



# 50 AÑOS DE LOS INVENTARIOS DE DEPRESIÓN DE BECK: CONSEJOS PARA LA UTILIZACIÓN DE LA ADAPTACIÓN ESPAÑOLA DEL BDI-II EN LA PRÁCTICA CLÍNICA

Jesús Sanz

Universidad Complutense de Madrid

En 2011 se publicó la primera adaptación española del Inventario de Depresión de Beck-II (BDI-II), justo cuando se cumplía el 50º aniversario de la publicación de su primera edición. En este tiempo, el BDI se ha convertido en el cuestionario autoaplicado más utilizado en España y en el mundo para evaluar la gravedad de la depresión. En este trabajo se presentan las características básicas del BDI-II en relación con sus versiones anteriores (BDI-I y BDI-IA) y el proceso de su adaptación a la población española, se resumen las propiedades psicométricas de dicha adaptación y se discute su utilización en la práctica clínica, proponiéndose pautas y puntuaciones de corte para la evaluación de la gravedad de la depresión, la evaluación de la significación clínica de los cambios terapéuticos, el cribado de personas con depresión y la ayuda en el diagnóstico diferencial de los trastornos depresivos.

**Palabras clave:** Depresión, Evaluación, BDI-II, Significación clínica, Cribado, Diagnóstico.

The first Spanish adaptation of the Beck Depression Inventory-II (BDI-II) was published in 2011. This year marks the 50<sup>th</sup> anniversary of the publication of its first edition. In this time, the BDI has become the most widely used self-report instrument for assessing the severity of depression in Spain and worldwide. This paper presents the basic characteristics of the BDI-II in relation with its previous versions (BDI-I and BDI-IA), describes the process of its adaptation to the Spanish population, summarizes the psychometric properties of this adaptation, and discusses its use in the clinical practice. To this use, guidelines and cut-off scores are proposed for measuring the severity of depression, for evaluating the clinical significance of therapeutic change, for screening for depression, and for aiding in the differential diagnosis of depressive disorders.

**Key words:** Depression, Assessment, BDI-II, Clinical significance, Screening, Diagnosis.

**E**n 2011 se publicó la primera adaptación española del Inventario de Depresión de Beck-II (Beck, Steer y Brown, 2011), internacionalmente conocido por el acrónimo de su nombre original en inglés: BDI-II (*Beck Depression Inventory-Second Edition*; Beck, Steer y Brown, 1996). Curiosamente, ese año se cumplía el 50º aniversario de la publicación de su primera edición (Beck, Ward, Mendelson, Mock y Erbaugh, 1961), efeméride que consagra al BDI como un clásico entre los instrumentos de evaluación de la depresión y que también refleja su adaptación a las concepciones más actuales de este constructo. En el presente trabajo se presentan las características básicas del BDI-II, se resumen su proceso de adaptación a la población española y las propiedades psicométricas de esta adaptación y se discute su utilización en la práctica clínica para la evaluación del progreso terapéutico y el cribado y diagnóstico de la depresión.

## 1961-2011: 50 AÑOS DE LOS INVENTARIOS DE DEPRESIÓN DE BECK

En los 50 años transcurridos desde su primera publicación, el BDI, tanto en su versión original (BDI-I) como en su versión revisada (BDI-IA; Beck, Rush, Shaw y Emery, 1979) y en su segunda edición (BDI-II), se ha convertido en el instrumento para evaluar la depresión más utilizado en el mundo y en España. Los datos bibliográficos de PsycINFO y Psycodoc (véase la Tabla 1) indican que el número de trabajos publicados en los últimos años sobre el BDI supera con creces a los publicados sobre la Escala de Valoración de la Depresión de Hamilton (HRSD; Hamilton, 1960), la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D; Radloff, 1977) o la Escala Autoaplicada para la Depresión de Zung (SDS; Zung, 1965), tres de los instrumentos con mayor reputación en el ámbito de la depresión (Nezu, Nezu, Friedman y Lee, 2009). Es más, recientemente, Muñiz y Fernández-Hermida (2010) encontraron que el BDI es el quinto test más utilizado por los psicólogos españoles en su práctica diaria, consolidándose en la misma posición que tenía diez años atrás (Muñiz y Fernández-Hermida, 2000), lo cual no debería sorprender ya que cuenta con

Correspondencia: Jesús Sanz. Facultad de Psicología. Universidad Complutense de Madrid. Campus de Somosaguas. 28223 Pozuelo de Alarcón (Madrid). España. E-mail: jsanz@psi.ucm.es



traducciones castellanas validadas en España para todas sus versiones: BDI-I (Conde y Useros, 1974), BDI-IA (Vázquez y Sanz, 1997) y BDI-II (Beck et al., 2011).

