

CRITERIOS Y BAREMOS PARA INTERPRETAR EL "INVENTARIO DE DEPRESIÓN DE BECK-II" (BDI-II)

Jesús Sanz, Sara Gutiérrez, Clara Gesteira y María Paz García-Vera
Universidad Complutense de Madrid (España)

Resumen

Se presentan dos estudios que ofrecen criterios y baremos para interpretar las puntuaciones de la adaptación española del BDI-II. El objetivo del estudio 1 fue validar las puntuaciones de corte originales del BDI-II para distinguir distintos niveles de gravedad de la depresión y buscar las mejores puntuaciones de corte para la adaptación española. Para ello, 215 pacientes españoles con diversos trastornos psicológicos completaron el BDI-II y se les aplicó una escala de heteroevaluación con la cual los psicólogos que les atendían valoraron la gravedad de su depresión. Los resultados de estas valoraciones validaron los niveles de depresión definidos por las puntuaciones de corte originales del BDI-II; además, mediante curvas ROC, se obtuvieron unas puntuaciones de corte para la adaptación española prácticamente iguales a aquellas. El objetivo del estudio 2 fue elaborar tablas de puntuaciones centiles a partir de tres muestras españolas: 712 pacientes con trastornos psicológicos, 569 adultos de la población general y 727 estudiantes universitarios. Se discuten las ventajas de la utilización combinada de los criterios y baremos propuestos, pues ambos ofrecen información complementaria.

PALABRAS CLAVE: evaluación de la depresión, Inventario de depresión de Beck, puntuaciones de corte, baremos.

Abstract

This paper reports two studies providing criteria and norms to interpret scores on the Spanish adaptation of the BDI-II. The first study was aimed at validating the original BDI-II cut-off scores to distinguish levels of depression severity and at finding the best cut-off scores for the Spanish adaptation. To get these aims, 215 Spanish patients with a variety of psychological disorders filled out the BDI-II and a clinician-rated scale was administered by their psychologists to assess the severity of patients' depression. The results of these clinician ratings validated the depression levels defined by the original BDI-II cut-off scores; in addition, by means of ROC curves, it was found that cut-off scores for the

Los autores quieren agradecer a los psicólogos de la Clínica Universitaria de Psicología de la Universidad Complutense de Madrid su inestimable colaboración en la recogida de datos y a Pearson Educación, S. A., a través de su división *Pearson Clinical and Talent Assessment España*, su permiso para utilizar los datos de las muestras de la adaptación española del BDI-II.

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

Spanish adaptation were practically equal to those for the original BDI-II. The aim of the second study was to develop tables of percentiles with three Spanish samples: 712 patients with psychological disorders, 569 adults from the general population and 727 university students. The advantages of the combined use of the proposed criteria and norms are discussed since they provide complementary information.

KEY WORDS: *depression assessment, Beck Depression Inventory, cutting scores, test norms.*

Introducción

En la actualidad, el "Inventario de depresión de Beck", internacionalmente conocido por el acrónimo de su nombre original en inglés (*Beck Depression Inventory*, BDI), es, en sus diferentes versiones, la original (BDI-I; Beck, Ward, Mendelson, Mock y Erbaugh, 1961), la revisada (BDI-IA; Beck, Rush, Shaw y Emery, 1979) y la segunda edición (BDI-II; Beck, Steer y Brown, 1996), el instrumento específico para medir la depresión más utilizado e investigado. Una búsqueda en *PsycINFO* obtuvo, para los años 2002 a 2013, 18.723 trabajos sobre el BDI, cantidad que superaba con creces los 9.528 trabajos sobre la "Escala de valoración de la depresión de Hamilton" (*Hamilton Rating Scale for Depression*, HRSD; Hamilton, 1967), los 8.505 sobre la "Escala de depresión del Centro de Estudios Epidemiológicos" (*Center for Epidemiologic Studies Depression Scale*, CES-D; Radloff, 1977) o los 1.150 sobre la "Escala autoaplicada para la depresión de Zung" (*Zung Self-Rating Depression Scale*, SDS; Zung, 1965), tres de los instrumentos que gozan de mayor reputación en el ámbito de la depresión (Nezu, Nezu, Friedman y Lee, 2009). Esa búsqueda se realizó el 1 de mayo de 2013 con las siguientes palabras y 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". Esa frecuente utilización e investigación también se extiende a España. Se realizó en *Psicodoc* una búsqueda en la misma fecha que la anterior y utilizando las siguientes combinaciones de palabras en el campo general de "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". Esta búsqueda localizó 352 trabajos sobre el BDI, cantidad que de nuevo sobrepasaba ampliamente los 109 trabajos sobre la HRSD, los 67 sobre la CES-D y los 50 sobre la SDS. Coherentemente, una investigación reciente (Muñiz y Fernández-Hermida, 2010) encontró que el BDI es el quinto test más utilizado por los psicólogos españoles en su práctica diaria, lo cual no debería sorprender dado que el BDI cuenta con traducciones castellanas validadas en España para todas sus versiones: BDI-I (Conde y Useros, 1974; Conde, Esteban y Useros 1976), BDI-IA (Sanz y Vázquez, 1998; Vázquez y Sanz, 1997, 1999) y BDI-II (Beck, Steer y Brown, 2011a)

e incluso para las diferentes formas breves que se han desarrollado de las dos versiones más recientes (Beck, Steer y Brown, 2011b; Sanz y García-Vera, 2007).

Interpretación del BDI-II basada en puntuaciones de corte

La interpretación de las puntuaciones en el BDI en cualquiera de sus versiones y formas se basa habitualmente en unos criterios o puntuaciones de corte que definen diferentes categorías o niveles de gravedad de sintomatología depresiva. Así, el manual original del BDI-II (Beck *et al.*, 1996, p. 11) propone las siguientes puntuaciones de corte y grados de depresión correspondientes: 0-13 indica depresión mínima, 14-19 depresión leve, 20-28 depresión moderada y 29-63 depresión grave. El uso de criterios o puntuaciones de referencia para distinguir distintos niveles de sintomatología es también habitual en muchos otros instrumentos psicopatológicos y responde a una tradición clínica sobre la utilidad de diferenciar niveles sindrómicos en función del número de síntomas presentes y su intensidad. Por ejemplo, el DSM-IV clasifica así el episodio depresivo mayor en función de su gravedad: “La gravedad se considera leve, moderada o grave en base al número de síntomas de los criterios, la gravedad de los síntomas y el grado de incapacidad funcional y malestar. Los episodios *leves* se caracterizan por la presencia de sólo cinco o seis síntomas depresivos y por una incapacidad leve o por una capacidad de función normal, pero a costa de un esfuerzo importante e inhabitual. Los episodios *graves sin síntomas psicóticos* se caracterizan por la presencia de la mayoría de los síntomas de los criterios y por una incapacidad evidente y observable (p. ej., incapacidad para trabajar o para cuidar de los hijos). Los episodios *moderados* tienen una gravedad intermedia entre los leves y los graves” (APA, 1995, p. 385).

