

# Sesgo de género en instrumentos de medida de la depresión. Revisión sistemática \*

## Gender Bias in Depression Measurement Instruments: A Systematic Review

Recibido: 24 septiembre 2025 | Aceptado: 11 noviembre 2025

BARTOLOMÉ CANTADOR TORIL<sup>a</sup>

Universidad de Córdoba, España

ORCID: <https://orcid.org/0000-0001-5904-6589>

ADORACIÓN ANTOLÍ CABRERA

Universidad de Córdoba, España

ORCID: <https://orcid.org/0000-0001-7259-5080>

### RESUMEN

Las mujeres presentan tasas de prevalencia de trastornos depresivos aproximadamente 1.7 veces superiores a las de los hombres. Ello podría estar influenciado, al menos en parte, por un sesgo de género, por el que mujeres y hombres obtienen puntuaciones distintas en los instrumentos de medida aun con niveles similares en el rasgo evaluado. Esta revisión sistemática resume la evidencia reciente sobre sesgo de género en ocho instrumentos de medida frecuentemente usados y sobre diferencias de género en otros aspectos de la medida de la depresión. Se identificaron 20 ítems con funcionamiento diferencial: las mujeres obtuvieron puntuaciones más altas en 15 ítems (13 en infancia y adolescencia) y los hombres en cuatro (tres en infancia y adolescencia) e inferiores en uno. Como conclusión, seis instrumentos frecuentemente usados evalúan la depresión de manera diferente según el género, presentando indicios de sesgo que podrían explicar las diferencias en las tasas de prevalencia.

### Palabras clave

depresión; género; sesgo; funcionamiento diferencial; instrumentos de medida.

### ABSTRACT

Depressive disorders prevalence rates in women are approximately 1.7 times higher than those in men. This could be influenced, at least in part, by gender bias due to the differential functioning of certain items in measurement instruments, which occurs when women and men are assigned different scores despite having similar levels of the trait that is evaluated. Recent evidence on gender bias in instruments for measuring depression due to the differential functioning of items in eight frequently used instruments was reviewed. The comparative analysis of scores obtained identified 20 items featuring differential functioning and gender differences in other aspects related to depression measurement. Women scored higher on 15 items (13 in children and adolescents), while men scored higher on four items (three in children and adolescents), and lower on one item. Conclusion: six frequently used instruments evaluate depression differently depending on gender, presenting signs of bias that could explain the differences in depressive disorders prevalence rates.

### Keywords

depression; gender; bias; differential functioning; measuring instruments.

<sup>a</sup> Autor de correspondencia. Correo electrónico: [oc2catob@uco.es](mailto:oc2catob@uco.es)

Como citar este artículo: Cantador, B., & Antolí, A. (2025). Sesgo de género en instrumentos de medida de la depresión: Revisión sistemática. *Universitas Psychologica*, 24, 1–15. <https://doi.org/10.11144/Javeriana.upsy24.sgim>

Los trastornos depresivos se caracterizan por un estado de ánimo persistentemente triste, vacío o irritable, acompañado de cambios somáticos y cognitivos que afectan significativamente la capacidad funcional (American Psychiatric Association [APA], 2022). El cuadro clínico principal es el trastorno depresivo mayor, que se caracteriza por episodios de al menos dos semanas de duración, con cambios claros en el afecto, la cognición y las funciones neurovegetativas. Estos episodios pueden asociarse con otras condiciones, como trastornos por abuso de sustancias o el uso de ciertos medicamentos.

Nolen-Hoeksema et al. (1999) identificaron las diferencias de género en la prevalencia de la depresión como uno de los fenómenos más consistentes de la epidemiología de los trastornos del estado de ánimo. Los estudios muestran mayor vulnerabilidad de las mujeres a la depresión que los hombres, independientemente del país, la cultura o el grupo étnico. Además, las mujeres son diagnosticadas con mayor frecuencia con trastornos depresivos y manifiestan más síntomas depresivos (Nolen-Hoeksema, 1987). Según Kessler et al. (1993), las tasas de prevalencia de depresión en mujeres son 1.7 veces superiores a las de los hombres. Seedat et al. (2009) encontraron que las mujeres presentan un riesgo significativamente mayor de padecer la mayoría de los trastornos del estado de ánimo.

Iriarte y Medina (2024) analizaron personas de 50 años o más de 35 países y encontraron una probabilidad prevista de depresión del 24.01% para los hombres y del 35.7% para las mujeres, con una brecha de género de 11.69%. Bacigalupe et al. (2020) señalan la mayor probabilidad de que las mujeres sean diagnosticadas de depresión y reciban prescripción de psicofármacos, pues son especialmente vulnerables a los mecanismos de medicalización de la salud mental.

Por otra parte, los hombres tienen menos probabilidades de ser diagnosticados con un trastorno depresivo (Salk et al., 2017), por lo que resulta necesario incorporar los síntomas de la depresión masculina en los criterios diagnósticos, con el fin de favorecer un tratamiento oportuno (Shi et al., 2021).

Se han analizado los instrumentos de evaluación de la depresión como una posible explicación de estas diferencias. Salokangas et al. (2002) encontraron que las mujeres obtenían puntuaciones más altas que los hombres en el Inventario de Depresión de Beck, BDI (Beck et al., 1961), en comparación con las obtenidas en la Escala de Depresión DEPS (Salokangas et al., 1995). Esta diferencia se explicaba por el funcionamiento diferencial de determinados ítems según el género, lo que sugiere la presencia de un sesgo en la medición. Otros estudios con el Inventario de Depresión de Beck-II, BDI-II (Beck et al., 1996), también encontraron puntuaciones superiores en mujeres (Osman et al., 1997; Carmody, 2005; You et al., 2009).

Asimismo, varios estudios han encontrado funcionamiento diferencial de ítems en otros instrumentos de medida, como los trabajos de Yang y Jones (2007), Cole et al. (2000) y Carleton et al. (2013) sobre la Escala de Depresión del Centro de Estudios Epidemiológicos, CES-D (Radloff, 1977), y los realizados por Broekman et al. (2008) y Marc et al. (2008) sobre la Escala de Depresión Geriátrica, GDS-15 (Yesavage & Sheikh, 1986).

Es preciso considerar también la aportación de Hiesinger et al. (2018), quienes señalan que la mayor prevalencia de depresión en mujeres podría explicarse en parte por la medición, dado que los instrumentos captan con mayor frecuencia la sintomatología más típica en mujeres. En relación con los posibles sesgos derivados de la aplicación diferencial de criterios diagnósticos según el género del paciente, Wrobel (1993) y Lewis et al. (2006) comprobaron, mediante estudios con viñetas clínicas, que era más probable diagnosticar depresión cuando la persona descrita era una mujer. Por otra parte, Langer y Abelson (1974) informaron que el proceso diagnóstico puede verse influido por las expectativas de los profesionales respecto a las personas que atienden. A estos datos se suman los aportados por Bruchmüller et al. (2012), quienes encontraron evidencia de que los profesionales no se adhieren estrictamente a los criterios diagnósticos de los sistemas de

clasificación, pudiendo verse afectado su juicio clínico por heurísticas y sesgos.

