TEST REVIEW:
Beck Depression Inventory – II
Catherine Smith & Bradley T. Erford
Loyola College in Maryland

1. Title: Beck Depression Inventory-II (BDI-II)

2. Authors: Aaron T. Beck, Gregory K. Brown, and Robert A. Steer

3. Publisher: The Psychological Corporation, San Antonio, TX

4. Forms, groups to which applicable: There have been two revisions of the Beck Depression Inventory. There exists the BDI, the BDI-IA, and the latest version, the BDI-II. Each inventory is an instrument for measuring the severity of depression in adolescents 13 years of age and up, as well as adults. The BDI-II contains DSM-IV criteria for depression not included in the two previous versions (Conoley, 1987).

5. Practical features: The BDI contains a four point scale for each item. The sum of the ratings on the 21 items is then simply compared to the cut score guidelines in order to identify the interpretive range (Beck, Brown, & Steer, 1996).

6. General type: The BDI-II serves as an indicator of the occurrence and severity of the symptoms of depression.

7. Date of publication: 1996.

8. Cost: booklets, answer sheets: The complete BDI-II kit consists of the manual and 25 record forms. The complete kit is priced at $53.00. The manual can be ordered for $25.50, record forms (quantity of 25) for $27.50, or @quantity of 100) for $104.50. Spanish record forms are also available for the same price.

9. Scoring services available and cost: There are no electronic scoring services available for the BDI-II. The BDI-II is scored by hand only.

10. Time required: Approximately 5-10 minutes is required for clients to complete the BDI-II.

11. Purpose for which evaluated: For use with adolescents and adults in assessing depression.

12. Description of test. Items and scoring: The BDI-II is a self-report analysis of depressive symptoms. It is not designed to be used for the actual diagnosis of depression (Sundberg, 1987). The wording of the BDI-II is clear and concise. The test contains 21 items, most of which assess depressive symptoms on a Likert scale of 0-3. The two exceptions to this are questions 16 and 18. Question 16 addresses changes in sleeping pattern, while question 18 addresses changes in appetite. The scale in these two items consist of 0, 1a, lb, 2a, 2b, 3a, & 3c. People are asked to report feelings consistent with
their own over the past 2 weeks instead of 1 week, as in the BDI and BDI-IA. The reason for this is to be consistent with the DSM-IV criteria for depression. There were also two items added to indicate any directional changes in eating and sleeping patterns. All forms of the inventory are written at the 5th grade reading level (Conoley, 1987). Clinical interpretation of scores is accomplished through criterion-referenced procedures utilizing the following interpretive ranges: 0-13 - minimal depression; 14-19 - mild depression; 20-28 - moderate depression; and 29-63 - severe depression (Beck et al., 1996).

13. Authors' purpose and basis for selecting item: The items in the BDI-II are designed to assess the severity of depression in adolescents and adults. This version of the test is specifically designed to address DSM-IV criteria for depression. The wording of all but three items has been changed from that of the BDI-IA for clarity. A series of factor and item analyses were performed in order to reduce the total number of items on the test from 27 to 21 (Beck et al., 1996).

14. Adequacy of directions, training required to administer: The instructions of the BDI-II are straightforward and clearly stated. Little training is required to administer or score the test. These two functions may be carried out by paraprofessionals. The interpretation of the final score requires a professional with clinical training and experience.

15. Mental functions or traits represented in each score: Only the total scale score, measuring clinical depression, is interpretable as subscale scores derived through factor analysis tend to be unreliable.

16. Comments regarding design of test: The simple 21-item rating scale format allows individuals to easily comprehend the questions and respond appropriately. The test is hand scored with little time and effort. Subtotals from pages one and two make up the Total Score.

17. Validation against criteria: BDI-II total scores have been correlated with scores on other psychological tests. The BDI-II is positively related to the Scale for Suicide Ideation (r = .37, n = 158) as well as the Beck Hopelessness Scale (r = .68, n = 158). The BDI-II was also positively correlated with the Hamilton Psychiatric Rating Scale for Depression (r = .71, n = 87) and the Hamilton Rating Scale for Anxiety (r = .47, n = 87; Beck et al., 1996). A diagnostic efficiency study using a clinical college sample of 127 students yielded a 93% true positive rate and 18% false positive rate (Beck et al., 1996).

18. Evidence of construct validity: The responses of 500 psychiatric outpatients were subjected to an unrotated principal components analysis and subsequent Promax-rotated iterated principal factor analysis yielding a two-factor solution: Somatic-Affective and Cognitive. Confirmation of this factor solution was attempted by analyzing responses of 120 typical college students. Using the same factor analytic procedures, two factors again emerged. However the two resultant factors represented the dimensions Cognitive- Affective and Somatic. It should be noted that while factor analysis of a 21 item scale using a sample of 500 participants meets currently accepted minimal
requirements of at least a 10:1 ratio of participants: items (actual ratio of 23.8:1), the study using 120 college students fall well short (actual ratio of 5.7:1). Such a low ratio of participants to item will likely result in an unreliable factor solution. Thus, future research should address factor stability, particularly through confirmatory factor analytic procedures. Also, as mentioned above, the BDI-II total score has been shown to correlate significantly with the scores of tests purporting to measure depression.

