*Papeles del Psicólogo, 2013. Vol. 34(3), pp. 161-168*
http://www.papelesdelpsicologo.es



# 50 YEARS OF THE BECK DEPRESSION INVENTORY: RECOMMENDATIONS FOR USING THE SPANISH ADAPTATION OF THE BDI-II IN CLINICAL PRACTICE

**Jesús Sanz**
*Complutense University of Madrid*

*The first Spanish adaptation of the Beck Depression Inventory-II (BDI-I) was published in 2011, which happened to be the 50th anniversary of the publication of its first edition. In this time, the BDI has become the most widely-used self-report questionnaire in Spain and in the world for evaluating the severity of depression. In this study, the basic characteristics of the BDI-II are presented and compared with its earlier versions (BDI-I and BDI-IA), the process of its adaptation to the Spanish population is described, the psychometric properties of this adaptation are summarized and its use in clinical practice is discussed. Concerning this use, guidelines and cut-off scores are proposed for measuring the severity of depression, evaluating the clinical significance of therapeutic change, screening for depression and assisting in the differential diagnosis of depressive disorders.*
**Keywords**: *Depression, Assessment, BDI-II, Clinical significance, Screening, Diagnosis*

*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 (Beck, Steer and Brown, 2011), internationally known by its acronym, the BDI-II, was published in 2011 (Beck Depression Inventory-Second Edition; Beck, Steer and Brown, 1996). Curiously, that year was the 50th anniversary of the publication of its first edition (Beck, Ward, Mendelson, Mock and Erbaugh, 1961), date which established the BDI as a classic among depression evaluation instruments and also reflects its adaptation to the most current conceptions of this construct. This article presents the basic characteristics of the BDI-II, summarizes the process of its adaptation to the Spanish population and the psychometric properties of this adaptation, and its use in clinical practice for evaluating the progress of therapy, screening and diagnosing depression is discussed.

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

## 1961-2011: 50 YEARS OF THE BECK DEPRESSION INVENTORY

In the 50 years that have gone by since its first publication, the BDI, in its original version (BDI-I), the revised version (BDI-IA; Beck, Rush, Shaw and Emery, 1979) and its second edition (BD-II), has become the most widely used instrument for assessing depression in Spain and in the world. The bibliographic data in PsycINFO and Psicodoc (see Table 1) show that the number of studies published in recent years on the BDI surpasses by far those published on the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960), the Center for Epidemiologic Studies Depression Scale (CESD; Radloff, 1977) or the Zung Self-Rating Depression Scale, 1965), three of the instruments with the best reputation in the field of depression (Nezu, Nezu, Friedman and Lee, 2009). Moreover, Muñiz and Fernández-Hermida (2010) found that the BDI is the fifth most widely-used test by Spanish psychologists in their daily practice, thereby consolidating it in the same position it had ten years before (Muñiz and

Fernández-Hermida, 2000), which is not surprising, since all its versions have been translated into Spanish and validated in Spain: BDI-I (Conde and Useros, 1974), BDI-IA (Vázquez and Sanz, 1997) and BDI-II (Beck et al., 2011).

It is hard to determine why it is so popular. The fact that the original BDI was published as an annex to an article in a widely-read scientific journal with an excellent reputation, and that version could be used for free may have initially contributed to its popularity. However, the two more recent versions (BDI-IA and BDI-II) have had all rights reserved since the beginning, and were marketed by a publisher. Furthermore, other self-rating instruments for depression, such as the SDS, are about as old, were originally published in the same journal or others with similar distribution and reputation, and could have been used for free, and still can be, however, they have never been as popular as the BDI. Other reasons for its popularity lie in the growing prestige acquired by its creator, Aaron Beck, and the gradual diffusion and success of his cognitive theory and therapy of depression. The BDI was adopted as the standard instrument in numerous studies designed to test cognitive depression models, including the Beck theory, which dominated the field of psychological theory in this field in the nineteen-eighties and nineties, and still does (Sanz and Vázquez, 2008; Tennen, Hall and Affleck, 1995), and the many studies done on the effectiveness of Beck's cognitive therapy, and of psychotherapy in general, for depression, studies which also began to proliferate at that time and have continually increased to date (Cuijpers, Li, Hofmann and Andersson, 2010). Finally, the BDI characteristics themselves (e.g., brief, easy application, scoring and interpretation) and its good psychometric properties have also contributed to its popularity, although perhaps only partially, since there other instruments, such as the CESD or the SDS which have similar characteristics and psychometric properties and are not, however, as popular.