Parece lógico que las propuestas de este tipo de criterios de referencia deberían estar basadas en estudios empíricos sobre la validez de las interpretaciones que se realizan sobre los resultados del instrumento. Así, la Comisión Internacional de Test (*International Test Commission* [ITC]; Colegio Oficial de Psicólogos e ITC, 2000) establece que se deberían “utilizar puntos de corte en la interpretación de las puntuaciones sólo cuando se disponga de datos empíricos sobre su validez” (directriz 2.7.9). A este respecto, las puntuaciones de corte del manual original del BDI-II se obtuvieron mediante distintas curvas de las características operativas del receptor (*receiver operating characteristic curves*), o curvas ROC, elaboradas a partir de las puntuaciones en el BDI-II de una muestra de 101 pacientes estadounidenses ambulatorios con trastornos psicológicos, 57 de los cuales tenían un diagnóstico de episodio depresivo mayor, único o recurrente, con el especificador de gravedad del DSM-III-R correspondiente (depresión leve, moderada o grave), y los restantes 44 un diagnóstico de otro trastorno psicológico no depresivo (Beck *et al.*, 1996). Con los datos de esta muestra, las puntuaciones de corte fueron calculadas tras obtener las tres curvas ROC que se mencionan a continuación y tras buscar en ellas las puntuaciones con valores más altos de especificidad y sensibilidad, aunque priorizando esta última: (1) la curva ROC para el grupo no depresivo frente al de depresión leve; (2) la curva ROC para los grupos no depresivo y con depresión leve frente al de depresión moderada y (c) la curva

ROC para los grupos no depresivo y con depresión leve o moderada frente al de depresión grave.

En el manual de la adaptación española del BDI-II (Beck, Steer y Brown, 2011a, adaptación española de J. Sanz y C. Vázquez) se asumieron las puntuaciones de corte de Beck *et al.* (1996) y, aunque se ofrecieron para las puntuaciones del BDI-II índices de rendimiento diagnóstico (p. ej., sensibilidad, especificidad) para detectar pacientes con trastorno depresivo mayor en una muestra española de pacientes externos con trastornos psicológicos, no se validaron explícitamente las puntuaciones de corte originales para distinguir distintos niveles de gravedad de la depresión en la población de pacientes españoles con trastornos psicológicos ni se buscaron las mejores puntuaciones de corte para esta población. Estos fueron, precisamente, los objetivos del estudio 1 de este trabajo.

Interpretación del BDI-II basada en normas

Junto con la interpretación basada en puntuaciones de corte o criterios, la interpretación de las puntuaciones de un test también puede hacerse comparándolas con las que obtienen un grupo de referencia, por ejemplo, una muestra de la población general, una muestra de pacientes con trastornos psicológicos, etc. Las puntuaciones de este grupo sirven como normas o baremos para interpretar la que obtiene una persona en particular y, por esa razón, a ese grupo de referencia se le denomina grupo normativo. Siguiendo a Urbina, “el término normas se refiere al rendimiento o conducta típica en el test de uno o más grupos de referencia. Las normas se presentan habitualmente en forma de tablas con estadísticos descriptivos —tales como medias, desviaciones típicas y distribuciones de frecuencia— que resumen el rendimiento del grupo o los grupos en cuestión” (Urbina, 2004, p. 78). Por esta razón es muy habitual que los test proporcionen medias, desviaciones típicas y procedimientos para convertir las puntuaciones directas en puntuaciones tipificadas, es decir, puntuaciones que expresan la posición relativa de la persona evaluada en la distribución del grupo de referencia, entre las cuales, las más habituales en test psicopatológicos, son las puntuaciones centiles (o percentiles), las típicas, las típicas normalizadas y varias derivadas de estas dos últimas (p. ej., las puntuaciones *T*).

El manual de la versión original del BDI-II presenta las medias y desviaciones típicas obtenidas en una muestra incidental de 500 pacientes estadounidenses con trastornos psicológicos y en otra de 120 estudiantes universitarios canadienses (Beck *et al.*, 1996, p. 35). En la misma línea, en el manual de la adaptación española del BDI-II se presentan las medias y desviaciones típicas obtenidas en tres muestras incidentales españolas: una de 712 pacientes externos con trastornos psicológicos, otra de 569 adultos de la población general y una tercera de 727 estudiantes universitarios (Beck, Steer y Brown, 2011a, p. 81). Sin embargo, y al contrario de lo que ocurre en algunos otros test psicopatológicos, ni el manual original del BDI-II ni el manual de su adaptación española presentan tablas de puntuaciones tipificadas, por ejemplo, tablas de puntuaciones *T* o de puntuaciones centiles. Estas tablas tienen varias ventajas. En primer lugar, aunque a partir de las

medias y desviaciones típicas de los manuales del BDI-II el evaluador podría calcular la puntuación típica o la puntuación *T* de una persona en concreto (aunque no sus puntuaciones típicas o *T* normalizadas), su propio cálculo podría suponer un obstáculo práctico para que dicho evaluador utilizara rutinariamente la información que proporciona esta otra forma de interpretar las puntuaciones del BDI-II. En segundo lugar, las puntuaciones centiles no pueden calcularse a partir de la media y desviación típica del grupo de referencia (salvo que la distribución de las puntuaciones directas de un test sea normal, lo cual, en el caso del BDI-II, es habitual en muestras clínicas –p. ej., Sanz, García-Vera, Espinosa, Fortún y Vázquez, 2005a; Sanz y García-Vera, 2013–, pero no en muestras no clínicas –p. ej., Sanz, Navarro y Vázquez, 2003; Sanz, Perdigón y Vázquez, 2003–) y, precisamente, tales puntuaciones destacan por su facilidad de interpretación (p. ej., decir que una persona tiene el centil 60 en el BDI-II es lo mismo que decir que su depresión es más grave que la que tiene el 60% de las personas de su población de referencia) y por su significación universal (p. ej., un centil 60 en el BDI-II significa lo mismo que un centil 60 en la SDS: la puntuación en cuestión es superior a la alcanzada por el 60% de la población respectiva). Debido a estas ventajas, Yela (1984) aconsejaba la tipificación de un test en normas centiles si este iba ser destinado al uso práctico, principalmente al clínico. En el caso de la adaptación española del BDI-II, Sanz, Perdigón *et al.* (2003, p. 265) publicaron una tabla de puntuaciones centiles para la población general española y Sanz, Navarro *et al.* (2003, p. 258) hicieron lo propio para la población de estudiantes universitarios españoles, pero, en cambio, no se ha publicado ninguna tabla de puntuaciones centiles para la población de pacientes españoles con trastornos psicológicos. En consecuencia, el objetivo del estudio 2 de este trabajo fue justamente elaborar esa tabla que pudiera ayudar en la interpretación de las puntuaciones de la adaptación española BDI-II. Además, se actualizaron las tablas de puntuaciones centiles para población general y estudiantes universitarios españoles al ampliar las muestras que participaron en los estudios de Sanz, Perdigón *et al.* (2003) y Sanz, Navarro *et al.* (2003).

Estudio 1

Para validar las puntuaciones de corte del manual original del BDI-II se utilizaron los datos de una muestra de pacientes con trastornos psicológicos que habían sido valorados por los psicólogos que los atendían en una escala de cuatro categorías en función de la gravedad de su depresión (mínima, leve, moderada y grave) y se analizó si los grupos de pacientes definidos según las puntuaciones de corte del manual original del BDI-II mostraban entre sí diferencias coherentes y significativas en su gravedad según la evaluación de los psicólogos. A partir de esos mismos datos y de forma parecida a como lo hicieron Beck *et al.* (1996), se buscaron las puntuaciones que, primando la sensibilidad, tenían un mejor equilibrio entre sensibilidad y especificidad, de manera que se esperaba obtener puntuaciones de corte para la adaptación española del BDI-II similares a las propuestas por Beck *et al.* (1996), lo cual serviría de validación adicional de estas últimas.

Método

Participantes

En este estudio se analizaron los datos de una parte de la muestra de pacientes con trastornos psicológicos que participó en el estudio de Sanz y García-Vera (2013) y que constituyó una de las muestras de la adaptación española del BDI-II (Beck *et al.*, 2011a). En dicho estudio se administró el BDI-II a 322 pacientes atendidos en la Clínica Universitaria de Psicología de la Universidad Complutense de Madrid (UCM), de los cuales, a 215 se les aplicó también el “Listado para el episodio depresivo” (LED; Sanz, García-Vera, Espinosa, Fortún y Vázquez, 2005b), una escala de heteroevaluación que permitió al psicólogo que les atendía valorar la gravedad de su depresión (Apéndice). Esta submuestra de 215 pacientes (67,8% mujeres), con edades comprendidas entre 18 y 64 años ($M= 30,9$; $DT= 11,5$), fue la que participó en el presente estudio (en la tabla 1 se resumen sus diagnósticos DSM-IV y su sexo y edad en función de los mismos).