19. Fairness: As Beck, Brown, & Steer (1989) point out, differences among men and women may exist regarding frequency and severity of expression of depressive symptoms. However only one set of criterion-referenced interpretive guidelines was offered in the manual, and this set is not broken out by sex. Future studies must explore this potential sex difference as current interpretation guidelines could jeopardize the BDI-II’s diagnostic efficiency, potentially leading to an overidentification of women and underidentification of men. In addition, no evidence was reported regarding fairness of items and total score across racial/cultural categories.

20. Comments regarding validity for particular purposes: The BDI-II is a flexible instrument which can be used in clinical or non-clinical settings. The hit rates reported in the diagnostic efficiency study above demonstrate the clinical utility of the BDI-II.

21. Generalizability: The authors’ suggest the results of previous versions of the BDI were generalizable across gender and cultures (Beck, et al., 1989). However, the authors’ also recommend developing local norms when using the test with new populations.

22. Reliability: The BDI-II yields a coefficient alpha of .92 for the outpatient population (n = 500) in the sample referred to in the manual. The coefficient alpha for the college students (n = 120) in the sample was .93. Both surpass the coefficient alphas for the preceding two versions of the BDI. In addition, a one-week test-retest correlation of .93 resulted from a study of 26 outpatients who had been referred for depression and took the BDI-II during their first and second therapy sessions (Beck et al., 1996). In a study with both white and Mexican- American subjects, an internal consistency coefficient of .80 was computed for the BDI-IA. No significant differences were found between participants from the two cultural backgrounds, therefore supporting the test’s reliability across ethnic groups and aging populations (Ames, Gatewood-Colwell, & Kaczmarek, 1989).

23. Norms: Interpretation of BDI-II responses is criterion- referenced. The standardization sample was comprised of 317 women and 183 men. Urban based populations make up two subsamples and rural based populations make up another two subsamples. Two hundred and seventy-seven outpatients were from Cherry Hill, New Jersey, 50 outpatients were from Bala Cynwyd, Pennsylvania, 127 outpatients were from Philadelphia, Pennsylvania, and 46 were from Louisville, Kentucky. The average age of the outpatients in the sample was 37.20 years, however, the ages ranged from 13-86 years. Caucasians made up ninety-one percent of the sample, while African- Americans and Asian-Americans made up only four and one percent, respectively (Beck et al., 1996).
24. Comments regarding adequacy of norms: In the standardization sample, minority populations were extremely under-represented. Only 2 groups, African-American and Asian-American, were included at all. Together, they comprise only five percent of the total sample. Containing only 500 individuals, the standardization sample is very small. There is no information regarding socioeconomic status or residential location (urban, suburban, rural) compared to the US census data. Also, the BDI-II's interpretation is criterion-referenced with cut score guidelines to differentiate among minimal, mild, moderate and severe categories of depression. However, a brief scan of the reported means and standard deviation raise some concern about the variation of a client's scores on the BDI-II and clinical severity estimates. For example, the manual reports that clients clinically diagnosed as severely depressed obtained a mean of 32.96 (SD = 12.0) on the BDI-II. The manual states the cutoff for the severely depressed range is 29-63. While some clinicians may find the offered cutoff guidelines helpful, caution is warranted. These inferences really only lead the clinician to conclude that higher scores on the BDI-II serve to indicate that a significant level of depressive symptoms is being reported by the client. Further study with samples diverse in sex and race are needed to enhance confidence in these recommended severity categories.

25. Aids to user: The BDI-II manual is concise and user-friendly. It clearly delineates the development of the inventory. Administration and scoring are discussed in sufficient detail. Under the administration procedure section of the manual, the choice of self-administration or oral administration is outlined. A bibliography of 36 research-based sources is included. Item option characteristic curves were presented in the manual as an interpretive aid for the sophisticated user interested in maximizing sensitivity or specificity.

26. Comments of reviewers: The BDI-II is a relatively new test, therefore little is available in the way of reviews. Beck's previous inventories, including the BDI and the BDI-IA, have been accepted as well-developed and useful tools. As Conoley (1987) reports, "The BDI (revised) is a well-researched assessment tool with substantial support for its reliability and validity. When used clinically, care should be taken to use it as an indicator of the extent of depression not as a diagnostic tool. Additionally, if used as a suicide screening tool its high fakability should be remembered (p. 79). Sundberg (1987) goes on to say, 'it (BDI-IA) is a simple, short, and specific measure for depression. For clinical purposes, of course, diagnosis must involve much more than this test alone"(p. 80).

27. General evaluation of the test: Overall, the BDI-II is a useful instrument. It provides a fast, efficient way to assess depression in either a clinical or non-clinical environment. One concern is that the standardization sample is not demographically representative of the U.S. population and little evidence has been provided regarding the sex and culture fairness of the items and total score. Predominantly white females from the east coast are used in the sample. Also, the standardization sample is somewhat small, containing only 500 individuals and the socioeconomic status of the participants is not reported. The fakability of the inventory has been an issue with all three versions of the Beck Depression Inventory. This should always be kept in mind during the administration and interpretation of the test. Additionally, caution is warranted when using the cutoff
guidelines presented for criterion-referenced interpretation. Psychometrically, studies of the BDI-II indicate excellent internal consistency and one-week test-retest reliability on clinical samples, as well as substantial diagnostic efficiency and correlations with other tests purporting to measure the construct of depression. However, further exploratory and confirmatory factor analytic work must be undertaken to further understand the dimensionality underlying the BDI-II.

References