Es difícil determinar las razones de esta popularidad. El hecho de que el BDI original se publicara como anexo de un artículo en una revista científica de amplia audiencia y reputación y de que esa versión pudiera utilizarse gratuitamente quizás contribuyó inicialmente a su popularidad. Sin embargo, las dos versiones más recientes (BDI-IA y BDI-II) tuvieron, desde sus inicios, todos los derechos reservados y fueron comercializadas por una editorial. Además, otros instrumentos autoaplicados para la depresión como la SDS tienen una antigüedad similar, fueron publicados originalmente en la misma revista o en otras de audiencia y reputación parecidas y también podían utilizarse de manera gratuita, e incluso aún pueden, y, sin embargo, no han alcanzado la popularidad del BDI. Otras razones para su popularidad residen en el creciente prestigio que fue adquiriendo su creador, Aaron Beck, y la progresiva difusión y éxito de su teoría y terapia cognitivas de la depresión. El BDI fue adoptado como el instrumento estándar en los numerosos estudios diseñados para comprobar los modelos cognitivos de la depresión, incluido la teoría de Beck, que dominaron el campo de las teorías psicológicas de este trastorno durante los años 80 y 90 del siglo XX, y aún siguen haciéndolo (Sanz y Vázquez, 2008; Tennen, Hall y Affleck, 1995), así como en los abundantes estudios realizados sobre la eficacia para la depresión de la terapia cognitiva de Beck y, en general, de la psicoterapia, estudios que también empezaron a proliferar en esas décadas y han aumentado constantemente hasta la actualidad

(Cuijpers, Li, Hofmann y Andersson, 2010). Finalmente, las propias características del BDI (p. ej., brevedad, fácil aplicación, corrección e interpretación) y sus buenas propiedades psicométricas también han contribuido a su popularidad, aunque quizás solo parcialmente, ya que hay otros instrumentos como la CES-D o la SDS que tienen características y propiedades psicométricas parecidas y no son, sin embargo, tan populares.

### LAS DIFERENTES VERSIONES DEL BDI: BDI-I, BDI-IA Y BDI-II

Muchas de las características del BDI se han mantenido constantes en sus tres versiones (véase la Tabla 2). Por ejemplo, todas contienen 21 ítems, cada uno de ellos compuesto por varias afirmaciones sobre un mismo síntoma depresivo que se presentan ordenadas de menor a mayor gravedad y entre las cuales la persona evaluada debe escoger la que mejor describa su estado; además, en cada ítem las diferentes afirmaciones se puntúan de 0 a 3, por lo que el rango de puntuación del inventario es el mismo, 0 a 63. Sin embargo, las tres versiones también presentan diferencias importantes en su contenido (p. ej., marco temporal de sus instrucciones, contenido de sus ítems, número de alternativas de cada ítem) que han dado lugar a diferencias en la interpretación de sus puntuaciones (p. ej., las puntuaciones de corte para distintos niveles de gravedad depresiva son ligeramente más altas en el BDI-II que en el BDI-I) o en sus propiedades psicométricas (p. ej., los índices de consistencia interna y validez factorial del BDI-II son superiores a los del BDI-I y BDI-IA; Beck et al., 2011; Botella y Ponte, 2011; Sanz y García-Vera, 2009) (véase la Tabla 2).

En cuanto al BDI-II, esta versión presenta modificaciones sustanciales respecto al BDI-I y BDI-IA encaminadas a cubrir todos los criterios diagnósticos sintomáticos de los trastornos depresivos según el DSM-IV (APA, 1995) (véase la Tabla 2). En concreto, las instrucciones del BDI-II preguntan por el estado de la persona evaluada durante las dos últimas semanas en lugar de preguntar por la última semana (BDI-IA) o por el mismo día en que se completa el instrumento (BDI-I). Así, las instrucciones del BDI-II evalúan la presencia de sintomatología depresiva durante el período mínimo necesario para el diagnóstico de un episodio depresivo mayor según el DSM-IV. Por otro lado, cuatro ítems del BDI-IA han sido suprimidos en el BDI-II y reemplazados por otros que evalúan síntomas de los criterios diagnósticos del DSM-IV que, además, son típicos de la depresión grave (agitación, sentimientos de inutilidad, dificultad de concentración y pérdida de energía). Además, en el BDI-II se han modificado los ítems de pérdida de

**TABLA 1**  
**NÚMERO DE TRABAJOS RECOGIDOS EN LAS BASES DE DATOS BIBLIOGRÁFICOS PSYCINFO Y PSYCODOC SOBRE LOS PRINCIPALES INSTRUMENTOS DE EVALUACIÓN DE LA DEPRESIÓN**

Instrumento	PsycINFO (2002-2013)	Psycodoc <sup>1</sup>
Inventario de Depresión de Beck	18723	352
Escala de Valoración de la Depresión de Hamilton	9528	109
Escala de Depresión del Centro de Estudios Epidemiológicos	8505	67
Escala Autoaplicada para la Depresión de Zung	1150	50

Nota. Búsquedas realizadas el 1 de mayo de 2013 y limitadas a los años 2002 a 2013, con las siguientes combinaciones de palabras en el campo "pruebas y medidas": "Beck Depression Inventory"; "Hamilton Rating Scale for Depression" o "Hamilton Depression Rating Scale"; "Self-Rating Depression Scale" o "Zung Depression Scale"; "CES-D" o "Center for Epidemiologic Studies Depression".