Tabla 1

Distribución de diagnósticos y características demográficas de la muestra del estudio 1 ($N= 215$)

Diagnósticos según el DSM-IV	N	%	Edad M (DT)	% Mujeres
Trastorno depresivo mayor	34	15,8	32,1 (12,9)	72,7
Otros trastornos depresivos	5	2,3	30,2 (14,4)	80,0
Trastorno de angustia/agorafobia	17	7,9	28,5 (11,6)	76,5
Fobia social	11	5,1	30,4 (9,5)	63,6
Obsesivo compulsivo	16	7,4	28,5 (12,2)	31,3
Ansiedad generalizada	10	4,6	30,5 (12)	90
Otros trastornos de ansiedad	13	6,1	29,2 (10,9)	84,6
Trastornos adaptativos	33	15,3	35,7 (13,6)	72,7
Trastornos de la conducta alimentaria	7	3,2	22,3 (2,4)	100
Trastornos de la personalidad	5	2,3	27,2 (7,4)	80
Problemas de relación (conyugales, paternofiliales, otros)	10	4,6	36,6 (10,1)	70
Otros trastornos o problemas	54	25,1	29,7 (9,4)	55,6

Instrumentos

- “Inventario de depresión de Beck-II” (BDI-II; Beck *et al.*, 1996). El BDI-II es un instrumento de autoinforme de 21 ítems diseñado para evaluar la gravedad de la sintomatología depresiva. En cada uno de sus ítems la persona evaluada tiene que elegir, entre cuatro alternativas ordenadas de menor a mayor gravedad, la frase que mejor describe su estado durante las últimas dos semanas (excepto en los ítems 16 y 18 que presentan siete alternativas). Cada ítem se valora de 0 a 3 puntos en función de la alternativa escogida y, tras sumar directamente la puntuación de cada ítem, se obtiene una puntuación

total que varía de 0 a 63. En este estudio se utilizó la versión española del BDI-II de Sanz y Vázquez (Sanz *et al.*, 2005a; Sanz y García-Vera, 2013; Sanz, Navarro *et al.*, 2003; Sanz, Perdigón *et al.*, 2003), la cual presenta índices de fiabilidad y validez adecuados para pacientes con trastornos mentales.

- “Listado para el episodio depresivo” (LED; Sanz *et al.*, 2005b). El LED es un instrumento de heteroevaluación para medir la presencia de síntomas depresivos y valorar la gravedad de la depresión en pacientes con trastornos mentales (Apéndice A). Consta de dos preguntas abiertas y de 12 preguntas con tres opciones de respuesta: “Sí”, “No” y “¿?” (sin información suficiente-no estoy seguro), que cubren los criterios diagnósticos del episodio depresivo mayor según el DSM-IV (APA, 1995). El LED fue diseñado para que un clínico, después de una entrevista de al menos 45 minutos con el paciente, coteje dichos criterios y determine la presencia de un diagnóstico de episodio depresivo. Además, el LED incluye una última pregunta en la que el clínico debe estimar la gravedad de los síntomas depresivos del paciente en una escala de tipo Likert de cuatro puntos (“sin depresión/depresión mínima”, “depresión leve”, “depresión moderada” y “depresión grave”) que se puntúa de 0 a 3 y que fue la que se utilizó como criterio de gravedad de la depresión en el presente trabajo. En el estudio sobre las propiedades psicométricas del BDI-II en pacientes españoles con trastornos psicológicos para el cual se desarrolló *ad hoc* el LED (Sanz *et al.*, 2005a), su ítem de gravedad de la depresión mostró una correlación positiva y estadísticamente significativa ($p < 0,001$) con el BDI-II ($r = 0,51$; $N = 101$), al igual que lo hizo en el presente estudio ($r = 0,66$; $N = 215$). Ambos coeficientes de correlación presentaban magnitudes superiores a 0,50, valor que constituye el estándar para una correlación de tamaño grande (Cohen, 1988) y que es considerado bueno según los criterios propuestos por Prieto y Muñiz (2000) para valorar los índices de validez convergente. Un análisis posterior de los datos del estudio de Sanz *et al.* (2005a) reveló que los ítems sintomáticos del LED (ítems 1 a 9) conformaban, con un alfa de Cronbach de 0,69 ($N = 95$), una escala con una adecuada fiabilidad de consistencia, aunque con algunas carencias (Prieto y Muñiz, 2000), y con la cual correlacionaba de forma significativa ($p < 0,001$) y grande ($r = 0,62$) el ítem de gravedad del LED. En la muestra de pacientes del presente estudio, la fiabilidad de consistencia interna del conjunto de los ítems sintomáticos del LED también fue adecuada con algunas carencias ($\alpha = 0,64$; $N = 130$), y su correlación con el ítem de gravedad fue también significativa y grande ($r = 0,65$). Aunque las dos correlaciones del conjunto de los ítems sintomáticos del LED con su ítem de gravedad podrían considerarse como índices adecuados de fiabilidad de consistencia interna de este último ítem, no se dispone hasta ahora de información sobre su fiabilidad interjueces, lo cual hubiera sido muy deseable dado que en los instrumentos de heteroevaluación como el LED es quizás el aspecto de la fiabilidad más importante a evaluar, máxime cuando el instrumento fue utilizado por varios evaluadores. Sobre esta limitación del presente trabajo se volverá más adelante en la Discusión.

Procedimiento

En el momento de admisión a la clínica, todos los pacientes pasaron por una evaluación individual realizada por un psicólogo clínico con una duración media de tres sesiones y tras la cual se efectuó un diagnóstico DSM-IV (APA, 1995). Durante esa evaluación se administró, por este orden, el BDI-II y el LED, junto a otros instrumentos específicos para los problemas concretos de los pacientes o que formaban parte de otras investigaciones en curso en la clínica y, en prácticamente todos los casos, el BDI-II y el LED fueron aplicados al finalizar la primera o segunda sesión de evaluación (28,4% y 67,1% de los casos, respectivamente). En todos los casos, los psicólogos completaron el LED sin conocimiento de la puntuación del paciente en el BDI-II. Todos los pacientes fueron atendidos en régimen ambulatorio y de todos ellos se recogió información demográfica y clínica básica. Todos los psicólogos clínicos tenían formación de postgrado a nivel de máster en la evaluación, diagnóstico y tratamiento psicológicos de trastornos mentales.

Análisis de datos

Para validar las puntuaciones de corte del manual original del BDI-II, se realizó, con el paquete estadístico SPSS v. 19 (al igual que el resto de análisis estadísticos, salvo que se indique lo contrario), un ANCOVA sobre las puntuaciones del LED con los niveles de gravedad del BDI-II como factor intersujetos y con el sexo y la edad como covariables. Posteriormente, si el factor intersujetos de los niveles o grupos de gravedad del BDI-II resultaba estadísticamente significativo, se realizaron pruebas *t* de comparación de pares de grupos, con el nivel de significación corregido según el procedimiento de Bonferroni, y se calculó, como medida del tamaño del efecto, la diferencia de medias estandarizada (*d*) de Cohen (1988).