**TABLE I**
**NUMBER OF STUDIES IN THE PSYCINFO AND PSICODOC DATABASES ON THE MAIN DEPRESSION ASSESSMENT INSTRUMENTS**

| Instrument | PsycINFO (2002-2013)* | Psicodoc··· |
|---|---|---|
| Beck Inventory of Depression | 18723 | 352 |
| Hamilton Depression Rating Scale | 9528 | 109 |
| Center for Epidemiologic Studies Depression Scale | 8505 | 67 |
| Zung Self-Rating Depression Scale | 1150 | 50 |

*Note. Searches on May 1, 2013 and limited to 2002 to 2013, with the following combinations of words in the field "tests and measures": "Beck Depression Inventory"; "Hamilton Rating Scale for Depression" or "Hamilton Depression Rating Scale"; "Self-Rating Depression Scale" or "Zung Depression Scale"; "CES-D" or "Center for Epidemiologic Studies Depression".
···Searches made on May 1, 2013 with the following combinations of words in the field "simple search": "inventory and Beck and depression" or "BDI and depression"; "Hamilton and scale and depression" or "HRSD and depression" or "HDRS and depression" or "HAM and depression"; "Zung and scale and depression" or "Zung and SDS" or "SDS and depression"; "CES-D and depression" or "scale and depression and center and studies and epidemiological".

## THE DIFFERENT VERSIONS OF THE BDI: BDI-I, BDI-IA AND BDI-II

Many of the BDI characteristics have remained constant in all three versions (see Table 2). For example, all of them contain 21 items, each of which is comprised of several statements about the same depressive symptom in order from least to most severe. The person assessed must choose the one that describes his condition best. In each item, the choices are scored from 0 to 3, so the inventory score range is 0-63. However, the content of the three versions is quite different (e.g., the time frame of the instructions, item content, number of choices on each item) which have led to differences in interpretation of their scores (e.g., the cut-off scores for different levels of severity of depression are slightly higher in the BDI-II than on the BDI-I) or their psychometric properties (e.g., the indices of internal consistency and factor validity of the BDI-II are higher than the BDI-I or BDI-IA; Beck et al., 2011; Botella and Ponte, 2011; Sanz and García-Vera, 2009) (see Table 2).

The BDI-II has been substantially modified from the BDI-I and BDI-IA to cover all the DSM-IV symptomatic diagnostic criteria for depressive disorders (APA, 1995) (see Table 2). Specifically, the BDI-II instructions ask about the state of the person assessed during the last two weeks instead of the last week (BDI-IA) or the same day on which the instrument is filled in (BDI-I). Thus, the BDI-II instructions assess the presence of depressive symptomology during the minimal period necessary for diagnosis of a major depressive episode according to the DSM-IV. In addition, four items on the BDI-IA were omitted from the BDI-II and replaced by others that evaluate DSM-IV diagnostic criteria symptoms, which furthermore, are typical of severe depression (agitation, worthlessness, concentration difficulty, loss of energy). Items on loss of appetite and insomnia are also modified in the BDI-II so that increases in appetite and sleep can

be assessed, not only decreases, and modifications have been made in some of the statements in other items. For example, the BDI-IA item on lack of social interest has been changed to also include lack of interest in people or activities, so that this BDI-II item assesses more appropriately DSM-IV symptomatic criteria on loss of interest in all or almost all activities. In fact, there are only three items on the BDI-IA that were not modified for the BDI-II (punishment feelings, suicidal thoughts and loss of interest in sex), and it is therefore an important revision of the BDI-I and BDI-IA.

All these modifications seem to have achieved their purpose. An analysis of content validity of the most popular self-report questionnaires for evaluating clinical depression in adults in Spain revealed that the BDI-II was the only instrument that covered all the DSM-IV symptoms that define major depression episode and dysthymic disorder, and moreover, its items reflect only depressive symptoms, so its scores are not influenced by symptoms not involved in the depression construct (Sanz et al., 2013; see Table 3).

## THE SPANISH ADAPTATION OF THE BDI-II

Given the improvements in the BDI-II over previous versions, a research group at the Complutense University of Madrid began studies for its validation in Spain in 2000 (Sanz, Navarro and Vázquez, 2003; Sanz, Perdigón and Vázquez, 2003; Sanz, García-Vera, Espinosa, Fortún and Vázquez, 2005; Sanz and García-Vera, 2009, 2013), which culminated in the Spanish adaptation of the instrument by that research group in collaboration with *Pearson Clinical & Talent Assessment España* who also published it (Beck et al., 2011).