Esta revisión sistemática tiene como objetivo actualizar y sintetizar la evidencia sobre el sesgo de género en la evaluación y diagnóstico de los trastornos depresivos, con el fin de comprender mejor su impacto en las diferencias de prevalencia entre hombres y mujeres. De este modo, buscamos contribuir a responder una pregunta planteada hace años sobre las causas de la mayor prevalencia del diagnóstico de depresión en mujeres, abordando el riesgo de sesgo de género en los instrumentos de medida más utilizados, y reconociendo que existen otros factores explicativos dada la complejidad de los trastornos depresivos.

Actualizamos la información reciente disponible en el periodo 2010–2022 mediante una síntesis narrativa que describe las características del funcionamiento diferencial de los ítems en los instrumentos de medida de la depresión, los cuales ofrecen posibles vías para comprender la brecha de género en el diagnóstico. Esperamos que esta revisión sistemática resulte útil para los profesionales dedicados al diagnóstico y tratamiento de los trastornos depresivos, y aporte elementos que contribuyan a la creciente corriente de pensamiento que cuestiona el carácter binario de la medición de los trastornos depresivos y promueve la incorporación de enfoques culturalmente sensibles.

Resaltamos la necesidad de aplicar el principio de justicia del Código de Conducta de la Asociación Americana de Psicología en el ámbito del diagnóstico (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), con el fin de garantizar el derecho de todas las personas a evaluaciones y diagnósticos psicológicos de igual calidad, equidad y sin sesgos, independientemente del origen, el estatus social u otras características individuales. Esto implica abordar las disparidades y barreras que dificultan que poblaciones marginadas o vulnerables reciban diagnósticos precisos y oportunos, y asegurar que el proceso diagnóstico —incluida

la formulación de un plan de tratamiento— se realice de manera rigurosa, libre de prejuicios y considerando el contexto cultural y social del individuo.

MÉTODO

Se ha realizado una revisión sistemática de la evidencia reciente sobre sesgo de género en el diagnóstico y la evaluación de los trastornos depresivos, diseñada y realizada siguiendo las pautas de la declaración PRISMA (PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. Page et al..2021).

Tras determinar los criterios de inclusión y exclusión, se diseñaron las estrategias de búsqueda de estudios en las la bases de datos Web of Science (WOS), Scopus y APA PsycInfo. Dos revisores realizaron la selección de los estudios y la extracción de datos y evaluaron el riesgo de sesgo de los artículos seleccionados. En la Tabla 1 figuran los criterios de inclusión y los criterios de exclusión.

Tabla 1  
Criterios de inclusión y exclusión

Estudios incluidos	Estudios excluidos
Artículos científicos	Libros Material editorial Tesis Documentos de trabajo
Estudios empíricos	No aportan hallazgos empíricos -revisiones o metaanálisis-
Tratan sobre diagnóstico y/o evaluación de la depresión	No incluyen información sobre diagnóstico y evaluación de la depresión
Analizan diferencias entre mujeres y hombres en el diagnóstico y/o la evaluación de la depresión	No informan sobre diferencias entre mujeres y hombres en la evaluación y/o el diagnóstico de la depresión. Informan sobre otras diferencias, p. ej. derivadas de la orientación sexual
Evalúan la posibilidad de sesgo de género en el diagnóstico y/o la evaluación de la depresión	No aportan datos sobre la posibilidad de sesgo de género en la evaluación y/o el diagnóstico de la depresión
Revisados por pares	No revisados por pares
Publicados entre 2010 y 2022 (última búsqueda en el 26 de mayo de 2023)	Publicados en otros periodos.

### *Estrategia de búsqueda*

Se llevó a cabo una estrategia de búsqueda sistemática utilizando una fórmula inicial: “gender AND bias AND depres\*” en las bases de datos de literatura científica mencionadas, estableciendo además como requisito la revisión por pares y la publicación entre 2010 y 2022.

La estrategia de búsqueda en WOS se diseñó para localizar artículos que tuvieran en su temática los términos “gender”, “bias” y “depres\*”, utilizando la fórmula de búsqueda “gender (tema) AND bias (tema) AND depres\* (tema)”. En el listado de artículos obtenido se seleccionaron aquellos que contenían los términos “Female” o “Male” o “Mood Disorders”, utilizando la fórmula de búsqueda AND “Female” OR “Male” OR “Mood Disorders”. Se excluyeron los artículos codificados como “Book” o como “Editorial Material” utilizando la fórmula de búsqueda NOT “Book” OR “Editorial Material”.

Se buscaron también artículos que tuvieran en su título el término “gender” y en su temática los términos “bias” y “depres\*”, utilizando la fórmula de búsqueda “gender (título) AND bias (tema) AND depres\* (tema)”. También se buscaron artículos que tuvieran en su título el término “bias” y en su temática los términos “gender” y “depres\*”, utilizando la fórmula de búsqueda “gender (tema) AND bias (título) AND depres\* (tema)”. Finalmente, se buscaron artículos que tuvieran en su título el término “depres\*” y en su temática los términos “gender” y “bias”, utilizando la fórmula de búsqueda “gender (tema) AND bias (tema) AND depres\* (título)”.

La estrategia de búsqueda en SCOPUS se diseñó para localizar artículos que contuvieran los términos “gender”, “bias” y “depres\*” en el título, resumen y palabras clave, utilizando la fórmula de búsqueda “gender AND bias AND depres\*”. En el listado de artículos obtenido se seleccionaron aquellos que contenían los términos “Female” o “Male” o “Depressive disorder” o “Depressive +Symptoms” o “Mood Disorders, utilizando la fórmula de búsqueda “AND Female OR Male OR

Depressive disorder OR Depressive+Symptoms OR Mood Disorders”. Se excluyeron los artículos codificados como “Book”.

La estrategia de búsqueda en APA PsycInfo se diseñó para localizar artículos que contuvieran en su título el término “gender” y que contuvieran los términos “bias” y “depres\*” en su temática. En el listado de artículos obtenido se seleccionaron aquellos que contenían el término “Depressive disorder”. Se excluyeron los artículos codificados como “Book” o “Editorial Material” o “Dissertation/Thesis”. También se buscaron artículos que contuvieran en su título el término “bias” y que contuvieran los términos “gender” y “depres\*” en su temática. En el listado de artículos obtenido se seleccionaron aquellos que contenían el término “Depressive disorder”. Se excluyeron los artículos codificados como “Book” o “Editorial Material” o “Dissertation/Thesis”. Finalmente, se buscaron artículos que contuvieran en su título el término “depres\*” y que contuvieran los términos “gender” y “bias”. En el listado de artículos obtenido se seleccionaron aquellos que contenían el término “Depressive disorder”. Se excluyeron los artículos codificados como “Book” o “Editorial Material” o “Dissertation/Thesis”.