<sup>1</sup>Búsquedas realizadas el 1 de mayo de 2013 con las siguientes combinaciones de palabras en el campo "búsqueda simple": "inventario y Beck y depresión" o "BDI y depresión"; "Hamilton y escala y depresión" o "HRSD y depresión" o "HDRS y depresión" o "HAM y depresión"; "Zung y escala y depresión" o "Zung y SDS" o "SDS y depresión"; "CES-D y depresión" o "escala y depresión y centro y estudios y epidemiológicos".



apetito e insomnio para que éstos evalúen tanto los descensos como los aumentos en apetito y sueño, y se han introducido modificaciones en algunas afirmaciones de otros ítems. Por ejemplo, la redacción del ítem sobre desinterés social del BDI-IA se ha cambiado para que incluya también el desinterés por personas y actividades, de manera que el ítem del BDI-II evalúa más adecuadamente el criterio sintomático de pérdida de interés en todas o casi todas las actividades del DSM-IV. En definitiva, en el BDI-II tan sólo tres ítems del BDI-IA (sentimientos de castigo, pensamientos de suicidio y pérdida de interés en el sexo) no han sido modificados y, por tanto, constituye una importante revisión del BDI-I y BDI-IA.

Todas esas modificaciones parecen haber conseguido su objetivo. Un análisis, desde la perspectiva de la validez de contenido, de los cuestionarios autoaplicados más populares en España para evaluar la depresión clínica en adultos reveló que el BDI-II era el único instrumento que cubría todos los síntomas que definen el

episodio depresivo mayor y el trastorno distímico en el DSM-IV y cuyos ítems, además, reflejan todos ellos síntomas depresivos, de manera que sus puntuaciones no están influidas por síntomas ajenos al constructo de depresión (Sanz et al., en prensa; véase la Tabla 3).

**LA ADAPTACIÓN ESPAÑOLA DEL BDI-II**

Dadas las mejoras del BDI-II respecto a versiones anteriores, un grupo de investigación de la Universidad Complutense de Madrid inició en el año 2000 diversos estudios para validarlo en España (Sanz, Navarro y Vázquez, 2003; Sanz, Perdígón y Vázquez, 2003; Sanz, García-Vera, Espinosa, Fortún y Vázquez, 2005; Sanz y García-Vera, 2009, 2013), estudios que culminaron con la adaptación española del instrumento realizada con *Pearson Clinical & Talent Assessment España* y publicada por esta editorial (Beck et al., 2011).

La adaptación española del BDI-II está basada en un análisis exhaustivo de sus propiedades psicométricas de

**TABLA 2  
PRINCIPALES SEMEJANZAS Y DIFERENCIAS ENTRE LAS VERSIONES DEL BDI**

Características	BDI o BDI-I	BDI-IA	BDI-II
Año de publicación	1961	1979 (1978) <sup>a</sup>	1996
Nº de ítems	21	21	21
Nº de alternativas por ítem	10 ítems = 4 8 ítems = 5 1 ítem = 6	Todos los ítems = 4	19 ítems = 4 2 ítems = 7
Nº de ítems modificados respecto a la versión previa	-	15	18
Marco temporal de las instrucciones	Día de hoy	Última semana	Dos últimas semanas
Rango de puntuaciones de los ítems/total	0-3/0-63	0-3/0-63	0-3/0-63
Puntuaciones de corte para la gravedad de la depresión <sup>b</sup>	Normal = 0-9 Leve = 10-15 Moderada = 16-23 Grave = 24-63	Mínima = 0-9 Leve = 10-16 Moderada = 17-29 Grave = 30-63	Mínima = 0-13 Leve = 14-19 Moderada = 20-28 Grave = 29-63
Adaptación española	Conde et al. (1974)	Vázquez y Sanz (1991, 1997)	Sanz y Vázquez (Beck et al., 2011)
% cubierto de síntomas de depresión mayor DSM-IV <sup>c</sup>		88.9%	100%
% cubierto de síntomas de distimia DSM-IV <sup>c</sup>		100%	100%
% cubierto de síntomas de depresión mayor CIE-10 <sup>c</sup>		90%	100%
% cubierto de síntomas de distimia CIE-10 <sup>c</sup>		75%	75%
% de ítems que evalúan síntomas depresivos <sup>c</sup>		95.2%	100%
Fiabilidad de consistencia interna <sup>d</sup>		Personas de 18-49 años = .85 Personas > 50 años = .80 Pacientes = .86	Universitarios = .87 Adultos = .88 Pacientes = .91
Validez factorial <sup>e</sup>		Estructura interna más variable (2, 3, 4 o más factores)	Estructura interna más consistente (normalmente 2 factores)

Nota. <sup>a</sup> Aunque el BDI-IA fue registrado como propiedad intelectual en 1978, apareció publicado por primera vez en Beck et al. (1979).  
<sup>b</sup> Las puntuaciones de corte del BDI-I aparecen recogidas en Loeb, Feshbach, Beck y Wolf (1964; véase también Bumberry, Oliver y McClure, 1978), las del BDI-IA en la última edición de su manual (Beck y Steer, 1993) y las del BDI-II en su manual correspondiente (Beck et al., 1996).  
<sup>c</sup> Adaptado de Sanz et al. (en prensa).  
<sup>d</sup> Los datos de fiabilidad del BDI-I/BDI-IA para adultos de distintas edades y para pacientes proceden de los estudios metaanalíticos de Yin y Fan (2000) y Beck, Steer y Garbin (1988), respectivamente, mientras que los del BDI-II han sido calculados a partir de los datos recogidos en Beck et al. (2011), tras calcular la media, ponderada por el tamaño de la muestra de participantes, de 10 muestras de pacientes con trastornos psicológicos que totalizaban 3348 pacientes, de 7 muestras de adultos de la población general que totalizaban 4088 adultos y de 16 muestras de estudiantes universitarios que totalizaban 15123 estudiantes.  
<sup>e</sup> Véase Beck et al. (1988), Beck et al. (2011) y Sanz y García-Vera (2009).