The Spanish adaptation of the BDI-II is based on an exhaustive analysis of its psychometric properties of internal consistency reliability, convergent and discriminant validity, factor validity, and criterion validity in three samples of patients with psychological disorders totaling 712 patients, two samples of adults from the general population totaling 569 adults and two samples of university students totaling 727 students. These analyses consistently suggest in all the samples, that the Spanish adaptation of the BDI-II has good psychometric

**TABLE 2**
**MAIN BDI VERSION SIMILARITIES AND DIFFERENCES**

| Characteristics | BDI or BDI-I | BDI-IA | BDI-II |
|---|---|---|---|
| Year published | 1961 | 1979 (1978)[a] | 1996 |
| Nº of items | 21 | 21 | 21 |
| Nº choices per item | 10 items = 4<br>8 items = 5<br>1 item = 6 | All items = 4 | 19 items = 4<br>2 items = 7 |
| Nº of items modified from the previous version | - | 15 | 18 |
| Instruction time frame | Today | Last week | Last two weeks |
| Range of item scores/total score | 0-3/0-63 | 0-3/0-63 | 0-3/0-63 |
| Cut-off scores for severity of depression[b] | Normal = 0-9<br>Mild = 10-15<br>Moderate = 16-23<br>Severe = 24-63 | Minimal = 0-9<br>Mild = 10-16<br>Moderate = 17-29<br>Severe = 30-63 | Minimal = 0-13<br>Mild = 14-19<br>Moderate = 20-28<br>Severe = 29-63 |
| Spanish adaptation | Conde et al. (1974) | Vázquez and Sanz (1991, 1997) | Sanz and Vázquez (Beck et al., 2011) |
| % coverage of major DSM-IV depression symptoms[c] | 88.9% | | 100% |
| % coverage of DSM-IV dysthymia symptoms[c] | 100% | | 100% |
| % coverage of major CIE-10 depression symptoms[c] | 90% | | 100% |
| % coverage of CIE-10 dysthymia symptoms[c] | 75% | | 75% |
| % of items that assess depressive symptoms[c] | 95.2% | | 100% |
| Reliability of internal consistency[d] | 18-49 yrs old = .85<br>People > 50 yrs old = .80<br>Patients = .86 | | University students = .87<br>Adults = .88<br>Patients = .91 |
| Factor validity[e] | More variable internal structure (2, 3, 4 or more factors) | | More consistent internal structure (usually 2 factors) |

*Note.* [a]Although the BDI-IA was copyrighted in 1978, it was published for the first time by Beck et al. (1979).
[b]The BDI-I cut-off scores appeared in Loeb, Feshbach, Beck and Wolf (1964; see also Bumberry, Oliver and McClure, 1978)
[c]Adapted from Sanz et al. (2013).
[d]Reliability data on the BDI-I/BDI-IA for adults of different ages and for patients are from meta-analytical studies by Yin and Fan (2000) and Beck, Steer and Garbin (1988), respectively, while those for the BDI-II were calculated based on data in Beck et al. (2011), after calculating the mean, weighted by participant sample size, from 10 universities totaling 15,123 students
[e]See Beck et al. (1988), Beck et al. (2011) and Sanz and García-Vera (2009).

indices for measuring the presence and severity of depressive symptomology, indices similar to those in the original version or the adaptations made in other countries (Beck et al., 2011). For the purposes of this study, some of these indices should be mentioned.

The Cronbach internal consistency reliability coefficients for patients with psychological disorders, adults from the general population and university students were .90, .87 and .89, respectively, all of which are excellent ($^3$ .85; Prieto and Muñiz, 2000) and comparable to those found in studies with similar samples in other countries (see Table 4).

Concerning the criterion validity for distinguishing between contrasted groups, the total mean score for the BDI-II for patients with psychological disorders was significantly higher than for adults in the general population or university students (Beck et al., 2011). In fact, the overall severity of depression in patients was about double found in the non-clinical samples (see Table 5), which means, in terms of effect size, standardized mean differences (Cohen *d*) equal to 1.07 and 1.21, respectively, which may be considered large (*d* > 0.80; Cohen, 1988).

