra et al. (2023)	Patient Health Questionnaire (PHQ)	PHQ 8	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population
Cjuno et al. (2023)	Patient Health Questionnaire (PHQ)	PHQ 9	Adaptación cultural al Quechua y análisis psicométrico del Patient Health Questionnaire (PHQ-9) en población peruana
Renteria et al. (2022)	Patient Health Questionnaire (PHQ)	PHQ 9	Efectos de estilos de crianza sobre la depresión en adolescentes escolarizados de una muestra peruana
Huarcaya-Victoria et al. (2020)	Patient Health Questionnaire (PHQ)	PHQ 9	Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú
Villarreal-Zegarra et al. (2020)	Patient Health Questionnaire (PHQ)	PHQ 9	Trends in the prevalence and treatment of depressive symptoms in Peru: A population-based study
Villarreal-Zegarra et al. (2019)	Patient Health Questionnaire (PHQ)	PHQ 9	Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics
Alvarez et al. (2022)	Patient Health Questionnaire (PHQ)	PHQ 9	Validación del Cuestionario sobre la Salud del Paciente-9 (PHQ-9) en internos de medicina humana de una universidad de referencia del Perú durante la pandemia covid-19
Zhong et al. (2014)	Patient Health Questionnaire (PHQ)	PHQ 9	Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression
Lingán-Huamán et al. (2023)	Patient Health Questionnaire (PHQ)	PHQ 9	Psychometric Properties of the Patient Health Questionnaire-9 in Peruvian University Students
Villarreal-Zegarra et al. (2023)	Patient Health Questionnaire (PHQ)	PHQ 9	Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population
Smith et al. (2022)	Patient Health Questionnaire (PHQ)	PHQ 9	Validation of the patient health Questionnaire-9 (PHQ-9) for detecting depression among pregnant women in Lima, Peru

Nota: La tabla está ordenada según la columna de instrumentos, algunas investigaciones abordan más de un instrumento, por lo que pueden repetirse.