fiabilidad de consistencia interna, validez convergente y discriminante, validez factorial y validez de criterio en tres muestras de pacientes con trastornos psicológicos que totalizaban 712 pacientes, dos muestras de adultos de la población general que totalizaban 569 adultos y dos muestras de estudiantes universitarios que totalizaban 727 estudiantes. Estos análisis sugieren, consistentemente en las diversas muestras, que la adaptación española del BDI-II muestra buenos índices psicométricos para medir la presencia y gravedad de sintomatología depresiva, índices similares a los de la versión original o de las adaptaciones realizadas en otros países (Beck et al., 2011). Para los objetivos del presente trabajo, cabe detallar algunos de esos índices.

Los coeficientes de Cronbach de fiabilidad de consistencia interna para los pacientes con trastornos psicológicos, los adultos de la población general y los estudiantes universitarios fueron, respectivamente, .90, .87 y .89, todos ellos excelentes ( $\geq .85$ ; Prieto y Muñiz, 2000) y comparables a los obtenidos en estudios con muestras semejantes de otros países (véase la Tabla 4).

Respecto a la validez de criterio para diferenciar grupos contrastados, la puntuación total media del BDI-II de los pacientes con trastornos psicológicos fue significativamente mayor que las de los adultos de la población general y los estudiantes universitarios (Beck et al., 2011). De hecho, la gravedad global de la depresión en los pacientes fue aproximadamente el doble de la encontrada en las mues-

tras no clínicas (véase la Tabla 5), lo que suponía, en términos del tamaño del efecto, unas diferencias estandarizadas de medias ( $d$  de Cohen) iguales a 1.07 y 1.21, respectivamente, que pueden considerarse grandes ( $d > 0.80$ ; Cohen, 1988).

En relación con la validez diagnóstica, un análisis de la curva de las características operativas del receptor o curva ROC reveló que el BDI-II mostraba un rendimiento diagnóstico (área bajo la curva ROC = .83) adecuado ( $> 0.70$ ; Swets, 1988) para detectar un trastorno depresivo mayor frente a un trastorno mental no depresivo en una muestra de pacientes con trastornos psicológicos evaluados mediante una entrevista diagnóstica estructurada y un listado de cotejo de síntomas depresivos (Beck et al., 2011). La capacidad de discriminación diagnóstica de cualquier instrumento depende de las características de la población en la que se utilice, por lo que el diagnóstico diferencial en una muestra de pacientes con trastornos psicológicos entraña más dificultad que si se pretende detectar personas con depresión en una muestra no clínica. Para este último objetivo, el análisis de la curva ROC mostró que el BDI-II tenía un rendimiento diagnóstico (área bajo la curva ROC = .91) alto ( $> 0.90$ ; Swets, 1988) para discriminar entre estudiantes universitarios con episodio depresivo mayor y estudiantes no deprimidos evaluados mediante una

**TABLA 3**  
**CORRESPONDENCIA ENTRE LOS ÍTEMS DE LOS PRINCIPALES CUESTIONARIOS, ESCALAS E INVENTARIOS DE DEPRESIÓN Y LOS CRITERIOS DIAGNÓSTICOS SINTOMÁTICOS PARA LOS TRASTORNOS DEPRESIVOS DEL DSM-IV (ADAPTADO DE SANZ ET AL., EN PRENSA)**

Instrumentos	% de síntomas de depresión mayor según el DSM-IV que cubren los ítems	% de síntomas de distimia según el DSM-IV que cubren los ítems	% de ítems que evalúan síntomas presivos
BDI-II	100%	100%	100%
CES-D	88.9%	85.7%	95%
MCMI-II/CC	77.8 %	57.1 %	67.7%
MCMI-III/CC	66.7%	71.4%	100%
MCMI-II/D	88.9 %	71.4%	100%
MCMI-III/D	44.4%	42.8%	92.8%
MMPI-D	88.9%	85.7%	45%
MMPI-2/D	88.9%	85.7%	47.4%
MMPI-2/DEP	77.8%	85.7%	75.8%
SCL-90-R/D	55.5%	42.8%	84.6%
SDS	100%	85.7%	90%

Nota. BDI-II: Inventario de Depresión de Beck-II; CES-D: Escala de Depresión del Centro de Estudios Epidemiológicos; MCMI-II/CC y MCMI-III/CC: escala de Depresión Mayor del Inventario Clínico Multiaxial de Millon; MCMI-II/D y MCMI-III/D: escala de Distimia del Inventario Clínico Multiaxial de Millon; MMPI-D y MMPI-2/D: escala clínica de Depresión del Inventario Multifásico de Personalidad de Minnesota; MMPI-2/DEP: escala de Depresión basada en el contenido del Inventario Multifásico de Personalidad de Minnesota-2; SCL-90-R/D: subescala de Depresión del Cuestionario de 90 Síntomas Revisado; SDS: Escala Autoaplicada de Depresión de Zung.