With respect to diagnostic validity, an analysis of the receiver operating characteristic curve or ROC curve, revealed adequate BDI-II diagnostic performance (area under the ROC curve = .83; adequate > 0.70; Swets, 1988) in detecting a major depressive disorder compared to a non-depressive mental disorder in a sample of patients with psychological disorders, as assessed by a structured diagnostic interview and a depression checklist (Beck et al., 2011). The diagnostic discrimination capacity of any instrument depends on the characteristics of the population on which it is used, so the differential diagnosis in a sample of patients with psychological disorders entails more difficulty than if it were intended to find people with depression in a nonclinical sample. For this, the ROC curve analysis showed high BDI-II diagnostic performance (area under the ROC curve = .91; high > 0.90; Swets, 1988) for discriminating between university students with a major depression episode and students who were not depressed assessed by a structured diagnostic interview. These two analyses of diagnostic performance enabled identification of several cut-off points in the BDI-II with adequate indices of diagnostic validity for finding persons with clinical depression (e.g., sensitivity and specificity > 70%; positive and negative predictive values > 50%; kappa > .40; see Table 6).

**TABLE 3**
**CORRESPONDENCE BETWEEN ITEMS IN THE MAIN DEPRESSION QUESTIONNAIRES, SCALES AND INVENTORIES AND DSM-IV DIAGNOSTIC CRITERIA FOR DEPRESSION (ADAPTED FROM SANZ ET AL., 2013)**

| Instruments | % DSM-IV major depression symptoms that are covered by items | % DSM-IV dysthymic symptoms that are covered by items | % items that assess depressive symptoms |
|---|---|---|---|
| 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% |

*Note.* BDI-II: Beck Depression Inventory-II; CES-D: Center for Epidemiologic Studies Depression Scale; MCMI-II/CC and MCMI-III/CC: Millon Clinical Multiaxial Inventory Major Depression Scale; MCMI-II/D and MCMI-III/D: Millon Clinical Multiaxial Inventory Dysthymia Scale; MMPI-D and MMPI-2/D: Minnesota Multiphasic Personality Inventory Depression Scale; MMPI-2/DEP: Depression Scale based on content of Minnesota Multiphasic Personality Inventory-2; SCL-90-R/D: Symptom Checklist 90-Revised Depression Subscale; SDS: Zung Self-Rating Depression Scale.

**TABLE 4**
**RELIABILITY OF INTERNAL CONSISTENCE (CRONBACH'S ALPHA) OF THE BDI-II IN DIFFERENT SAMPLES AND WITH DIFFERENT ADAPTATIONS (ADAPTED FROM BECK ET AL., 2011)**

| Population/BDI-II adaptation | *N* | *alpha* |
|---|---|---|
| Outpatients with mental disorders | | |
| Original or adaptations to other countries | 2636 | .92 |
| Spanish adaptation[a] | 712 | .90 |
| University students | | |
| Original or adaptations to other countries | 14396 | .87 |
| Spanish adaptation[a] | 727 | .89 |
| Adults from the general population | | |
| Original or adaptations to other countries | 3519 | .88 |
| Spanish adaptation[a] | 569 | .87 |

*Nota.* [a]Mean coefficient of the samples on the Spanish adaptation of the BDI-II (Beck et al., 2011) after weighting them by sample size.

**TABLE 5**
**MEANS AND STANDARD DEVIATIONS OF BDI-II SCORES ON DIFFERENT SPANISH SAMPLES (ADAPTED FROM BECK ET AL., 2011)**

| | *N* | *M* | *SD* |
|---|---|---|---|
| Outpatients with mental disorders | 712 | 19.98 | 10.96 |
| University students | 727 | 8.75 | 7.34 |
| Adults from the general population | 569 | 9.61 | 7.76 |

## RECOMMENDATIONS FOR USING THE SPANISH ADAPTATION OF THE BDI-II IN CLINICAL PRACTICE

BDI score interpretation is usually based on cut-off scores defining different depressive symptom severity levels. The original BDI-II manual proposes the following: 0-13 shows minimal depression, 14-19 mild depression, 20-28 moderate depression and 29-63 severe depression (Beck et al., 1996). These scores were also used in the Spanish adaptation of the BDI-II, and an empirical study has recently shown its validity for distinguishing levels of severity of depression in Spanish patients with psychological disorders (Sanz, Gutiérrez, Gesteira and García-Vera, 2014).