Para obtener las mejores puntuaciones de corte para la adaptación española del BDI-II se elaboraron, mediante el programa *MedCalc* v. 12.7.2, distintas curvas ROC que distinguían, de acuerdo al criterio de las categorías de gravedad del LED, los pacientes con depresión mínima de los pacientes con depresión leve, estos dos grupos de los pacientes con depresión moderada, y estos tres últimos grupos de pacientes de los pacientes con depresión grave. A partir de estas curvas se buscaron las puntuaciones que, primando la sensibilidad, tenían un mejor equilibrio entre sensibilidad y especificidad. Para ello, además, se calculó, para cada una de esas puntuaciones, el índice de Youden (*J*), una medida conjunta de eficacia diagnóstica que refleja la diferencia entre la tasa de verdaderos positivos y la de falsos positivos y que, por tanto, tiene en cuenta simultáneamente la sensibilidad y la especificidad, tal y como refleja su expresión algebraica: $J = \text{sensibilidad} - (1 - \text{especificidad})$, de modo que cuanto más cerca está de 1, mejor es la eficacia de la puntuación de corte para distinguir entre grupos de pacientes con diferentes niveles de depresión.

Finalmente, para examinar la validez de criterio concurrente de la adaptación española del BDI-II, se realizó un ANCOVA sobre las puntuaciones del BDI-II con las categorías de gravedad del LED como factor intersujetos y con el sexo y la edad

como covariables, y, en el caso de que dicho factor fuera estadísticamente significativo, se realizaron pruebas *t* de comparación de pares de grupos, con el nivel de significación corregido según el procedimiento de Bonferroni, y se calcularon las correspondientes *d* de Cohen.

Resultados

En la tabla 2 se recogen las puntuaciones medias que obtuvieron en el LED los grupos de pacientes con diferentes niveles de gravedad de su depresión definidos por las puntuaciones de corte del BDI-II. Tal y como se esperaba, un ANCOVA sobre las puntuaciones del LED con los niveles de gravedad del BDI-II como factor intersujetos reveló que existían diferencias estadísticamente significativas en el LED entre los grupos con distintos niveles de gravedad del BDI-II, $F(3, 208) = 59,30$; $p < 0,001$. Las posteriores pruebas *t* de comparación de pares de grupos indicaron que, como media, el grupo de pacientes con depresión mínima puntuó significativamente más bajo en el LED que el de depresión leve (0,54 frente a 1,22), éste, a su vez, puntuó más bajo que el de depresión moderada (1,22 frente a 1,54), aunque esta diferencia no alcanzó la significación estadística, y, a su vez, el de depresión moderada puntuó significativamente más bajo que el de depresión grave (1,54 frente a 2,15) (tabla 2). Es más, todas estas diferencias entre grupos alcanzaron, según los estándares de Cohen (1988), tamaños del efecto grandes ($d > 0,80$), salvo por la diferencia entre los grupos con depresión leve y moderada cuya diferencia fue moderada ($d \approx 0,50$) (tabla 2).

Por otro lado, en la tabla 2 también se presentan las puntuaciones medias que obtuvieron en el BDI-II los grupos de pacientes con diferentes niveles de gravedad de su depresión definidos por las puntuaciones o categorías de gravedad del LED. Tal y como cabría esperar, un ANCOVA realizado sobre las puntuaciones del BDI-II con las categorías de gravedad del LED como factor intersujetos reveló que existían diferencias estadísticamente significativas en el BDI-II entre los grupos con distintos niveles de gravedad del LED, $F(3, 208) = 57,65$; $p < 0,001$. Las posteriores pruebas *t* indicaron que, como media, el grupo de pacientes con depresión mínima puntuó significativamente más bajo en el BDI-II que el de depresión leve (10,0 frente a 16,7), éste, a su vez, puntuó significativamente más bajo que el de depresión moderada (16,7 frente a 26,4), y, a su vez, éste último puntuó significativamente más bajo que el de depresión grave (26,4 frente a 34,8) (tabla 2). Es más, todas estas diferencias entre grupos alcanzaron tamaños del efecto grandes ($d > 0,80$) (tabla 2).

Tabla 2
Diferencias entre las puntuaciones medias en el LED y el BDI-II para las categorías de gravedad de la depresión

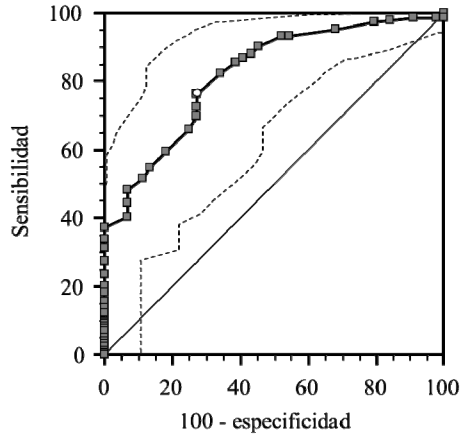
Instrumentos	Categorías de gravedad de la depresión [†]				F del ANCOVA [‡]	Diferencias entre categorías adyacentes de gravedad de la depresión	Categorías adyacentes de gravedad de la depresión comparadas		
	Mínima	Leve	Moderada	Grave			Leve vs. mínima	Moderada vs. leve	Grave vs. moderada
LED									
<i>M</i>	0,54 _a	1,22 _b	1,54 _b	2,15 _c	59,30***	<i>t</i>	5,71	2,27	4,29
<i>DT</i>	0,53	0,72	0,71	0,69		<i>p</i>	0,001	0,025	0,001
<i>n</i>	68	46	54	46		<i>d</i>	1,11	0,45	0,87
BDI-II									
<i>M</i>	10,0 _a	16,7 _b	26,4 _c	34,8 _d	57,65***	<i>t</i>	4,40	6,58	3,51
<i>DT</i>	6,6	8,9	9,2	8,9		<i>p</i>	0,001	0,001	0,001
<i>n</i>	44	83	69	18		<i>d</i>	0,82	1,07	0,92

Notas: Tras comparar dos a dos todas las medias, las medias con diferentes subíndices, tomadas dos a dos, difieren con $p < 0,05$ ajustada con el procedimiento de Bonferroni (0,05/6 comparaciones= 0,0083). *d*= tamaño del efecto *d* de Cohen (diferencia de medias estandarizada). [†]Para el LED, las categorías de gravedad están definidas por los rangos de puntuación del BDI-II propuestos por Beck *et al.* (1996), mientras que para el BDI-II las categorías de gravedad están definidas por las valoraciones de los clínicos en el LED. [‡] Resultados del ANCOVA realizado sobre las puntuaciones del BDI-II o del LED con las categorías de gravedad como factor intersujetos y con el sexo y la edad como covariables. *** $p < 0,001$.

En la figura 1 se presenta la curva ROC elaborada para distinguir, de acuerdo al criterio de las categorías de gravedad del LED, los pacientes con depresión leve de los pacientes con depresión mínima. El área bajo esta curva ROC (*area under the curve*, AUC) fue 0,82 (error típico= 0,034; intervalo de confianza asintótico al 95%= 0,76-0,87) y fue estadísticamente significativa ($z = 9,45$; $p < 0,0001$). De forma similar, en las figuras 2 y 3 se presentan las curvas ROC elaboradas para distinguir, respectivamente, los pacientes con depresión moderada de los pacientes con depresión mínima o leve, y los pacientes con depresión grave de los pacientes con depresión mínima, leve o moderada. Los AUC de estas dos curvas ROC fueron, respectivamente, 0,86 (error típico= 0,024; intervalo de confianza asintótico al 95%= 0,81-0,90) y 0,88 (error típico= 0,031; intervalo de confianza asintótico al 95%= 0,83-0,92) y ambas fueron también estadísticamente significativas ($z = 14,94$ y $12,08$, respectivamente, con $p < 0,0001$). Es más, las tres AUC superaron el criterio mínimo de exactitud recomendado para un instrumento diagnóstico (AUC > 0,70; Swets, 1988) y alcanzaron valores superiores a 0,80, lo que implicaba que existía más de un 80% de probabilidad de que un paciente seleccionado al azar de la categoría de depresión más grave obtuviera una puntuación directa en el BDI-II superior al de otro paciente seleccionado de la categoría de depresión menos grave.