**TABLA 4**  
**FIABILIDAD DE CONSISTENCIA INTERNA (ALFA DE CRONBACH) DEL BDI-II EN DIFERENTES MUESTRAS Y CON DIFERENTES ADAPTACIONES (ADAPTADO DE BECK ET AL., 2011)**

Población/adaptación del BDI-II	N	alfa
Pacientes externos con trastornos mentales		
Original o adaptaciones de otros países	2636	.92
Adaptación española <sup>a</sup>	712	.90
Estudiantes universitarios		
Original o adaptaciones de otros países	14396	.87
Adaptación española <sup>a</sup>	727	.89
Adultos de la población general		
Original o adaptaciones de otros países	3519	.88
Adaptación española <sup>a</sup>	569	.87

Nota. <sup>a</sup>Valor medio de los coeficientes de las diferentes muestras de la adaptación española del BDI-II (Beck et al., 2011) tras ponderarlos por el tamaño de cada muestra.

**TABLA 5**  
**MEDIAS Y DESVIACIONES TÍPICAS DE LAS PUNTUACIONES DEL BDI-II EN DIFERENTES MUESTRAS ESPAÑOLAS (ADAPTADO DE BECK ET AL., 2011)**

	N	M	DT
Pacientes externos con trastornos mentales	712	19.98	10.96
Estudiantes universitarios	727	8.75	7.34
Adultos de la población general	569	9.61	7.76



entrevista diagnóstica estructurada. Estos dos análisis de rendimiento diagnóstico permitieron identificar varias puntuaciones de corte del BDI-II con índices adecuados de validez diagnóstica para detectar personas con depresión clínica (p. ej., sensibilidad y especificidad > 70%; valores predictivos positivo y negativo > 50%; kappa > .40; véase la Tabla 6).

**RECOMENDACIONES PARA UTILIZAR LA ADAPTACIÓN ESPAÑOLA DEL BDI-II EN LA PRÁCTICA CLÍNICA**

La interpretación de las puntuaciones del BDI se suele basar en unas puntuaciones de corte que definen diferentes niveles de gravedad de sintomatología depresiva. El manual original del BDI-II propone las siguientes: 0-13 indica depresión mínima, 14-19 depresión leve, 20-28 depresión moderada y 29-63 depresión grave (Beck et al., 1996). Estas puntuaciones fueron asumidas en la adaptación española del BDI-II y, recientemente, un estudio ha demostrado empíricamente su validez para distinguir niveles de gravedad de la depresión en pacientes españoles con trastornos psicológicos (Sanz, Gutiérrez, Gesteira y García-Vera, 2013).

Complementariamente, la interpretación de las puntuaciones en el BDI-II puede hacerse comparándolas con los baremos de un grupo de referencia. En Sanz et al. (2013) se pueden encontrar tablas de puntuaciones centiles del BDI-II para pacientes con trastornos psicológicos, adultos de la población general y estudiantes universitarios españoles, las cuales permiten evaluar con mayor precisión la gravedad de la depresión.

A partir de las propiedades psicométricas de la adaptación española del BDI-II, se pueden proponer pautas de interpretación para dos objetivos muy importantes en la práctica clínica: evaluar la significación clínica de los cambios terapéuticos e identificar personas con depresión clínica.

**EVALUACIÓN DE LA SIGNIFICACIÓN CLÍNICA DE LOS CAMBIOS TERAPÉUTICOS**

Dado que el BDI contiene sólo 21 ítems que se pueden responder en 5-10 minutos y que su corrección es tan fácil y rápida que puede realizarse en 1-2 minutos, es posible aplicar el BDI repetidamente durante el proceso terapéutico para monitorizar los progresos del paciente y evaluar los resultados de la terapia. En este contexto, el BDI-II puede ayudar a responder a la pregunta de si un paciente está igual, ha mejorado o ha empeorado de su depresión o incluso si ya se ha recuperado, es decir, si se ha producido una reducción o aumento clínicamente significativo en su sintomatología depresiva.

Para ello, se puede utilizar la aproximación estadística a la significación clínica de Jacobson y Truax (1991) que asume que un cambio clínicamente significativo supondría la vuelta a una población funcional de un paciente que antes del tratamiento pertenecía a una población disfuncional, es decir, que dicho cambio supondría que la puntuación de un paciente en el BDI-II ya no pertenece a la distribución de puntuaciones en el BDI-II de una población disfuncional (p. ej., los pacientes españoles con trastornos psicológicos) sino a la distribución de una población funcional (p. ej., la población general española).