En las búsquedas en WOS se incluyeron las bases de datos Web of Science Core Collection, MEDLINE, Current Contents Connect y SciELO. En las búsquedas en APA PsycInfo se incluyeron las bases de datos: Psychology Database, Publicly Available Content Database, Health & Medical Collection, Nursing & Allied Health Premium y APA PsycArticles.

### *Selección de los estudios*

Dos revisores realizaron la selección de los estudios. Se realizaron sesiones de debate para analizar los hallazgos realizados durante el proceso de selección. Tras la fase inicial de identificación, se consensó en varias sesiones la exclusión de aquellos cuyos títulos no cumplían los criterios de inclusión. A continuación, se consensó en una nueva sesión la selección de artículos tras la lectura independiente de los

resúmenes de los artículos seleccionados en la fase anterior. Finalmente, se realizó una sesión de debate para seleccionar de entre los restantes artículos aquellos cuyos textos completos no cumplieran los criterios de inclusión. Resultaron incluidos 12 artículos, su proceso de selección se describe en la Figura 1.

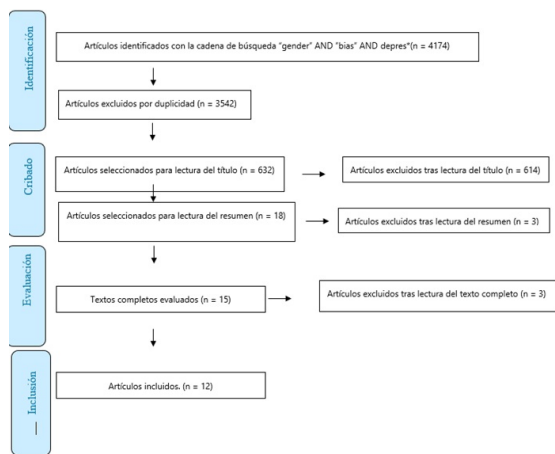


Figura 1.  
Resumen del proceso de selección de artículos

### Extracción de datos

Se codificaron las variables extraídas de cada estudio: país, instrumento de medida, tamaño y características demográficas de la muestra, porcentajes de mujeres y hombres, y modo de administración. Se identificaron y sintetizaron los hallazgos, y se analizaron las características demográficas de los participantes y los métodos de obtención de datos. Teniendo en cuenta que los estudios sobre el funcionamiento diferencial de ítems aportaron información relevante sobre el riesgo de sesgo, se analizaron los ítems con funcionamiento diferencial según el género, en función de la puntuación obtenida y del tipo de población. Asimismo, se examinó la sintomatología asociada con los ítems con funcionamiento diferencial según el género y la edad, así como la incidencia de estos ítems.

### Evaluación de riesgo de sesgo y estrategia de síntesis

Se siguió el procedimiento de *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins et al., 2023), analizando los dominios de proceso de aleatorización, desviaciones de las intervenciones previstas, falta de datos de resultados, medición de los resultados y selección de los resultados informados.

Se empleó un procedimiento de codificación temática, agrupándose los estudios en dos categorías: estudios sobre funcionamiento diferencial de ítems y estudios sobre diferencias de género en puntuaciones obtenidas. Los instrumentos de medida se compararon transversalmente para establecer parámetros relacionados con la sintomatología asociada a los ítems con funcionamiento diferencial. La heterogeneidad estadística de los estudios impidió realizar un metaanálisis, optándose por realizar una síntesis narrativa.

## RESULTADOS

El hallazgo más relevante es el funcionamiento diferencial de 20 ítems en instrumentos de medida de la depresión que se encuentran entre los más frecuentemente utilizados. Las mujeres obtuvieron mayor puntuación en comparación con los hombres en 15 ítems, la mayor parte de ellos en población infantil y adolescente. Los hombres obtuvieron mayor puntuación en 4 ítems, igualmente en su mayoría en población infantil y adolescente; además, presentaron una puntuación menor en un ítem. Estos hallazgos aportan una perspectiva particular sobre el sesgo de género en el diagnóstico y la evaluación de los trastornos depresivos, derivado de la existencia, en los instrumentos de medida, de ítems en los que las mujeres tienden a obtener puntuaciones superiores a las de los hombres.

Es importante tener en cuenta que algunos resultados son contradictorios según el estudio, como ocurre con el PHQ-9: se encontró funcionamiento diferencial en un ítem en un estudio (Thibodeau & Asmundson, 2014) y ausencia de funcionamiento diferencial en otro

(Cameron et al., 2013). Asimismo, se observó funcionamiento diferencial en algunos ítems, con mayor puntuación unas veces en mujeres y otras veces en hombres, según el instrumento utilizado. Esto ocurrió en los ítems de pensamientos o deseos de muerte (mayor puntuación en adolescentes mujeres en el BDI-II C, pero mayor en adolescentes hombres en el CDI) y pesimismo (mayor puntuación en adolescentes mujeres en el CDI, pero mayor en adolescentes hombres en el BDI-II C).

Otros estudios informan de aspectos adicionales de la medición de la depresión que completan el panorama sobre el sesgo de género en este ámbito. El estudio de Berger et al. (2012) mostró que los hombres reportan menos síntomas depresivos que las mujeres en los instrumentos de medida cuando el término depresión se asociaba con una teoría causal incontrolable —esto es, cuando los síntomas se atribuían a factores fuera del control personal, como aspectos biológicos o eventos ambientales—.

Dos estudios adicionales comprobaron la ausencia de evidencia de diferencias de género en relación con el modo de administración —autoevaluación o entrevista— en los instrumentos MINI, PHQ-9 e IDAS-II (Levin-Aspenson & Watson, 2018), así como en función de los procedimientos diagnósticos empleados —listas de verificación frente a juicio clínico abierto— (Cwik et al., 2016). En la Tabla 2 se presenta un resumen de los hallazgos sobre funcionamiento diferencial.