Complementing this, BDI-II scores can be interpreted by comparing them with norms from a reference group. Sanz et al. (2014) provide tables of BDI-II percentile scores for patients with psychological disorders, adults in the general population and Spanish university students, which enable more precise assessment of the severity of depression.

Based on the psychometric properties of the Spanish adaptation of the BDI-II, interpretation guidelines may be proposed for two very important clinical goals: assessing the clinical significance of therapeutic changes and identifying persons with clinical depression.

## ASSESSING THE CLINICAL SIGNIFICANCE OF THERAPEUTIC CHANGE

Since the BDI contains only 21 items which can be answered in 5-10 minutes and they can be scored so easily and quickly that it can be done in 1-2 minutes, it is possible to apply the BDI repeatedly during the therapy to monitor patient progress and evaluate results of therapy. In this context, the BDI-II can assist in answering the question of whether the patient's depression is the same, has improved or has worsened or even if he as recovered, that is, if there has been a clinically significant reduction or increase in his depressive symptomology.

The statistical approach to clinical significance by Jacobson and Truax (1991), which assumes that a clinically significant change would mean the patient, who before therapy belonged to a dysfunctional population, returns to a functional population, can be used to do this. That is, the change would mean that the patient's BDI-II score is no longer in the BDI-II score distribution of a dysfunctional population (e.g., Spanish patients with psychological disorders) but the distribution of a functional population (e.g., the general Spanish population).

To find out whether there has been a clinically significant change in a patient, the Jacobson and Truax method (1991; McGlinchey, Atkins and Jacobson, 2002) involves, in the first place, establishing a cut-off score which the patient must meet to go on from a dysfunctional distribution to a functional one. When the two distributions overlap, as is the case in the BDI-II (see Table 5), the best cut-off score (*C*) is the midpoint weighted between the functional and dysfunctional means:

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

where $DT_n$ and $DT_p$ represent the standard deviations for the BDI-II of the normal population and patients, respectively, and $M_n$ and $M_p$ are the means for the BDI-II of the normal population and patients, respectively.

In the second place, the method involves estimating whether the change indicated by the scores on an instrument such as the BDI-II is not due to measurement error, but reliably shows a real change in the patient's depressive symptomology. A reliable change index is proposed for this (*RCI*) which takes the standard error of the difference between the two scores on the instrument ($S_{dif}$), depending on its standard error of measurement (Se), which in turn, depends on its reliability ($r_{xx}$):

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

**TABLE 6**
**DIAGNOSTIC PERFORMANCE INDICES OF BDI-II SCORES IN FINDING PERSONS WITH DEPRESSIVE DISORDERS (ADAPTED FROM BECK ET AL., 2011)**

| BDI-II Score | Sensitivity | Specificity | Positive predictive value | Negative predictive value | Kappa |
|---|---|---|---|---|---|
| Sample of university students (*N=165*) | | | | | |
| 12 | 92.94 | 84.15 | 35.27 | 99.23 | .44 |
| 19 | 57.14 | 96.72 | 61.54 | 96.09 | .56 |
| Sample of outpatients with mental disorders (*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 |

$$s_{dif} = \sqrt{2(s_e)^2} = \sqrt{2\left(s_x\sqrt{(1 - r_{xx})}\right)^2}$$

Where $x_2$ would be the score on the BDI-II of a patient at any given time (e.g., post treatment), $x_1$ the score on the BDI-II at any previous given moment (e.g., pre-treatment), and $s_x$ the standard deviation of the scores on the BDI-II in the patient population and $r_{xx}$ the reliability of internal consistency of the BDI-II in this population. The standard error of the difference between the two scores ($S_{dif}$) would describe the distribution amplitude of the change scores that would be expected if there were no real change, so an RCI over 1.96 would not be very likely (p< .05), which would happen if there were no real change. Each patient, then, should pass that *RCI* to demonstrate that his change is not due to BDI-II measurement errors:

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

Based on these two criteria, the Jacobson-Truax method (McGlinchey et al., 2002) classifies a patient as *recovered* if his score on an instrument has surpassed both the *RCI* of 1.96 and the C score, *improved* if the score has surpassed the RCI of 1.96, but not the C score, *unchanged* if the score has not surpassed either of the two criteria, and *worsened* if the score has surpassed the *RCI* of 1.96, but in the direction of worsening.