Para determinar la existencia de un cambio clínicamente significativo en un paciente, el método de Jacobson y Truax (1991; McGlinchey, Atkins y Jacobson, 2002) implica, en primer lugar, establecer una puntuación de corte que el paciente debe alcanzar para pasar de una distribución disfuncional a una funcional. Cuando las dos distribuciones se solapan, como ocurre en el BDI-II (véase la Tabla 5), la mejor puntuación de corte (C) es el punto medio ponderado entre las medias de las distribuciones funcional y disfuncional:

$$C = \frac{(DT_n \times M_p) + (DT_p \times M_n)}{(DT_n + DT_p)}$$

en la que  $DT_n$  y  $DT_p$  representan las desviaciones típicas en el BDI-II de la población normal y de pacientes, respectivamente, y  $M_n$  y  $M_p$  las medias en el BDI-II de la población normal y de pacientes, respectivamente.

En segundo lugar, el método implica estimar si el cambio que indican las puntuaciones de un instrumento como el BDI-II no se debe a su error de medida, sino que refleja un cambio fiable, real, en el nivel de sintomatología depresiva del paciente. Para ello, se propone un índice de cambio fiable (*reliable change index; RCI*) que tiene en cuenta el error típico de la diferencia entre dos puntuaciones del instrumento ( $S_{diff}$ ), el cual depende de su

**TABLA 6**  
**ÍNDICES DE RENDIMIENTO DIAGNÓSTICO DE LAS PUNTUACIONES DEL BDI-II PARA DETECTAR PERSONAS CON TRASTORNOS DEPRESIVOS (ADAPTADO DE BECK ET AL., 2011)**

Puntuación BDI-II	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo	Kappa
Muestra de estudiantes universitarios (N=165)					
12	92.94	84.15	35.27	99.23	.44
19	57.14	96.72	61.54	96.09	.56
Muestra de pacientes externos con trastornos mentales (N=171)					
15	94.12	46.72	30.48	96.97	.23
22	82.35	73.72	43.75	94.39	.42
30	55.88	91.24	61.29	89.29	.49



error típico de medida ( $S_e$ ) que, a su vez, depende de su fiabilidad ( $r_{xx}$ ):

$$RCI = \frac{x_2 - x_1}{S_{dif}}$$

$$S_{dif} = \sqrt{2(S_e)^2} = \sqrt{2(s_x\sqrt{1-r_{xx}})^2}$$

en las que  $x_2$  sería la puntuación en el BDI-II de un paciente en un momento dado (p. ej., postratamiento),  $x_1$  la puntuación en el BDI-II en un momento anterior (p. ej., pretratamiento),  $s_x$  la desviación típica de las puntuaciones del BDI-II en la población de pacientes y  $r_{xx}$  la fiabilidad de consistencia interna del BDI-II en dicha población. El error típico de la diferencia entre las dos puntuaciones ( $S_{dif}$ ) describiría la amplitud de la distribución de las puntuaciones de cambio que se esperaría si no ocurriera ningún cambio real, de manera que un  $RCI$  mayor que 1.96 sería muy poco probable ( $p < .05$ ) que sucediera sin que ocurriera un cambio real. Cada paciente, pues, debería pasar ese  $RCI$  para demostrar que su cambio no se debe a los errores de medida del BDI-II:

$$RCI > 1.96 \Rightarrow \frac{x_2 - x_1}{S_{dif}} > 1.96 \Rightarrow x_2 - x_1 > S_{dif} \times 1.96$$

Basándose en estos dos criterios, el método de Jacobson y Truax (McGlinchey et al., 2002) clasifica a un paciente como *recuperado* si su puntuación en un instrumento ha superado tanto el  $RCI$  de 1.96 como la puntuación  $C$ , *mejorado* si dicha puntuación ha superado el  $RCI$  de 1.96, pero no la puntuación  $C$ , *sin cambios* si la puntuación no ha superado ninguno de los dos criterios, y *empeorado* si la puntuación ha superado el  $RCI$  de 1.96, pero en la dirección de un empeoramiento.

Con los datos de la adaptación española del BDI-II (véanse las Tablas 4 y 5), la puntuación  $C$  y el cambio que supone un  $RCI$  de 1.96 serían:

$$C = \frac{(7.76 \times 19.98) + (10.96 \times 9.61)}{(7.76 + 10.96)} = 13.91$$

$$S_{dif} = \sqrt{2(10.96\sqrt{(1-.90)})^2} = 4.89$$

$$x_2 - x_1 > 4.89 \times 1.96 \Rightarrow x_2 - x_1 > 9.58$$

En consecuencia, con dicha adaptación, un paciente cuya puntuación en el BDI-II ha descendido 10 puntos o más y dicha puntuación es menor de 14 se podría considerar *recuperado* de su depresión; si su puntuación ha descendido 10 puntos o más, pero la misma no llega a 14, se podría considerar *mejorado*; si su puntuación no ha descendido 10 puntos y tampoco llega a 14, se podría considerar *sin cambios*, y si la puntuación refleja un aumento de 10 puntos o más se podría considerar que ha *empeorado*.

No obstante, es importante recordar que la valoración del estado de un paciente debería tener en cuenta otros criterios sobre sus problemas depresivos (p. ej., para hablar de recuperación, habría que considerar si la puntuación  $< 14$  en el BDI-II se sigue repitiendo durante al menos dos meses) así como tener en cuenta otra información relevante sobre otros problemas y trastornos del paciente, sobre su nivel de funcionamiento en las actividades laborales o sociales habituales o en las relaciones con los demás, etc.