**Tabla 2**

*Funcionamiento diferencial de ítems y puntuación comparativa según género*

Instrumento	Ítem	Puntuación	Población
<b>Hallazgos en mujeres</b>			
BDI-II-C	Disconformidad consigo mismo	Más puntuación	Adolescente
	Llanto	Más puntuación	Adolescente
	Pérdida de interés en otros	Más puntuación	Adolescente
	Pensamientos o deseos de muerte	Más puntuación	Adolescente
CDI	Falta de diversión en la escuela	Más puntuación	Infantil Adolescente
	Alteración del apetito	Más puntuación	Infantil. Adolescente
	Autopercepción negativa	Más puntuación	Infantil. Adolescente
	Culpa	Más puntuación	Infantil. Adolescente
	Falta de amigos	Más puntuación	Infantil. Adolescente
	Fastidio	Más puntuación	Infantil. Adolescente
	Llanto	Más puntuación	Infantil. Adolescente
	Pesimismo	Más puntuación	Infantil. Adolescente
CES-D	Tristeza	Más puntuación	Infantil. Adolescente.
	Llanto	Más puntuación	Adulta
PHQ-9	Alteración del apetito	Más puntuación	Adulta
<b>Hallazgos en hombres</b>			
BDI-II-C	Dificultades de concentración	Más puntuación	Adolescente
	Fracaso	Más puntuación	Adolescente
	Pesimismo	Más puntuación	Adolescente
CDI	Pensamientos o deseos de muerte	Más puntuación	Infantil. Adolescente
GDS-15	Aburrimiento	Menos puntuación	Adulta

Nota: BDI-II-C= Inventario de Depresión de Beck, versión china; CDI = Inventario de Depresión Infantil; CES-D = Escala de Depresión del Centro de Estudios Epidemiológicos; PHQ-9 = Cuestionario de Salud del Paciente; GDS-15 = Escala de Depresión Geriátrica

### *Sintomatología asociada a los ítems con funcionamiento diferencial.*

Las adolescentes tuvieron mayor puntuación en ítems asociados a síntomas de aburrimiento, alteración del apetito, autopercepción negativa, culpa, disconformidad consigo mismo, falta de amigos, fastidio, llanto, pérdida de interés en otros y tristeza. Los adolescentes tuvieron mayor puntuación en ítems asociados a síntomas de dificultades de concentración y sentimiento de fracaso. Los adultos también tuvieron menor puntuación en un ítem asociado a síntomas de aburrimiento. Los síntomas de pensamientos o

deseos de muerte y pesimismo se asocian a mayor puntuación en ambos géneros, en función del instrumento de medida (BDI-II-C o CDI). En la Tabla 3 figura la sintomatología asociada a ítems con funcionamiento diferencial.

**Tabla 3**  
*Sintomatología asociada a ítems con funcionamiento diferencial*

Ítem	Puntuación	Género	Edad	Instrumento de medida
Aburrimiento	Menos puntuación	Hombres	Adulta	GDS-15
Alteración del apetito	Más puntuación	Mujeres	Infancia	PHQ-9
Autopercepción negativa	Más puntuación	Mujeres	Infancia	BDI-II-C
Culpa	Más puntuación	Mujeres	Adolescencia	BDI-II-C
Dificultades de concentración	Más puntuación	Hombres	Adolescencia	BDI-II-C
Disconformidad consigo mismo	Más puntuación	Mujeres	Adolescencia	BDI-II-C
Falta de amigos	Más puntuación	Mujeres	Infancia	BDI-II-C
Falta de diversión en la escuela	Más puntuación	Mujeres	Infancia	BDI-II-C
Fastidio	Más puntuación	Mujeres	Adolescencia	BDI-II-C
Fracaso	Más puntuación	Hombres	Adolescencia	BDI-II-C
Llanto	Más puntuación	Mujeres	Infancia	BDI-II-C
Pensamientos o deseos de muerte	Más puntuación	Hombres	Adolescencia	BDI-II-C
Pérdida de interés en otros	Más puntuación	Mujeres	Adolescencia	BDI-II-C
Pesimismo	Más puntuación	Mujeres	Infancia	BDI-II-C
Pesimismo	Más puntuación	Hombres	Adolescencia	BDI-II-C
Tristeza	Más puntuación	Mujeres	Infancia	BDI-II-C

*Incidencia de ítems con funcionamiento diferencial en los instrumentos de medida.*

El ítem sobre llanto figura en 3 instrumentos de medida: CDI, CES-D y BDI-II-C. Otros ítems figuran en 2 instrumentos de medida: aburrimiento (CDI y GDS-15 -en este caso con menor puntuación-), alteración del apetito (CDI y PHQ-9), pesimismo (CDI y BDI-II-C) y pensamientos o deseos de muerte (BDI-II-C y CDI). En la Tabla 4 figura la sintomatología asociada a ítems con funcionamiento diferencial.

**Tabla 4**  
*Incidencia de ítems con funcionamiento diferencial en instrumentos de medida de la depresión*

Ítem	Instrumento de medida
1. En tres instrumentos de medida	
Llanto	CDI BDI-II-C CES-D
2. En dos instrumentos de medida	
Alteración del apetito	CDI PHQ-9
Pesimismo	CDI BDI-II-C
Pensamientos o deseos de muerte	BDI-II-C CDI
3. En un instrumento de medida	
Aburrimiento en la escuela	CDI
Aburrimiento en centros geriátricos	GDS-15 Menor puntuación
Autopercepción negativa	CDI
Culpa	CDI
Dificultades de concentración	BDI-II-C
Disconformidad consigo mismo	BDI-II-C
Falta de amigos	CDI
Fastidio	CDI
Fracaso	BDI-II-C
Pérdida de interés en otros	BDI-II-C
Tristeza	CDI

Nota: BDI-II-C= Inventario de Depresión de Beck, versión china; CDI = Inventario de Depresión Infantil; CES-D = Escala de Depresión del Centro de Estudios Epidemiológicos; PHQ-9 = Cuestionario de Salud del Paciente; GDS-15 = Escala de Depresión Geriátrica

La Tabla 5 presenta la incidencia de ítems con funcionamiento diferencial en los instrumentos de medida.

**Tabla 5**  
*Instrumentos de medida analizados*

Instrumento	Finalidad	Tipo	Administración	Ítems	Población
BDI-II C	Evaluar. gravedad	Inventario	Autoinforme	21	Adolescente adulta
CDI	Evaluar. síntomas	Inventario	Autoinforme	28	Infantil adolescente
				20	
CES-D	Cribado	Escala	Autoinforme	10	A partir de seis años
				12	
GDS-15	Cribado	Escala	Entrevista	15	A partir de 60 años
HADS	Cribado	Escala	Entrevista	7	Adulta
IDAS-II	Evaluar. síntomas	Escala	Autoinforme	99	Adulta
MINI	Evaluar. síntomas	Entrevista	Entrevista	174	Adulta
PHQ-9	Diagnóstico	Cuestionario	Entrevista	9	Adulta

Nota: BDI-II = Inventario de Depresión de Beck; CDI = Inventario de Depresión Infantil; CES-D = Escala de Depresión del Centro de Estudios Epidemiológicos; GDS-15 = Escala de Depresión Geriátrica; HADS = Escala de Ansiedad y Depresión Hospitalaria; IDAS-II = Inventario de Síntomas de Depresión y Ansiedad; MINI = Mini Entrevista Neuropsiquiátrica Internacional; PHQ-9 = Cuestionario de Salud del Paciente.