Figura 1

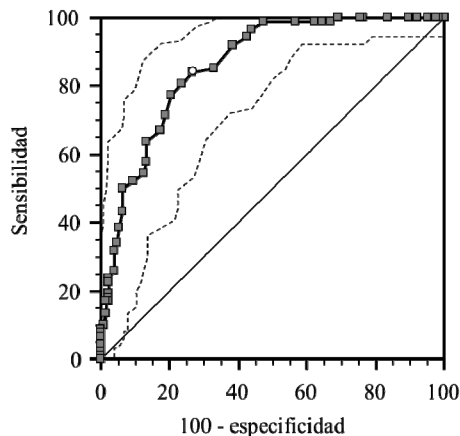
Curva ROC de las puntuaciones del BDI-II para diferenciar los pacientes con depresión leve de los pacientes con depresión mínima



Nota. La diagonal representa una diferenciación aleatoria (AUC= 0,50). Las líneas punteadas representan el intervalo de confianza al 95% de la curva ROC. El punto en blanco en la curva ROC indica el punto en el que apareció un mayor equilibrio entre sensibilidad y especificidad y que correspondió a una puntuación en el BDI-II ≥ 14 .

Figura 2

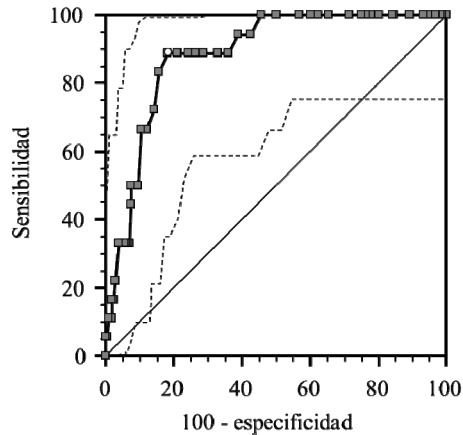
Curva ROC de las puntuaciones del BDI-II para diferenciar los pacientes con depresión moderada de los pacientes con depresión mínima o leve



Nota. La diagonal representa una diferenciación aleatoria (AUC= 0,50). Las líneas punteadas representan el intervalo de confianza al 95% de la curva ROC. El punto en blanco en la curva ROC indica el punto en el que apareció un mayor equilibrio entre sensibilidad y especificidad y que correspondió a una puntuación en el BDI-II ≥ 19 .

Figura 3

Curva ROC de las puntuaciones del BDI-II para diferenciar los pacientes con depresión grave de los pacientes con depresión mínima, leve o moderada



Nota. La diagonal representa una diferenciación aleatoria ($AUC= 0,50$). Las líneas puntuadas representan el intervalo de confianza al 95% de la curva ROC. El punto en blanco en la curva ROC indica el punto en el que apareció un mayor equilibrio entre sensibilidad y especificidad y que correspondió a una puntuación en el BDI-II ≥ 28 .

Las puntuaciones del BDI-II que mostraban un mejor equilibrio entre sensibilidad y especificidad, tal y como mostraban las distintas curvas ROC y los índices Youden (J) resultantes, y que, a su vez, siguiendo las recomendaciones de Beck *et al.* (1996), primaban la sensibilidad frente a la especificidad, fueron las puntuaciones de 14 (figura 1), 19 (figura 2) y 28 (figura 3), con sus respectivos índices de rendimiento diagnóstico: $J= 0,49, 0,57$ y $0,71$; sensibilidad= $0,76, 0,84$ y $0,89$ y especificidad= $0,73, 0,73$ y $0,82$. Estas puntuaciones, además, superaban simultáneamente los valores mínimos de sensibilidad y especificidad ($> 0,70$) que se consideran aceptables para cualquier prueba diagnóstica (Lenderink *et al.*, 2012; Sheehan *et al.*, 1997). Con esas puntuaciones se elaboraron los rangos de puntuaciones de gravedad de la depresión para la adaptación española del BDI-II y, como puede verse en la tabla 3, tales rangos fueron muy similares a los propuestos para la versión original del BDI-II, habiendo una total igualdad para la categoría de depresión mínima y apenas una variación de un punto para las categorías de depresión leve, moderada y grave.

Tabla 3

Puntuaciones de corte para distintos grados de depresión en el BDI-II original y en su adaptación española

Grados de depresión	BDI-II original	Adaptación española del BDI-II
	Basados en las curvas ROC (Beck <i>et al.</i> , 1996, p. 11)	Basados en las curvas ROC (este estudio)
Mínima	0-13	0-13
Leve	14-19	14-18
Moderada	20-28	19-27
Grave	29-63	28-63

Discusión

Los resultados del estudio 1 mostraron, en primer lugar, que entre grupos de pacientes españoles con trastornos psicológicos y con distintos niveles de gravedad de su sintomatología depresiva según las puntuaciones de corte de la versión original del BDI-II (Beck *et al.*, 1996), existían diferencias en la valoración que de la gravedad de su depresión hacían sus psicólogos, diferencias que eran estadísticamente significativas, de un tamaño moderado o grande y coherentes con su nivel de gravedad según su puntuación en el BDI-II.

En segundo lugar, los resultados también mostraron que era posible obtener puntuaciones de corte para la adaptación española del BDI-II que tenían simultáneamente índices adecuados de sensibilidad y especificidad ($> 0,70$) para discriminar entre pacientes españoles que tenían niveles de sintomatología depresiva de distinta gravedad según las valoraciones de sus psicólogos, y que tales puntuaciones de corte eran muy similares a las puntuaciones de corte de la versión original del BDI-II, con una total igualdad para la categoría de depresión mínima y apenas una variación de un punto para las categorías de depresión leve, moderada y grave. Esta similitud fue superior incluso a la encontrada con la propia versión original del BDI-II en el estudio de Sprinkle *et al.* (2002). En este estudio la variación en las puntuaciones de corte fue de 2 puntos para la categoría de depresión leve y de 4 puntos para la categoría de depresión moderada, aunque en este caso el hecho de que el estudio se realizara con una muestra de 137 estudiantes universitarios que recibían tratamiento en un centro de consejo psicológico universitario y, por tanto, probablemente con un nivel de psicopatología menor al de los pacientes del estudio de Beck *et al.* (1996) y del presente estudio, podría explicar las diferencias. De hecho, en el estudio de Sprinkle *et al.* (2002) no se pudieron examinar las puntuaciones de corte para discriminar la depresión grave porque sólo encontraron cuatro estudiantes con ese nivel de depresión. En este sentido, hay que recordar que aunque los pacientes del presente estudio también estaban siendo atendidos en un centro clínico universitario, la Clínica Universitaria de Psicología de la UCM, este centro no sólo presta atención psicológica ambulatoria a estudiantes de la universidad, sino también a personal de la universidad y a personas ajenas a la misma (García-Vera y Sanz, 2009), y, quizás por esa razón, el perfil de pacientes y problemas clínicos que

atiende es comparable en muchos aspectos al que se puede encontrar en centros privados españoles de psicología clínica (Estupiñá, Labrador y García-Vera, 2012). De hecho, la muestra de pacientes del presente estudio obtuvo una puntuación media en el BDI-II de 20 ($DT= 11,4$), muy semejante a la media de 22,8 ($DT= 11,9$) que Sanz y García-Vera (2009) encontraron tras promediar ponderadamente los resultados de 9 estudios (8 realizados en EE. UU. y uno en Argentina) que totalizaban 2.978 pacientes con trastornos psicológicos procedentes de centros públicos o privados de salud mental o psicología o de servicios u hospitales psiquiátricos con consulta externa.