## IDENTIFICACIÓN DE PERSONAS CON DEPRESIÓN CLÍNICA

Aunque el BDI-II no es un instrumento diseñado para diagnosticar trastornos depresivos, la información que proporciona puede ayudar a identificar personas con un probable trastorno depresivo para su posterior confirmación diagnóstica con un instrumento más adecuado (p. ej., una entrevista diagnóstica estructurada) o servir como fuente complementaria de datos que ayude a realizar un diagnóstico de trastorno depresivo. Para ello, se pueden utilizar los índices de rendimiento diagnóstico de la Tabla 6. Así, cuando se use la adaptación española del BDI-II como cribado y, por tanto, con confirmación diagnóstica posterior de los casos detectados con el inventario, sería recomendable escoger puntuaciones de corte con una alta sensibilidad, es decir, con una alta capacidad para detectar un trastorno depresivo en pacientes que realmente lo están sufriendo, de manera que se asegure que ningún paciente que sufre dicho trastorno quede sin detectar. En este caso es preferible cometer un error debido a un falso positivo, ya que dicho error sería luego fácil de corregir mediante la confirmación diagnóstica posterior, que cometer un error debido a un falso negativo, ya que esta persona con trastorno depresivo no sería posteriormente evaluada y pasaría desapercibida. Para la población española no clínica, esa puntua-



ción de corte sería igual o mayor que 12, ya que esta puntuación, obteniendo un adecuado índice de especificidad ( $> 70\%$ ) y de concordancia diagnóstica ( $kappa > .40$ ), podría detectar al 93% de las personas que tienen un episodio depresivo mayor (véase la Tabla 6). Para la población española de pacientes con trastornos psicológicos, aunque la puntuación de corte de 15 tendría una sensibilidad del 93%, su especificidad y concordancia diagnóstica no serían adecuadas ( $< 70\%$  y  $< .40$ , respectivamente; véase la Tabla 6), por lo que sería más recomendable utilizar la puntuación de corte de 22, la cual podría detectar al 82% de los pacientes con trastornos depresivos mostrando a la vez una sensibilidad (74%) y concordancia diagnóstica (.42) adecuadas.

Cuando se use la adaptación española del BDI-II como ayuda para realizar un diagnóstico de trastorno depresivo que ya se sospechaba por la información de otros instrumentos, lo recomendable sería escoger puntuaciones de corte con una alta especificidad, es decir, con una alta capacidad para detectar la ausencia de trastorno depresivo en personas que realmente no padecen este trastorno, de manera que se asegure que un paciente al que se le asigna ese diagnóstico tiene realmente dicho trastorno (es preferible cometer un error debido a un falso negativo que debido a un falso positivo). De forma relacionada, sería recomendable que dicha puntuación tuviera un valor predictivo positivo mayor de 50%, es decir, que la probabilidad de que una persona que obtenga esa puntuación en el BDI-II padezca realmente un trastorno depresivo mayor esté por encima del 50%. Para las poblaciones españolas no clínicas y clínicas, esas puntuaciones de corte serían iguales o mayores que 19 y 30, respectivamente, ya que esas puntuaciones, obteniendo índices adecuados de concordancia diagnóstica ( $kappa > .40$ ), mostrarían especificidades superiores a 90% y valores predictivos positivos de 61% (véase la Tabla 6).

## CONCLUSIONES

En un artículo de 2010 sobre el uso de los tests en España, Muñiz y Fernández-Hermida subrayaban la amplia utilización del BDI a pesar de ser un instrumento que todavía no estaba comercializado en España y recomendaban que "sería altamente aconsejable que esta prueba tan utilizada por los profesionales dispusiese de un proceso de validación más sistemático y riguroso en nuestro país" (pp. 116-117). Afortunadamente, un año más tarde, justo cuando se cumplía el 50º aniversario de su primera publicación, la recomendación de Muñiz y Fernández-Hermida era seguida y los profesionales españoles cuentan ya con una adaptación española de la

última versión del BDI, el BDI-II, adaptación que es fruto de un proceso sistemático y riguroso de validación y está comercializada (Beck et al., 2011).

El BDI-II presenta modificaciones sustanciales respecto al BDI-I y BDI-IA diseñadas para evaluar los síntomas diagnósticos del DSM-IV para los trastornos depresivos y, de hecho, entre los instrumentos autoaplicados más utilizados en España para evaluar la depresión en adultos, el BDI-II es el único que cubre todos los síntomas del episodio depresivo mayor y la distimia según el DSM-IV. Además, el BDI-II muestra índices de consistencia interna y validez factorial superiores a los del BDI-I y BDI-IA. La adaptación española del BDI-II también muestra buenos índices de fiabilidad y validez en muestras clínicas y no clínicas, índices que permiten proponer pautas y puntuaciones de corte útiles para la evaluación de la gravedad de la depresión, la valoración de la significación clínica de los cambios terapéuticos, el cribado de personas con depresión y la ayuda en el diagnóstico diferencial de los trastornos depresivos.