With the data from the Spanish adaptation of the BDI-II (see Tables 4 and 5) the *C* score and the change with an *RCI* of 1.96 would be:

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

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

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

Therefore, with this adaptation, a patient whose score on the BDI-II goes down 10 points or more and is less than 14 may be considered *recovered* from his depression; if his score goes down 10 points or more, but does not reach 14, he may be considered *improved*; if his score does not go down 10 points and does not reach 14, he may be considered *unchanged*, and if the score shows a 10-point or more increase he could be considered to have *worsened*.

However, it is important to remember that evaluation of the patient's condition should take other criteria about his depression into account (e.g., to talk about recovery, the score would have to be < 14 on the BDI-II for at least two months) and relevant information on any other patient problems and disorders, such as how well he functions on the job or in his usual social activities, or relates to others, etc.

## IDENTIFYING PERSONS WITH CLINICAL DEPRESSION

Although BDI-II is not an instrument designed to diagnose depressive disorders, the information it provides can be helpful in identifying persons with a probable depressive disorder for later confirmation with a more adequate instrument (e.g., a structured diagnostic interview) or serve as a source of additional information that could assist in diagnosing a depressive disorder. The diagnostic performance indices in Table 6 can be used for this. Thus when the Spanish adaptation of the BDI-II is used for screening with later confirmation of the diagnosis in cases identified by the inventory, it is recommended that highly sensitive cut-off scores be chosen. That is, one that is able to detect a depressive disorder in patients who really have it to ensure that no patient who has this disorder remains undetected. In this case, it is preferable to commit a false positive error, since this error would then be easy to correct by later diagnostic confirmation, than a false negative error, since a person with a depressive disorder would not be reevaluated and would go unnoticed. For the nonclinical Spanish population, the cut-off would be equal to or higher than 12, since this score, with an adequate specificity index (>70%) and diagnostic concordance (kappa > .40), could detect 93% of the persons who have a major depressive episode (see Table 6). For the Spanish population of patients with psychological disorders, although the cut-off score of 15 would have a sensitivity of 93%, its specificity and diagnostic concordance would not be adequate (<70% and < .40, respectively; see Table 6), so it would be more recommendable to use a cut-off score of 22, which could detect 82% of patients with depressive disorders showing adequate specificity (74%) and diagnostic concordance (.42) at the same time.

When the Spanish adaptation of the BDI-II is used to assist in diagnosing a depressive disorder suspected from information from other instruments, it is recommendable to select highly specific cut-off scores, that is, with a high capacity for finding absence of depressive disorder in persons who really do not have it, ensuring that a patient who is diagnosed with it really does have the disorder (it is preferable to commit a false negative error than false positive). In this regard, it would be recommendable for this score to have a positive predictive value over 50% since it indicates that for a person with  that score the probability of having a truly major depressive disorder is over 50%. For nonclinical and clinical Spanish populations, those cut-off scores would be equal to or higher than 19 and 30 respectively, since those scores, with adequate diagnostic concordance indices (kappa > .40), would show specificities over 90% and positive predictive values of 61% (see Table 6).

## CONCLUSIONS

In an article published in 2010 on the use of tests in Spain, Muñiz and Fernández-Hermida underlined the wide use of the BDI even though it was an instrument that had not yet been marketed in Spain, and recommended that "It would be highly advisable for this test, which is so widely used by professionals, to be subject to a more systematic and rigorous validation process in our country" (pp. 116-117). Fortunately, one year later, on the 50th anniversary of its first publication, the recommendation of Muñiz and Fernández-Hermida was followed, and Spanish professionals now have a Spanish adaptation of the latest version of the BDI, the BDI-II, an adaptation which is the fruit of a systematic and rigorous validation process, and is now being marketed (Beck et al., 2011). The BDI-II shows substantial modifications over the BDI-I and BDI-IA. It is designed to assess DSM-IV diagnostic symptoms for depressive disorders, and in fact, among the self-rating instruments most widely used in Spain for assessing depression in adults, the BDI-II is the only one that covers all of the DSM-IV symptoms of major depressive episode and dysthymia. Furthermore, the BDI-II shows higher internal consistency and factor validity indices than the BDI-I or BDI-IA. The Spanish adaptation of the BDI-II also shows good reliability and validity indices in clinical and nonclinical samples, indices that enable useful guidelines and cut-off points to be suggested for assessing the severity of depression, evaluating the clinical significance of therapeutic changes, screening persons with depression and assisting in the differential diagnosis of depressive disorders.