Conjuntamente, esos dos resultados del presente estudio corroboran la validez de las puntuaciones de corte propuestas en el manual original del BDI-II (Beck *et al.*, 1996) y asumidas en el manual de la adaptación española del BDI-II (Beck *et al.*, 2011a), en el sentido de que, en un paciente con trastorno psicológico, sea estadounidense o sea español, se puede considerar que una puntuación entre 0 y 13 indicaría un nivel de sintomatología depresiva mínimo, entre 14 y 19 leve, entre 20 y 28 moderado, y entre 29 y 63 grave.

Que se sepa, éste es el primer estudio que, siguiendo la metodología de las curvas ROC utilizada por Beck *et al.* (1996), ha sometido a comprobación empírica la validez de las puntuaciones de corte de una adaptación del BDI-II. Por tanto, sería deseable que futuros estudios replicaran con la adaptación española del BDI-II o con otras adaptaciones del BDI-II la generalización de sus resultados. Es más, sería también deseable que en esos nuevos estudios la valoración de la gravedad de la depresión realizada por el psicólogo y que servía de criterio para el BDI-II fuera realizada para cada paciente por más de un clínico y que, además, se examinara la fiabilidad interjueces de dichas valoraciones. Mientras tanto, puesto que la generalización de los índices de validación a otros fines y a otras poblaciones no se puede asumir automáticamente, sino que debería confirmarse empíricamente (Colegio Oficial de Psicólogos e ITC, 2000), es importante constatar que, a pesar de las limitaciones del presente estudio, las tan famosas puntuaciones de corte del BDI-II y que tan frecuentemente han sido utilizadas para estimar la gravedad de la sintomatología depresiva de los pacientes españoles con trastornos psicológicos, son efectivamente válidas para este fin y para esta población.

Por otro lado, y de forma complementaria, los resultados del estudio 1 muestran índices adecuados de validez de criterio concurrente para la adaptación española del BDI-II, ya que las puntuaciones medias que obtuvieron en el BDI-II los grupos de pacientes con diferentes niveles de gravedad de su depresión, según las valoraciones de sus psicólogos, eran significativamente diferentes, y tales diferencias eran grandes y coherentes con su nivel de gravedad según la valoración del clínico. Estos resultados son consistentes con los obtenidos por Beck *et al.* (1996, p. 23) con la versión original del BDI-II y con pacientes estadounidenses externos con trastornos psicológicos. De hecho, las puntuaciones medias en el BDI-II para los diferentes niveles de gravedad de depresión (mínimo, leve, moderado y grave) fueron similares en ambos estudios: respectivamente, 10, 16,7, 26,4 y 34,8 en el presente estudio frente a 7,6, 19,4, 27,4 y 32,9 en el estudio de Beck *et al.* (1996).

Estudio 2

Para obtener una tabla de puntuaciones centiles del BDI-II para pacientes españoles con trastornos psicológicos y para actualizar las tablas de puntuaciones centiles para población general y estudiantes universitarios españoles de estudios previos, se analizaron los datos de las muestras de participantes de la adaptación española del BDI-II (Beck *et al.*, 2011a), las cuales en realidad eran muestras ampliadas de las que habían participado en los estudios previos sobre dicha adaptación (Sanz *et al.*, 2005a; Sanz y García-Vera, 2013; Sanz, Navarro *et al.*, 2003; Sanz, Perdigón *et al.*, 2003).

Método

Participantes

En este estudio se analizaron los datos de tres muestras de participantes (Beck *et al.*, 2011a). Una primera muestra formada por 712 pacientes externos españoles con trastornos psicológicos, de los cuales 627 (88%) habían sido atendidos en la Clínica Universitaria de Psicología de la UCM y los restantes 85 (12%) habían sido atendidos en las consultas ambulatorias de psicología de los centros de salud mental vinculados al Hospital Universitario 12 de Octubre de Madrid. La edad de los pacientes oscilaba entre 18 y 79 años ($M= 32,6$; $DT= 12,1$) y la mayoría era mujer (69,9%). El 19,2% de los pacientes tenía un diagnóstico, según el DSM-IV, de trastorno del estado de ánimo (10,7% de trastorno depresivo mayor, 2,7% de trastorno distímico y 5,9% de otro trastorno del estado de ánimo), el 36,4% de trastorno de ansiedad, el 11,7% de trastorno adaptativo y el restante 32,7% de otros trastornos o problemas psicológicos. La segunda muestra estaba formada por 569 adultos de la población general española, de los cuales 470 (82,6%) habían sido reclutados mediante la técnica de la "bola de nieve" por estudiantes de Psicología de la UCM y los restantes 99 (17,4%) habían sido reclutados por psicólogos en varias ciudades de diferentes zonas de España. La edad de los participantes de esta segunda muestra oscilaba entre 17 y 92 años ($M= 42,6$; $DT= 17,7$) y aproximadamente la mitad eran mujeres (52,4%). A pesar de que los procedimientos utilizados para la selección de los participantes de esta segunda muestra no permitían la obtención de una muestra aleatoria, su estratificación demográfica en cuanto a las variables sexo y edad no resultó diferente de forma estadísticamente significativa del perfil demográfico de la población de la Comunidad de Madrid, en el caso de los participantes reclutados por los estudiantes de Psicología de la UCM, y del perfil demográfico de la población de España, en el caso de los participantes reclutados por psicólogos de diferentes zonas de España (véase Beck *et al.*, 2011a). Finalmente, la tercera muestra estaba formada por 727 estudiantes universitarios que cursaban estudios en Madrid, de los cuales 511 (70,3%) estudiaban Psicología, 57 (7,8%) Derecho, 49 (6,7%) Periodismo, 29 (4%) Ingeniería y los restantes 81 (11,1%) otros estudios. La edad de los estudiantes variaba entre 18 y 58 años ($M= 21,9$; $DT= 3,5$) y la mayoría era mujer (80,3%).

Instrumento

- “Inventario de depresión de Beck-II” (BDI-II; Beck *et al.*, 1996), en su adaptación española (Beck *et al.*, 2011a), y que ya fue descrito en el estudio 1.

Procedimiento

Todos los pacientes de la primera muestra fueron atendidos en régimen ambulatorio y de todos ellos se recogió información demográfica y clínica básica, analizándose el BDI-II que completaron en la evaluación inicial individual realizada por un psicólogo en el momento de admisión al centro y durante la cual, además, se aplicaron otros instrumentos que o bien formaban parte del protocolo de evaluación clínica habitual o de algún protocolo de investigación o bien eran específicos para el problema concreto del paciente. A todos los participantes de la muestra de la población general se les aplicó el BDI-II de forma individual y anónima junto a otros test psicológicos, contrabalanceándose el orden de aplicación del BDI-II respecto al de los otros test, de manera que se construyeron varios tipos de cuadernillos y en la mitad de los mismos el BDI-II precedía a los otros cuestionarios y en la otra mitad los seguía. Más de la mitad (58,4%) de los estudiantes universitarios completaron el BDI-II de forma individual y contrabalanceada con otro instrumento de ansiedad y tras solicitar su participación voluntaria y anónima en una investigación sobre ansiedad y depresión. Los restantes estudiantes universitarios, todos ellos estudiantes de la Facultad de Psicología de la UCM, completaron, de forma contrabalanceada, el BDI-II y un instrumento de ansiedad o bien en grupos de 20-30 personas y como parte de las prácticas de una de sus asignaturas (165 estudiantes) o bien en grupos de 40-50 personas y tras solicitar su participación voluntaria y anónima en una investigación sobre ansiedad y depresión (137 estudiantes). Información más detallada sobre los procedimientos de aplicación del BDI-II en las tres muestras se puede consultar en Beck *et al.* (2011a).