## REFERENCIAS

- American Psychiatric Association (1995). *DSM-IV. Manual diagnóstico y estadístico de los trastornos mentales*. Barcelona: Masson. (Orig. 1994).
- Beck, A. T., Rush, A. J., Shaw, B. F., y Emery, G. (1979). *Cognitive therapy of depression*. Nueva York: Guilford. (Traducción española en Bilbao: DDB, 1983).
- Beck, A. T., y Steer, R. A. (1993). *Beck Depression Inventory. Manual*. San Antonio, TX: The Psychological Corporation.
- Beck, A. T., Steer, R. A., y Brown, G. K. (1996). *BDI-II. Beck Depression Inventory-Second Edition manual*. San Antonio, TX: The Psychological Corporation.
- Beck, A. T., Steer, R. A., y Brown, G. K. (2011). *Manual. BDI-II. Inventario de Depresión de Beck-II* (Adaptación española: Sanz, J., y Vázquez, C.). Madrid: Pearson.
- Beck, A. T., Steer, R. A., y Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77-100.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., y Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Botella, J., y Ponte, G. (2011). Effects of the heterogeneity of the variances on reliability generalization: an example with the Beck Depression Inventory. *Psicothema*, 23, 516-522.
- Bumberry, W., Oliver, J., y McClure, J. N. (1978). Validation of the Beck Depression Inventory in a university



- population using psychiatric estimate as the criterion. *Journal of Consulting and Clinical Psychology*, 46, 150-155.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2<sup>a</sup> ed. Hillsdale, NJ: LEA.
- Conde, V., y Useros, E. (1974). El inventario para la medida de la depresión de Beck. *Revista de Psiquiatría y Psicología Médica de Europa y América Latina*, 12, 153-167.
- Cuijpers, P., Li, J., Hofmann, S. G., y Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. *Clinical Psychology Review*, 30, 768-778.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23, 56-61.
- Jacobson, N. S., y Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Loeb, A., Feshbach, S., Beck, A. T., y Wolf, A. (1964). Some effects of reward upon the social perception and motivation of psychiatric patients varying in depression. *Journal of Abnormal and Social Psychology*, 68, 609-616.
- McGlinchey, J. B., Atkins, D. C., y Jacobson, N. S. (2002). Clinical significance methods: which one to use and how useful are they? *Behavior Therapy*, 33, 529-550.
- Muñiz, J., y Fernández-Hermida, J. R. (2000). La utilización de los test en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J., y Fernández-Hermida, J. R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests. *Papeles del Psicólogo*, 31, 108-121.
- Nezu, A. M., Nezu, C. M., Friedman, J., y Lee, M. (2009). Assessment of depression. En I. H. Gotlib y C. L. Hammer (Eds.), *Handbook of depression* (2<sup>a</sup> ed.) (pp. 44-68). Nueva York, NY: Guilford Press.
- Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Radloff, L. (1977). The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Sanz, J., y García-Vera, M. P. (2009). The Beck Depression Inventory-second edition (BDI-II): factor congruence and generalizability of its indexes of internal consistency. En E. Rehulka (Ed.), *School and health* 21. *General issues in health education* (pp. 331-342). Brno: MSD.
- Sanz, J., y García-Vera, M. P. (2013). Rendimiento diagnóstico y estructura factorial del Inventario para la Depresión de Beck-Segunda Edición (BDI-II). *Anales de Psicología*, 29, 66-75.
- Sanz, J., García-Vera, M. P., Espinosa, R., Fortún, M., y Vázquez, C. (2005). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. *Clínica y Salud*, 16, 121-142.
- Sanz, J., Gutiérrez, S., Gesteira, C., y García-Vera, M. P. (2013). Criterios y baremos para interpretar las puntuaciones en la adaptación española del "Inventario de Depresión de Beck-II" (BDI-II). Manuscrito bajo revisión editorial.
- Sanz, J., Izquierdo, A., y García-Vera, M. P. (en prensa). Una revisión desde la perspectiva de la validez de contenido de los cuestionarios, escalas e inventarios autoaplicados más utilizados en España para evaluar la depresión clínica en adultos. *Psicopatología Clínica, Legal y Forense*.
- Sanz, J., Navarro, M. E., y Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 1. Propiedades psicométricas en estudiantes universitarios. *Análisis y Modificación de Conducta*, 29, 239-288.
- Sanz, J., Perdígón, L. A., y Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clínica y Salud*, 14, 249-280.
- Sanz, J., y Vázquez, C. (2008). Trastornos del estado de ánimo: teorías psicológicas. En A. Belloch, B. Sandín y F. Ramos (Eds.), *Manual de Psicopatología*, vol. 2. Edición revisada (pp. 271-298). Madrid: McGraw-Hill.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Tennen, H., Hall, J. A., y Affleck, G. (1995). Depression research methodologies in the Journal of Personality and Social Psychology: a review and critique. *Journal of Personality and Social Psychology*, 68, 870-884.
- Vázquez, C., y Sanz, J. (1997). Fiabilidad y valores normativos de la versión española del Inventario para la Depresión de Beck de 1978. *Clínica y Salud*, 8, 403-422.
- Yin, P., y Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: reliability generalization across studies. *Educational and Psychological Measurement*, 60, 201-223.
- Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.