### *Instrumentos de medida analizados*

Se analizaron 8 instrumentos de medida de la depresión: Inventario de Depresión de Beck-II (BDI-II) (Beck et al., 1996), Inventario de depresión infantil CDI (Kovacs, 1992), Escala de Depresión del Centro de Estudios Epidemiológicos CES-D (Radloff, 1977), Escala de depresión geriátrica GDS-15 (Yesavage y Sheikh, 1986), Escala Hospitalaria de Ansiedad y Depresión HADS (Zigmond y Snaith, 1983), Inventario de Síntomas de Depresión y Ansiedad IDAS-II (Watson et al., 2012), Mini Entrevista Neuropsiquiátrica Internacional MINI (Sheehan et al., 1998) y Cuestionario de Salud del Paciente PHQ-9 (Kroenke et al., 2001).

Se encuentran entre los más usados para el diagnóstico y evaluación de la depresión, por lo que los datos obtenidos ofrecen una amplia panorámica sobre su funcionamiento según se apliquen a mujeres u hombres. En la tabla 6 figura

una descripción de las características de estos instrumentos de medida.

**Tabla 6**  
*Estudios sobre funcionamiento diferencial de ítems*

Estudio	País	Instrumento	Muestra	% mujeres	% hombres	Administración
Cameron et al. (2013)	Reino Unido	HADS PHQ-9	N = 1.087 Adultos Atención Primaria	72%	28%	Entrevista
Chiesi et al. (2018)	Italia	GDS-15	N = 1.302 Personas mayores Clínicas geriátricas	59%	41%	Entrevista
Fong & Ho (2014)	Hong Kong	HADS	N = 753 Adultos Comunitaria	57 %	43 %	Entrevista
Goodwill (2021)	EE. UU.	CES-D 12	N = 3.570 Adultos Encuesta nacional	64%	36%	Entrevista
Suh et al. (2017)	EE. UU.	CES-D 10	N = 468 Universitarios	33%	67%	Cuestionario en línea
Thibodeau & Asmundson (2014)	EE. UU.	PHQ-9	N = 5.996 Niños y adultos Encuesta nacional	51%	49%	Entrevista
Van Beek et al. al. (2012)	Países Bajos	CDI	N = 4.048 Escolares	52%	48%	Autoinforme
Wu & Huang (2014)	China	BDI-II C	N = 827 Escolares	50%	50%	Autoinforme
Zhu & Chiu (2021)	China	CES-D	N = 34.762 Niños, adolescentes y adultos Encuesta nacional	51% 1% sin datos	48%	Entrevista

Nota: BDI-II C = Inventario de Depresión de Beck, versión china; CDI = Inventario de Depresión Infantil; CES-D = Escala de Depresión del Centro de Estudios Epidemiológicos; GDS-15 = Escala de Depresión Geriátrica; HADS = Escala Hospitalaria de Ansiedad y Depresión; IDAS-II = Inventario de Síntomas de Depresión y Ansiedad; MINI = Mini Entrevista Neuropsiquiátrica Internacional; PHQ-9 = Cuestionario de Salud del Paciente

### *Características de los estudios*

Predominan los realizados en E.E. U.U. (cinco de los 12 estudios, el 42%); otros estudios se realizaron en Europa (Alemania, Italia, Países Bajos y Reino Unido) y en Asia (China y Hong Kong). No constan estudios realizados en amplias zonas del mundo (África, América -excepto E.E. U.U.-, Australia y Asia -excepto China y Hong Kong). Contienen datos de 54.195 participantes. La mayoría de los datos proceden de tres estudios a partir de encuestas nacionales (44.328); el resto pertenecían a colectivos específicos (universitarios, escolares, psicoterapeutas, pacientes de atención primaria, personas mayores en clínicas geriátricas). Hay



predominio de los datos de mujeres (53%); en 5 estudios superan el 60%. La mayoría de los datos se obtuvieron mediante entrevista (47.910 personas, el 88%).

En las tablas 7 y 8 se ofrece un resumen de los datos más relevantes.

**Tabla 7**  
*Estudios sobre diferencias de género en puntuaciones obtenidas según etiquetas diagnósticas y teorías causales de la depresión empleadas, así como según el modo de administración.*

Estudio	País	Instrumento	Muestra	% mujeres	% hombres	Administración
Berger et al. (2012)	EE. UU.	CES-D	N = 467 Adultos Comunitaria Conveniencia	70%	30%	Cuestionario en línea
Levin-Aspenson & Watson (2018)	EE. UU.	MINI PHQ-9 IDAS-II	N = 440 Adultos Comunitaria	68 %	32 %	Entrevista Autoinforme

Nota: BDI-II = Inventario de Depresión de Beck; : BDI-II C= Inventario de Depresión de Beck, versión china; CDI = Inventario de Depresión Infantil; CES-D = Escala de Depresión del Centro de Estudios Epidemiológicos; CES-D 10 = Escala de Depresión del Centro de Estudios Epidemiológicos , versión abreviada de 10 ítems; CES-D 12 = Escala de Depresión del Centro de Estudios Epidemiológicos , versión abreviada de 12 ítems; CMNI-22 = Inventario de Conformidad a Normas Masculinas; GDS-15 = Escala de Depresión Geriátrica; HADS = Escala de Ansiedad y Depresión Hospitalaria; IDAS-II = Inventario de Síntomas de Depresión y Ansiedad; MINI = Mini Entrevista Neuropsiquiátrica Internacional; PHQ-9 = Cuestionario de Salud del Paciente.