## REFERENCES

American Psychiatric Association (1995). *DSM-IV. Manual diagnóstico y estadístico de los trastornos mentales [DSM-IV. Diagnostic and Statistical Manual of Mental Disorders]*. Barcelona: Masson. (Orig. 1994).

Beck, A. T., Rush, A. J., Shaw, B. F., y Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford. (Spanish translation in Bilbao: DDB, 1983).

Beck, A. T., and 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 [Manual. BDI-II. Beck Depression Inventory-II]* (Spanish adaptation: Sanz, J., and 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., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Botella, J., and 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., and 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, 2nd ed.* Hillsdale, NJ: LEA.

Conde, V., and Useros, E. (1974). El inventario para la medida de la depresión de Beck [The Beck Depression Inventory]. *Revista de Psiquiatría y Psicología Médica de Europa y América Latina, 12*, 153-167.

Cuijpers, P., Li, J., Hofmann, S. G., and 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., and 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., and 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., and Jacobson, N. S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy, 33*, 529-550.

Muñiz, J., and Fernández-Hermida, J. R. (2000). La utilización de los test en España [Use of tests in Spain]. *Papeles del Psicólogo, 76*, 41-49.

Muñiz, J., and Fernández-Hermida, J. R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests [The opinion of Spanish psychologists on the use of tests]. *Papeles del Psicólogo, 31*, 108-121.

Nezu, A. M., Nezu, C. M., Friedman, J., and Lee, M. (2009). Assessment of depression. In I. H. Gotlib and C. L. Hammer (Eds.), *Handbook of depression* (2ª ed.) (pp. 44-68). New York, NY: Guilford Press.

Prieto, G., and Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model for assessing the quality of tests used in Spain]. *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., and García-Vera, M. P. (2009). The Beck Depression Inventory-second edition (BDI-II): Factor congruence and generalizability of its indexes of internal consistency. In E. Rehulka (Ed.), *School and Health 21. General issues in Health Education* (pp. 331-342). Brno: MSD.

Sanz, J., and García-Vera, M. P. (2013). Rendimiento diagnóstico y estructura factorial del Inventario para la Depresión de Beck-Segunda Edición (BDI-II) [Diagnostic performance and factor structure of the Beck Depression Inventory-Second Edition (BDI-II)]. *Anales de Psicología, 29*, 66-75.

Sanz, J., García-Vera, M. P., Espinosa, R., Fortún, M., and Vázquez, C. (2005). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos [Spanish adaptation of the Beck Depression Inventory-II (BDI-II): Psychometric properties in patients with psychological disorders] *Clínica y Salud, 16*, 121-142.

Sanz, J., Gutiérrez, S., Gesteira, C., and García-Vera, M. P. (2014). Criterios y baremos para interpretar el "Inventario de Depresión de Beck-II" (BDI-II) [Criteria and scales for interpreting the Beck Depression Inventory-II (BDI-II)]. *Behavioral Psychology/Psicología Conductual, 22*, 37-59.

Sanz, J., Izquierdo, A., and García-Vera, M. P. (2013). 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 [A review of the questionnaires, scales and self-rating inventories most widely used in Spain for evaluating clinical depression in adults from the perspective of content validity]. *Psicopatología Clínica, Legal y Forense, 13.*

Sanz, J., Navarro, M. E., and Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 1. Propiedades psicométricas en estudiantes universitarios [Spanish adaptation of the Beck Depression Inventory-II (BDI-II): Psychometric properties in university students]. *Análisis y Modificación de Conducta, 29*, 239-288.

Sanz, J., Perdigón, L. A., and Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general [Spanish adaptation of the Beck Depression Inventory-II (BDI-II): Psychometric properties in the general population]. *Clínica y Salud, 14*, 249-280.

Sanz, J., and Vázquez, C. (2008). Trastornos del estado de ánimo: teorías psicológicas [Mood Disorders: Psychological theories]. In A. Belloch, B. Sandín and F. Ramos (Eds.), *Manual de Psicopatología*, vol. 2. [Psychopathology Manual, vol. 2] Revised edition (pp. 271-298). Madrid: McGraw-Hill.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285-1293.

Tennen, H., Hall, J. A., and 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., and Sanz, J. (1997). Fiabilidad y valores normativos de la versión española del Inventario para la Depresión de Beck de 1978 [Reliability and normative values in the Spanish version of the Beck Depression Inventory]. *Clínica y Salud, 8*, 403-422.

Yin, P., and 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.