Análisis de datos

Para cada una de las muestras de participantes, se calcularon las puntuaciones centiles (o percentiles) del BDI-II con el método por defecto (HAVERAGE) del SPSS. Para examinar la posible existencia de diferencias entre sexos en el BDI-II independientes del efecto de la edad, se calculó, para cada muestra de participantes, la correlación parcial de las puntuaciones directas del BDI-II con el sexo, controlando la edad. Cuando dicha correlación fue estadísticamente significativa, se realizó una prueba *t* de comparación de las submuestras de varones y mujeres, se calculó la correspondiente *d* de Cohen y se calcularon los percentiles del BDI-II para cada una de esas submuestras.

Resultados

En la tabla 4 se presentan las puntuaciones centiles del BDI-II para la muestra de pacientes externos con trastornos psicológicos. En dicha muestra, la correlación parcial de las puntuaciones directas del BDI-II con el sexo fue estadísticamente significativa (r parcial= 0,07; $p= 0,045$), de manera que la puntuación total media en el BDI-II de las 497 pacientes mujeres ($M= 20,53$; $DT= 11,31$) fue significativamente mayor que la de los 214 pacientes varones ($M= 18,70$; $DT= 10,02$), $t(709)= 2,04$; $p= 0,041$; d de Cohen= 0,17. Por tanto, en la tabla 4 también se presentan las puntuaciones centiles del BDI-II para las submuestras de pacientes varones y mujeres.

Tabla 4
Puntuaciones centiles del BDI-II para la muestra total y por sexo de pacientes externos con trastornos psicológicos

Centiles	Total ($N= 712$)	Varones ($n= 214$)	Mujeres ($n= 498$)
1	1	0	2
5	4	3	5
10	6	6	6
20	10	9	11
25	12	12	12
30	13	13	13
40	16	15	17
50	19	18	19
60	22	21	22
70	25	24	25
75	27	26	28
80	29	27	30
90	34	32	37
95	42	36	43
99	48	45	49
M	19,98	18,70	20,53
DT	10,96	10,02	11,31

En la tabla 5 se recogen las puntuaciones centiles del BDI-II para las muestras de adultos de la población general y de estudiantes universitarios. En la muestra de estudiantes universitarios, la correlación parcial del BDI-II con el sexo no fue estadísticamente significativa (r parcial= 0,001; $p= 0,983$), pero en la muestra de la población general dicha correlación sí lo fue (r parcial= 0,11; $p= 0,012$), de forma que, en promedio, las 297 mujeres de la población general ($M= 10,41$; $DT= 8,0$) puntuaban más alto en el BDI-II que los 271 varones ($M= 8,73$; $DT= 7,41$), $t(566)= 2,58$; $p= 0,01$; d de Cohen= 0,22. Por esta razón, la tabla 5 también recoge las puntuaciones centiles del BDI-II para las submuestras de varones y mujeres de la población general.

Tabla 5

Puntuaciones centiles del BDI-II para la muestra total y por sexo de adultos de la población general y para la muestra total de estudiantes universitarios

Centiles	Adultos de la población general			Estudiantes universitarios
	Total (N= 569)	Varones (n= 271)	Mujeres (n= 298)	Total (N= 727)
1	0	0	0	0
5	0	0	0	0
10	1	1	1	1
20	3	2	3	3
25	4	3	4	3
30	5	4	5	4
40	6	6	7	5
50	8	7	9	7
60	10	9	11	9
70	12	11	14	11
75	14	12	16	12
80	16	14	17	14
90	19	18	22	18
95	25	23	26	24
99	36	36	36	32
Media	9,61	8,73	10,41	8,75
DT	7,76	7,41	8,00	7,34

Discusión

En el estudio 2 se han elaborado tablas de puntuaciones centiles del BDI-II para pacientes españoles con trastornos psicológicos, para adultos españoles de la población general y para estudiantes universitarios españoles, y, en el caso de las dos primeras poblaciones, puesto que se han encontrado diferencias estadísticamente significativas entre varones y mujeres, también se han elaborado tablas de puntuaciones centiles para cada sexo por separado. Es verdad que hubiera sido deseable que, para la elaboración de esas tablas, se hubieran seleccionado muestras mediante un procedimiento aleatorio, no incidental y, en todo caso, de procedencia geográfica más heterogénea. Es más, también hubiera sido conveniente contar con muestras de mayor tamaño. Sin embargo, los tamaños de las muestras utilizadas en su elaboración (712, 569 y 727 participantes para las muestras de pacientes, adultos de la población general y estudiantes universitarios, respectivamente) pueden valorarse, de acuerdo a los estándares propuestos por Prieto y Muñiz (2000), como moderados ($300 \leq N < 600$) o grandes ($600 \leq N < 1000$). Por tanto, a pesar de las limitaciones señaladas, las tablas de puntuaciones centiles de la adaptación española del BDI-II que aquí se ofrecen pueden servir como normas o baremos para interpretar las puntuaciones en el BDI-II de las personas evaluadas que puedan considerarse parte de las tres poblaciones señaladas.

Tan sólo se ha encontrado un estudio que haya elaborado normas centiles del BDI-II para algunos de los tipos de poblaciones aquí estudiadas. En concreto, Roelofs *et al.* (2013) elaboraron normas para la población general holandesa, encontrando, por ejemplo, que los percentiles 60, 80 y 95 correspondían a las puntuaciones 11, 18 y 32, respectivamente, todas ellas más altas que las correspondientes a los mismos percentiles en la muestra de la población general española del presente estudio (10, 16 y 25, respectivamente), lo cual subraya la necesidad de obtener normas y criterios para cada población con la que se pretenda utilizar un instrumento, en este caso, el BDI-II, en lugar de asumir la posible generalización de tales normas y criterios.

Discusión general

El objetivo general de los dos estudios de este trabajo era proporcionar información que ayudara a los profesionales e investigadores españoles que utilizan el BDI-II a interpretar mejor las puntuaciones que ofrece el instrumento. Tradicionalmente, las puntuaciones del BDI-II, al igual que ha ocurrido con las puntuaciones de sus versiones previas (BDI-I y BDI-IA), han sido interpretadas a partir de ciertos criterios o puntuaciones de corte que indican distintos niveles de gravedad de la sintomatología depresiva, en particular, las puntuaciones de corte propuestas por Beck *et al.* (1996) en el manual original del inventario. Estas puntuaciones de corte también fueron asumidas por la adaptación española del BDI-II (Beck *et al.*, 2011a), aunque no existía ningún estudio que hubiera comprobado empírica y específicamente la plausibilidad de esta asunción. Los resultados del Estudio 1 del presente trabajo permiten cubrir esa laguna en la literatura científica al corroborar la validez de las puntuaciones de corte del manual original del BDI-II para distinguir distintos niveles de gravedad de la sintomatología depresiva en pacientes españoles con trastornos psicológicos evaluados mediante la adaptación española del BDI-II.

Sin embargo, la interpretación de las puntuaciones en el BDI-II también puede hacerse comparando dichas puntuaciones con los baremos o normas de un grupo de referencia, las cuales si se expresan en forma de puntuaciones centiles tienen varias ventajas para los usuarios de los test debido a que son fáciles de interpretar y tienen un significado universal. En el estudio 2 del presente trabajo se han elaborado tablas de puntuaciones centiles del BDI-II para pacientes españoles con trastornos psicológicos, para adultos españoles de la población general y para estudiantes universitarios españoles, las cuales pueden ayudar a los clínicos e investigadores españoles a evaluar con mayor precisión la gravedad de la depresión con la adaptación española del BDI-II.