**Tabla 8**  
*Evaluación de riesgo de sesgo.*

Estudios	Dominios				
	Aleatorización	Desviaciones de las intervenciones previstas	Falta de datos de resultados	Medición del resultado	Selección del resultado informado
Berger et al. (2012)	Poco claro	Bajo	Poco claro	Al menos, "Algunas preocupaciones"	Bajo
Cameron et al. (2013)	Poco claro	Bajo	"Algunas preocupaciones"	"Algunas preocupaciones"	"Algunas preocupaciones"
Chiesi et al. (2017)	Poco claro	Poco claro	Poco claro	"Algunas preocupaciones"	Bajo
Cwik et al. (2016)	Poco claro	Bajo	"Algunas preocupaciones"	"Algunas preocupaciones"	Bajo
Fong y Ho (2014)	Poco claro	Bajo	Poco claro	"Algunas preocupaciones"	Bajo
Goodwill (2021)	Poco claro	Bajo	"Algunas preocupaciones"	"Algunas preocupaciones"	Bajo
Levin-Aspenson y Watson (2018)	Poco claro	"Algunas preocupaciones"	Poco claro	"Algunas preocupaciones"	Bajo
Suh et al. (2017)	Poco claro	Bajo	Poco claro	"Algunas preocupaciones"	Bajo
Thibodeau y Asmundson (2014)	Poco claro	Bajo	Poco claro	"Algunas preocupaciones"	Bajo
van Beek et al. (2012)	Poco claro	Bajo	Poco claro	"Algunas preocupaciones"	Bajo
Wu & Huang (2014)	Poco claro	Bajo	Bajo	"Algunas preocupaciones"	Bajo
Zhu & Chiu (2021)	Poco claro	"Algunas preocupaciones"	Bajo	"Algunas preocupaciones"	Bajo

En torno a los estudios sobre diferencias de género en función de los procedimientos diagnósticos, se halló el trabajo de Cwik et al. (2016), realizado en Alemania, en el cual se emplearon listas de verificación junto con juicio clínico abierto como procedimiento diagnóstico. La muestra estuvo conformada por 475 psicoterapeutas, de los cuales 69 % eran mujeres y 31 % hombres, y la administración del estudio se llevó a cabo mediante un cuestionario en línea.

El análisis del riesgo de sesgo muestra que los estudios presentan un riesgo poco claro derivado del proceso de aleatorización, ya que no informan cómo fueron asignados los participantes a los grupos según género ni describen medidas orientadas a garantizar la aleatoriedad. En la mayoría de los estudios se observa un riesgo bajo de sesgo por desviaciones de las intervenciones previstas y por datos faltantes. La medición de los resultados podría estar influenciada por el hecho de que son los propios participantes quienes

los reportan, lo que implica un juicio de, al menos, “algunas preocupaciones” respecto del riesgo de sesgo asociado a esta causa. Asimismo, se identifica un riesgo bajo de sesgo por la selección del resultado.

## DISCUSIÓN

Se ha analizado y sintetizado la evidencia sobre funcionamiento diferencial según el género en ítems de seis de ocho instrumentos de medida que se encuentran entre los más frecuentemente usados. El funcionamiento diferencial condiciona riesgo de sesgo de género: las mujeres, especialmente niñas y adolescentes, obtuvieron mayor puntuación que los hombres en 15 ítems, lo que puede propiciar que sean diagnosticadas de trastornos depresivos con mayor frecuencia. Hay riesgo de sesgo de género para los hombres, aunque en menor número de ítems. Tuvo mayor incidencia el ítem sobre llanto, con funcionamiento diferencial en tres instrumentos de medida.

Es reseñable el hallazgo de funcionamiento diferencial en los adolescentes o en las adolescentes, en función del instrumento de medida utilizado. Esto ocurre en los ítems de pensamientos o deseos de muerte (en los adolescentes con BDI-II C y en las adolescentes con CDI) y pesimismo (en las adolescentes en CDI y en los adolescentes en BDI-II C). En PHQ-9, un estudio (Cameron et al., 2013) encontró invarianza de medida, mientras que otro (Thibodeau & Asmundson, 2014) encontró funcionamiento diferencial, lo que pone de manifiesto el carácter controvertido del asunto.

No conocemos estudios sobre cómo esa combinación de riesgos de sesgo en mujeres y hombres puede contribuir a la disparidad de las tasas de prevalencia. Tampoco conocemos estudios sobre cómo esa combinación de riesgos de sesgo pudiera interactuar con la perspectiva del desarrollo, dada la mayor presencia de ítems con funcionamiento diferencial en la infancia y en la adolescencia (Van Beek et al., 2012).

Estos hallazgos pueden contribuir a un mejor conocimiento de las características diferenciales

de la sintomatología depresiva en mujeres y hombres. Los síntomas de llanto, aburrimiento, autopercepción negativa, alteración del apetito, culpa, disconformidad consigo mismo, falta de amigos, fastidio, pérdida de interés en otros y tristeza se asocian a ítems en los que obtienen mayores puntuaciones las adolescentes. Los adolescentes tuvieron mayor puntuación en ítems asociados a síntomas de dificultades de concentración y sentimiento de fracaso. Por su parte Berger et al. (2012) mostraron que los hombres informan menos síntomas en los instrumentos de medida de la depresión cuando el término depresión se usa junto con una teoría causal que atribuye los síntomas a factores fuera del control personal (p. ej., biología o eventos ambientales).

Enfatizamos que los estudios incluidos en nuestra revisión, aun incluyendo análisis de funcionamiento diferencial, no abordan medidas específicas para garantizar la equidad en el diagnóstico de la depresión desde la perspectiva de género. Además, ni los estudios revisados ni los instrumentos de medida de la depresión analizados incorporan una conceptualización de género con perspectiva no binaria y no hacen un reconocimiento explícito del espectro identitario y la influencia cultural en la expresión del malestar psicológico, lo que contrasta con los enfoques actuales sobre diversidad y espectro identitario. Una revisión sistemática de estudios sobre evaluación psicológica publicados recientemente (Rivera, 2025) sostiene que, aun incluyendo análisis de funcionamiento diferencial, son escasos los que estudios que ofrecen datos por género y medidas para garantizar la equidad desde esta perspectiva.

Han aparecido importantes retos para el concepto binario de género en los últimos años (Hyde et al., 2018), entre otros que la tendencia a ver el género / sexo como una categoría binaria está determinada culturalmente y es modificable, como sugiere la investigación sobre las personas transgénero y no binarias.

### *Implicaciones teóricas y epistemológicas del sesgo de género en la medición de la depresión*

El funcionamiento diferencial en ítems en los que las mujeres obtuvieron mayor puntuación, especialmente ítems de llanto y culpa en niñas y adolescentes, no solo puede propiciar que sean diagnosticadas de trastornos depresivos con mayor frecuencia, sino que refleja desigualdades estructurales en la producción del conocimiento psicológico y en la concepción normativa de los afectos que pueden reproducir narrativas históricas de vulnerabilidad femenina. Existe un amplio debate sobre la construcción social del malestar y la medicalización del sufrimiento, al que no son ajenas las críticas feministas al diagnóstico psiquiátrico. Un aspecto especialmente relevante es la evidencia de la influencia cultural en la definición como patológico del sufrimiento (Bacigalupe et al., 2020), que conduce al desarrollo de procesos de medicalización de la salud mental y sobrediagnóstico en las mujeres.

Por otra parte, la menor expresión emocional en hombres puede asociarse con normas de masculinidad hegemónica relacionadas con la reticencia a buscar ayuda psicológica (Mahalik, Di Bianca, & Harris, 2022; Mokhwelepa & Sumbane, 2025) y con la ocultación de síntomas (Smith et al., 2016), lo que podría explicar sus puntuaciones menores. Los hallazgos encontrados en los estudios incluidos en esta revisión sistemática evidencian la ineludible necesidad de reflexionar sobre cómo los sesgos de género en los instrumentos de medida reproducen estereotipos culturales y sus consecuencias en la detección, el diagnóstico y la intervención en los trastornos depresivos, así como la necesidad de incorporarlos a las teorías sobre la depresión y a la epistemología de la evaluación psicológica. Enfatizamos la necesidad de una reflexión sobre la dimensión ética y política de la evaluación psicológica de los trastornos depresivos y subrayamos la responsabilidad científica de revisar las herramientas diagnósticas a la luz de los principios de equidad.

### **Limitaciones**

Los estudios incluidos en esta revisión sistemática presentan limitaciones. La heterogeneidad en las muestras de participantes y en las condiciones de aplicación, así como la ausencia de estudios en amplias zonas geográficas (África, América -excepto EE. UU.-, Europa -excepto Alemania, Italia y Países Bajos- y Asia -excepto República Popular China y Hong Kong-) limitan la posibilidad de acumular y generalizar evidencias. Se debe tener en cuenta, además, la evidencia de riesgo poco claro de sesgo derivado del proceso de aleatorización y desequilibrio en la distribución de los participantes en las muestras según género.

### **Conclusiones**

Los hallazgos evidencian que existen 20 ítems con funcionamiento diferencial en función del género de los participantes en seis de los instrumentos de medida de la depresión más frecuentemente usados, lo que implica un riesgo de sesgo de género. El funcionamiento diferencial afecta a 15 ítems en las respuestas de mujeres, cuyas puntuaciones fueron mayores comparadas con las obtenidas por hombres y a cuatro ítems en las respuestas de hombres. En un solo ítem los hombres tuvieron menor puntuación.

Es preciso tomar conciencia del riesgo de que se produzcan falsos diagnósticos que darían lugar al etiquetado patológico depresivo en mujeres. Enfatizamos que los sesgos de género identificados en la medida de la depresión constituyen no solo un error de medida, sino que reflejan un fenómeno epistemológico más amplio que implica valores normativos sobre la emocionalidad y la normalidad. Futuras líneas de investigación deberían reconsideren las categorías diagnósticas desde una perspectiva crítica, promoviendo un lenguaje diagnóstico no androcéntrico. La revisión de los instrumentos de medida de la depresión debería ser parte de una agenda de investigación crítica orientada a la justicia diagnóstica y la erradicación de la injusticia epistémica en salud.

## REFERENCIAS

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. [https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards\\_2014edition.pdf](https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf)
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425787>
- Bacigalupe, A., Cabezas, A., Bueno, M. B., & Martín, U. (2020). El género como determinante de la salud mental y su medicalización. Informe SESPAS 2020. *Gaceta Sanitaria*, 34, 61–67. <https://doi.org/10.1016/j.gaceta.2020.06.013>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory–II manual*. The Psychological Corporation.
- 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. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
- Berger, J. L., Addis, M. E., Reilly, E. D., Syzdek, M. R., & Green, J. D. (2012). Effects of gender, diagnostic labels, and causal theories on willingness to report symptoms of depression. *Journal of Social and Clinical Psychology*, 31(5), 439–457. <https://doi.org/10.1521/jscp.2012.31.5.439>
- Broekman, B. F. P., Nyunt, S. Z., Niti, M., Jin, A. Z., Ko, S. M., Kumar, R., Fones, C. S. L., & Ng, T. P. (2008). Differential item functioning of the Geriatric Depression Scale in an Asian population. *Journal of Affective Disorders*, 108(3), 285–290. <https://doi.org/10.1016/j.jad.2007.10.005>
- Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of Consulting and Clinical Psychology*, 80(1), 128–138. <https://doi.org/10.1037/a0026582>
- Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2013). Differential item functioning of the HADS and PHQ-9: An investigation of age, gender and educational background in a clinical UK primary care sample. *Journal of Affective Disorders*, 147(1–3), 262–268. <https://doi.org/10.1016/j.jad.2012.11.015>
- Carleton, R. N., Thibodeau, M. A., Teale, M. J. N., Welch, P. G., Abrams, M. P., Robinson, T., & Asmundson, G. J. (2013). The Center for Epidemiologic Studies Depression Scale: A review with theoretical and empirical evaluation of item content and factor structure. *PLoS ONE*, 8(3), e58067. <https://doi.org/10.1371/journal.pone.0058067>
- Carmody, D. P. (2005). Psychometric characteristics of the Beck Depression Inventory-II with college students of diverse ethnicity. *International Journal of Psychiatry in Clinical Practice*, 9(1), 22–28. <https://doi.org/10.1080/13651500510014800>
- Cole, S. R., Kawachi, I., Maller, S. J., & Berkman, L. F. (2000). Test of item-response bias in the CES–D Scale. *Journal of Clinical Epidemiology*, 53(3), 285–289. [https://doi.org/10.1016/S0895-4356\(99\)00151-1](https://doi.org/10.1016/S0895-4356(99)00151-1)
- Cwik, J. C., Papen, F., Lemke, J. E., & Margraf, J. (2016). An investigation of diagnostic accuracy and confidence associated with diagnostic checklists as well as gender biases in relation to mental disorders. *Frontiers in Psychology*, 7, 1813. <https://doi.org/10.3389/fpsyg.2016.01813>
- Hiesinger, K., Tophoven, S., & March, S. (2018). Geschlechtsspezifische Verzerrungen bei der Erfassung von Depressivität. *Prävention und Gesundheitsförderung*, 13(3), 211–217. <https://doi.org/10.1007/s11553-017-0634-x>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2023). *Cochrane handbook for systematic reviews of interventions* (Version