Ambas formas de interpretar las puntuaciones del BDI-II, la basada en las puntuaciones de corte y la basada en las normas o baremos, son complementarias y deberían combinarse para obtener una visión más adecuada del nivel de sintomatología depresiva de una persona. Así, aunque dos pacientes españoles con trastornos psicológicos pueden puntuar en el BDI-II en el rango que indica un nivel moderado de depresión, por ejemplo, 22 y 27, una consulta a la tabla 4 del presente trabajo sugeriría que la primera puntuación indica también que el primer

paciente sufre una sintomatología depresiva más grave que el 60% de los pacientes españoles con trastornos psicológicos, mientras que la segunda puntuación indica que el segundo paciente sufre una sintomatología depresiva más grave que el 75% de los pacientes españoles con trastornos psicológicos. En la medida que el BDI-II se ha convertido no sólo en el instrumento más utilizado para evaluar la sintomatología depresiva en población clínica, sino también en población no clínica, la posibilidad de contar con normas centiles para adultos españoles de la población general y para estudiantes universitarios españoles permite apreciar mucho mejor las diferencias individuales en depresión en las poblaciones no clínicas, ya que en ellas es más difícil encontrar personas que superan las puntuaciones de corte del BDI-II correspondientes a una depresión moderada o grave y, por tanto, los percentiles del grupo de referencia correspondiente permitirían apresar pequeñas diferencias individuales dentro los niveles mínimos y leves de gravedad de la depresión. Por otro lado, expresar la puntuación de una persona en el BDI-II en términos de su percentil, en lugar de simplemente indicar si está por encima o por debajo de una determinada puntuación de corte, está en consonancia con la idea de que la depresión debe entenderse como un constructo dimensional en lugar de como un constructo categorial, una idea que cuenta con gran aval empírico (Flett, Vredenburg y Krames, 1997; Prisciandaro y Roberts, 2009; Ruscio y Ruscio, 2000).

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.
- 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. (2011a). *Manual. BDI-II. Inventario de depresión de Beck-II*. Madrid: Pearson.
- Beck, A. T., Steer, R. A. y Brown, G. K. (2011b). *Manual. BDI-FS. Inventario de depresión de Beck para pacientes médicos*. Madrid: Pearson.
- 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.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2ª ed. Hillsdale, NJ: LEA.
- Colegio Oficial de Psicólogos y Comisión Internacional de Tests (ITC) (2000). Directrices internacionales para el uso de los tests. *Infocop*, 77, 21-32.
- 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.
- Conde, V., Esteban, T. y Useros, E. (1976). Revisión crítica de la adaptación castellana del Cuestionario de Beck. *Revista de Psicología General y Aplicada*, 31, 469-497.
- Estupiñá, F., Labrador, F. J. y García-Vera, M. P. (2012). A study of patients who go to a psychology clinic seeking treatment. *Spanish Journal of Psychology*, 15, 275-285.
- Flett, G. L., Vredenburg, K. y Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin*, 121, 395-416.

- García-Vera, M. P. y Sanz, J. (2009). La Clínica Universitaria de Psicología de la Universidad Complutense de Madrid: historia, características y situación actual. *Acción Psicológica*, 6, 123-134.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, 278-296.
- Lenderink, A. F., Zoer, I., van der Molen, H. F., Spreeuwiers, D., Frings-Dresen, M. H. y van Dijk, F. J. (2012). Review on the validity of self-report to assess workrelated diseases. *International Archives of Occupational and Environmental Health*, 85, 229-251.
- 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 (dirs.), *Handbook of depression* (2ª ed.) (pp. 44-68). Nueva York: Guilford.
- Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Prisciandaro, J. J. y Roberts, J. E. (2009). A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. *Psychological Medicine*, 39, 1087-1096.
- Radloff, L. (1977). The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Roelofs, J., van Breukelen, G., de Graaf, L. E., Beck, A. T., Arntz, A. y Huibers, M. J. H. (2013). Norms for the Beck Depression Inventory (BDI-II) in a large Dutch community sample. *Journal of Psychopathology and Behavioral Assessment*, 35, 93-98.
- Ruscio, J. y Ruscio, A. M. (2000). Informing the continuity controversy: a taxometric analysis of depression. *Journal of Abnormal Psychology*, 109, 473-487.
- Sanz, J. y García-Vera, M. P. (2007). Análisis psicométrico de las versiones breves del «Inventario para la depresión de Beck» de 1978 (BDI-IA). *Psicología Conductual*, 15, 191-214.
- 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. Řehulka (dir.), *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 de depresión de Beck-II (BDI-II). *Anales de Psicología*, 29, 66-75.
- Sanz, J., García-Vera, M. P., Espinosa, R., Fortún, M. y Vázquez, C. (2005a). 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., García-Vera, M. P., Espinosa, R., Fortún, M. y Vázquez, C. (2005b). *Listado para el episodio depresivo (LED)*. Documento no publicado, Universidad Complutense de Madrid, España.
- 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., Perdigón, 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. (1998). Fiabilidad, validez y datos normativos del Inventario para la depresión de Beck. *Psicothema*, 10, 303-318.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M. F. y Dunbar G. C. (1997). The validity of the Mini International

- Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, 12, 232-241.
- Sprinkle, S., Lurie, D., Insko, S., Atkinson, G., Jones, G., Logan, A. R. y Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of Counseling Psychology*, 49, 381-385.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- 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.
- Vázquez, C. y Sanz, J. (1999). Fiabilidad y validez de la versión española del Inventario para la depresión de Beck de 1978 en pacientes con trastornos psicológicos. *Clínica y Salud*, 10, 59-81.
- Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons.
- Yela, M. (1984). *Psicometría y teoría de los tests*. Manuscrito no publicado, Universidad Complutense de Madrid, España.
- Zung, W.W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.

RECIBIDO: 5 de mayo de 2013

ACEPTADO: 19 de septiembre de 2013

APÉNDICE**LISTADO PARA EL EPISODIO DEPRESIVO (LED)**
(Sanz, García-Vera, Espinosa, Fortún y Vázquez, 2005)

Por favor, complete la siguiente escala después de haber tenido una entrevista con el/la paciente sobre sus problemas de una duración mínima de 45 minutos. Si al contestar alguna pregunta no dispone de suficiente información o no está seguro/a, marque “¿?”.

A. DURANTE LAS ÚLTIMAS 2 SEMANAS EL/LA PACIENTE:

1. ¿Se ha sentido casi todos los días triste o deprimido/a? Sí No ¿?
2. ¿Perdió el interés por las cosas que le interesaban o ya no disfrutaba de ellas? Sí No ¿?

SI HA MARCADO NO A LAS DOS PREGUNTAS ANTERIORES PASE A LA SECCIÓN B**DURANTE LAS ÚLTIMAS 2 SEMANAS EL/LA PACIENTE:**

3. ¿Tuvo cambios significativos en su apetito o peso? Sí No ¿?
4. ¿Tuvo cambios significativos en su patrón de sueño? Sí No ¿?
5. ¿Hablaba o se movía más lento/a o estaba más agitado/a de lo normal para él/ella? Sí No ¿?
6. ¿Le faltaban energías o se sentía cansado/a todo el tiempo? Sí No ¿?
7. ¿Se sentía inútil, inferior a los demás, culpable o lleno/a de faltas? Sí No ¿?
8. ¿Tuvo dificultades para pensar, concentrarse o tomar decisiones? Sí No ¿?
9. ¿Pensó mucho en la muerte o sintió que quería morirse? Sí No ¿?
10. ¿Estaba tomando alguna sustancia (alcohol, drogas, medicamento) que pudiera afectar a su estado de ánimo? Sí No ¿?
• ¿Qué sustancias? _____
11. ¿Sufrió recientemente la muerte de algún ser querido? Sí No ¿?
• ¿Hace cuántos meses? _____
12. ¿Sufrió malestar significativo o deterioro en las áreas social, laboral u otras importantes debido a sus síntomas? Sí No ¿?

B. SI CREE QUE EL/LA PACIENTE TIENE ALGÚN TIPO DE PROBLEMA DEPRESIVO (AUNQUE NO SE CORRESPONDA CON NINGÚN DIAGNÓSTICO), ¿CUÁL ESTIMA QUE ES LA GRAVEDAD DE SU DEPRESIÓN?

- Sin depresión/Mínima Leve Moderada Grave