- 6.4). Cochrane. <https://training.cochrane.org/handbook>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & Van Anders, S. M. (2018). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171–193. <https://doi.org/10.1037/amp0000307>
- Iriarte, A. D. V., & Medina, J. (2024). Cross-national prevalence of risk of depression among older adults: Are there gender and age differences? *Innovation in Aging*, 8(Suppl. 1), 355–356. <https://doi.org/10.1093/geroni/igae098.1159>
- Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G., & Nelson, C. B. (1993). Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity and recurrence. *Journal of Affective Disorders*, 29(2–3), 85–96. [https://doi.org/10.1016/0165-0327\(93\)90026-G](https://doi.org/10.1016/0165-0327(93)90026-G)
- Kovacs, M. (1992). *Children's Depression Inventory (CDI) manual*. Multi-Health Systems.
- 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>
- Langer, E. J., & Abelson, R. P. (1974). A patient by any other name: Clinician group difference in labeling bias. *Journal of Consulting and Clinical Psychology*, 42(1), 4–9. <https://doi.org/10.1037/h0036054>
- Levin-Aspenson, H. F., & Watson, D. (2018). Mode of administration effects in psychopathology assessment: Analyses of gender, age, and education differences in self-rated versus interview-based depression. *Psychological Assessment*, 30(3), 287–299. <https://doi.org/10.1037/pas0000474>
- Lewis, R., Lamdan, R. M., Wald, D., & Curtis, M. (2006). Gender bias in the diagnosis of a geriatric standardized patient: A potential confounding variable. *Academic Psychiatry*, 30(5), 392–396. <https://doi.org/10.1176/ap.30.5.392>
- Mahalik, J. R., Di Bianca, M., & Harris, M. P. (2022). Conformity to masculine norms and men's responses to the COVID-19 pandemic. *Psychology of Men & Masculinity*, 23(4), 445–449. <https://doi.org/10.1037/me n0000401>
- Marc, L. G., Raue, P. J., & Bruce, M. L. (2008). Screening performance of the 15-item Geriatric Depression Scale in a diverse elderly home care population. *The American Journal of Geriatric Psychiatry*, 16(11), 914–921. <https://doi.org/10.1097/JGP.0b013e318186bd67>
- Mokhwelepa, L. W., & Sumbane, G. O. (2025). Men's mental health matters: The impact of traditional masculinity norms on men's willingness to seek mental health support—A systematic review of literature. *American Journal of Men's Health*, 19(3). <https://doi.org/10.1177/15579883251321670>
- Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, 101(2), 259–282. <https://doi.org/10.1037/0033-2909.101.2.259>
- Nolen-Hoeksema, S., Larson, J., & Grayson, C. (1999). Explaining the gender difference in depressive symptoms. *Journal of Personality and Social Psychology*, 77(5), 1061–1072. <https://doi.org/10.1037/0022-3514.77.5.1061>
- Osman, A., Downs, W. R., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiro, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *Journal of Psychopathology and Behavioral Assessment*, 19(4), 359–376. <https://doi.org/10.1007/BF02229026>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., ... McKenzie, J. E. (2021). PRISMA 2020 Explanation and Elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <https://doi.org/10.1136/bmj.n160>
- 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>
- Rivera, F. (2025). Equidad de los test psicológicos desde una perspectiva de género: Análisis de buenas prácticas en psicometría. *Apuntes de Psicología*, 43(1), 107–120. <https://doi.org/10.70478/apuntes.psi.2025.43.10>
- Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, 143(8), 783–822. <https://doi.org/10.1037/bul0000102>
- Salokangas, R. K., Poutanen, O., & Stengård, E. (1995). Screening for depression in primary care: Development and validation of the Depression Scale. *Acta Psychiatrica Scandinavica*, 92(1), 10–16. <https://doi.org/10.1111/j.1600-0447.1995.tb09536.x>
- Salokangas, R. K., Vaahtera, K., Paciriev, S., Sohlman, B., & Lehtinen, V. (2002). Gender differences in depressive symptoms: An artefact caused by measurement instruments? *Journal of Affective Disorders*, 68(2–3), 215–220. [https://doi.org/10.1016/S0165-0327\(00\)00315-3](https://doi.org/10.1016/S0165-0327(00)00315-3)
- Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., Demyttenaere, K., de Girolamo, G., Haro, J. M., Jin, R., Karam, E. G., Kovess-Masfety, V., ... Kessler, R. C. (2009). Cross-national associations between gender and mental disorders in the WHO World Mental Health Surveys. *Archives of General Psychiatry*, 66(7), 785–795. <https://doi.org/10.1001/archgenpsychiatry.2009.36>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(20), 22–33.
- Shi, P., Yang, A., Zhao, Q., Chen, Z., Ren, X., & Dai, Q. (2021). A hypothesis of gender differences in self-reporting symptoms of depression: Implications to solve under-diagnosis and under-treatment of depression in males. *Frontiers in Psychiatry*, 12, 589687. <https://doi.org/10.3389/fpsyt.2021.589687>
- Smith, D. T., Mouzon, D. M., & Elliott, M. (2016). Reviewing the assumptions about men's mental health: An exploration of the gender binary. *American Journal of Men's Health*, 12(1), 78–89. <https://doi.org/10.1177/1557988316630953>
- Suh, H., van Nuenen, M., & Rice, K. G. (2017). The CES-D as a measure of psychological distress among international students: Measurement and structural invariance across gender. *Assessment*, 24(7), 896–906. <https://doi.org/10.1177/1073191116632337>
- Thibodeau, M. A., & Asmundson, G. J. (2014). The PHQ-9 assesses depression similarly in men and women from the general population. *Personality and Individual Differences*, 56, 149–153. <https://doi.org/10.1016/j.paid.2013.08.039>
- Van Beek, Y., Hessen, D. J., Hutteman, R., Verhulp, E. E., & Van Leuven, M. (2012). Age and gender differences in depression across adolescence: Real or 'bias'? *Journal of Child Psychology and Psychiatry*, 53(9), 973–985. <https://doi.org/10.1111/j.1469-7610.2012.02553.x>
- Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., Stasik, S. M., & Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*, 19(4), 399–420. <https://doi.org/10.1177/1073191112449857>
- Wrobel, N. H. (1993). Effect of patient age and gender on clinical decisions. *Professional Psychology: Research and Practice*, 24(2), 206–213. <https://doi.org/10.1037/0735-7028.24.2.206>
- Yang, F. M., & Jones, R. N. (2007). CES-D item response bias found with Mantel-Haenszel method successfully replicated using latent variable modeling. *Journal*

- of Clinical Epidemiology*, 60(11), 1195–1200. <https://doi.org/10.1016/j.jclinepi.2007.02.008>
- Yesavage, J. A., & Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter scale. *Clinical Gerontologist*, 5(1–2), 165–173. [https://doi.org/10.1300/J018v05n01\\_09](https://doi.org/10.1300/J018v05n01_09)
- You, S., Merritt, R. D., & Conner, K. R. (2009). Do gender differences in the role of dysfunctional attitudes in depressive symptoms depend on depression history? *Personality and Individual Differences*, 46(2), 218–223. <https://doi.org/10.1016/j.paid.2008.10.002>
- Zhu, J., & Chiu, M. M. (2021). Gender- and age-bias in CES-D when measuring depression in China: A Rasch analysis. *Current Psychology*, 42(10), 8186–8196. <https://doi.org/10.1007/s12144-021-01991-2>
- Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6), 361–370. <https://doi.org/10.1111/j.1600-0447.1983.tb09716.x>

## Notas

\* Artículo de investigación

**Declaración de intereses** No existe ningún interés competitivo de los autores de la revisión.