Review of the Beck Depression Inventory-II by RICHARD F. FARMER, Associate Professor of Psychology, Idaho State University, Pocatello, ID:

The Beck Depression Inventory-II (BDI-II) is the most recent version of a widely used self-report measure of depression severity. Designed for persons 13 years of age and older, the BDI-II represents a significant revision of the original instrument published almost 40 years ago (BDI-I; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) as well as the subsequent amended version copyrighted in 1978 (BDI-IA; Beck, Rush, Shaw, & Emery, 1979; Beck & Steer, 1987, 1993). Previous editions of the BDI have considerable support for their effectiveness as measures of depression (for reviews, see Beck & Beamesderfer, 1974; Beck, Steer & Garbin, 1988; and Steer, Beck, & Garrison, 1986).

Items found in these earlier versions, many of which were retained in modified form for the BDI-II, were clinically derived and neutral with respect to a particular theory of depression. Like previous versions, the BDI-II contains 21 items, each of which assesses a different symptom or attitude by asking the examinee to consider a group of graded statements that are weighted from 0 to 3 based on intuitively derived levels of severity. If the examinee feels that more than one statement within a group applies, he or she is instructed to circle the highest weighting among the applicable statements. A total score is derived by summing weights corresponding to the statements endorsed over the 21 items. The test authors provide empirically informed cut scores (derived from receiver operating characteristic [ROC] curve methodology) for indexing the severity of depression based on responses from outpatients with a diagnosed episode of major depression (cutoff scores to index the severity of dysphoria for college samples are suggested by Dozois, Dobson, & Ahnberg, 1998).

The BDI-II can usually be completed within 5 to 10 minutes. In addition to providing guidelines for the oral administration of the test, the manual cautions the user against using the BDI-II as a diagnostic instrument and appropriately recommends that interpretations of test scores should only be undertaken by qualified professionals. Although the manual does not report the reading level associated with the test items, previous research on the BDI-IA suggested that items were written at about the sixth-grade level (Berndt, Schwartz, & Kaiser, 1983).

A number of changes appear in the BDI-II, perhaps the most significant of which is the modification of test directions and item content to be more consistent with the major depressive episode concept as defined in the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV; American Psychiatric Association, 1994). Whereas the BDI-I and BDI-IA assessed symptoms experienced at the present time and during the past week, respectively, the BDI-II instructs the examinee to respond in terms of how he or she has 'been feeling during the past two weeks, including today' (manual, p. 8, emphasis in original) so as to be consistent with the DSM-IV time period for the assessment of major depression. Similarly, new items included in the BDI-II address psychomotor agitation, concentration difficulties, sense of worthlessness, and loss of energy so as to make the BDI-II item set more consistent with DSM-IV criteria. Items that appeared in the BDI-I and BDI-IA that were dropped in the second edition were those that assessed weight loss, body image change, somatic preoccupation, and work difficulty. All but three of the items from the BDI-IA retained for inclusion in the BDI-II were reworded in some way. Items that assess changes in sleep patterns and appetite now address both increases and decreases in these areas.

Two samples were retained to evaluate the psychometric characteristics of the BDI-II: (a) a clinical sample (n = 500; 63% female; 91% White) who sought outpatient therapy at one of four outpatient clinics on the U.S. east coast (two of which were located in urban areas, two in suburban areas), and (b) a convenience sample of Canadian college students (n = 120; 56% women; described as 'predominantly White'). The average ages of the clinical and student samples were, respectively, 37.2 (SD = 15.91; range = 13-86) and 19.58 (SD = 1.84).

Reliability of the BDI was evaluated with multiple methods. Internal consistency was assessed using corrected item-total correlations (ranges: .39 to .70 for outpatients; .27 to .74 for students) and coefficient alpha (.92 for outpatients; .93 for students). Test-retest reliability was assessed over a 1-week interval among a small subsample of 26 outpatients from one clinic site (r = .93). There was no significant change in scores noted among this outpatient sample between the two testing occasions, a finding that is different from those often obtained with college students who, when tested repeatedly with earlier versions of the BDI, were often observed to have lower scores on subsequent testing occasions (e.g., Hatzenbuehler, Parpal, & Matthews, 1983).

Following the method of Santor, Ramsay, and Zuroff (1994), the test authors also examined the item-option characteristic curves for each of the 21 BDI-II items as endorsed by the 500 outpatients. As noted in a previous review of the BDI (1993 Revised) by Waller (1998), the use of this method to evaluate item performance represents a new standard in test revision. Consistent with findings for depressed outpatients obtained by Santor et al. (1994) on the BDI-IA, most of the BDI-II items performed well as evidenced by the individual item-option curves. All items were reported to display monotonic relationships with the underlying dimension of depression severity. A minority of items were somewhat problematic, however, when the degree of correspondence between estimated and a priori weights associated with item response options was evaluated. For example, on Item 11 (agitation), the response option weighted a value of 1 was more likely to be endorsed than the option weighted 3 across all levels of depression, including depression in the moderate and severe ranges. In general, though, response option weights of the BDI-II items did a good job of discriminating across estimated levels of depression severity. Unfortunately, the manual does not provide detailed discussion of item-option characteristic curves and their interpretation.

The validity of the BDI-II was evaluated with outpatient subsamples of various sizes. When administered on the same occasion, the correlation between the BDI-II and BDI-IA was quite high (n = 101, r = .93), suggesting that these measures yield similar patterns of scores, even though the BDI-II, on average, produced equated scores that were about 3 points higher. In support of its convergent validity, the BDI-II displayed moderately high correlations with the Beck Hopelessness Scale (n = 158, r = .68) and the Revised Hamilton Psychiatric Rating Scale for Depression (HRSD-R; n = 87, r = .71). The correlation between the BDI-II and the Revised Hamilton Anxiety Rating Scale (n = 87, r = .47) was significantly less than that for the BDI-II and HRSD-R, which was cited as evidence of the BDI-II's discriminant validity. The BDI-II, however, did share a moderately high correlation with the Beck Anxiety Inventory (n = 297; r = .60), a finding consistent with past research on the strong association between self-reported anxiety and depression (e.g., Kendall & Watson, 1989). Additional research published since the manual's release (Steer, Ball, Ranieri, & Beck, 1997) also indicates that the BDI-II shares higher correlations with the SCL-90-R Depression subscale (r = .89) than with the SCL-90-R Anxiety subscale (r = .71), although the latter correlation is still substantial. Other data presented in the test manual indicated that of the 500 outpatients, those diagnosed with mood disorders (n = 264) had higher BDI-II scores than those diagnosed with anxiety (n = 88), adjustment (n = 80), or other (n = 68) disorders. The test authors also cite evidence of validity by separate factor analyses performed on the BDI-II item set for outpatients and students. However, findings from these analyses, which were different in some significant respects, are questionable evidence of the measure's validity as the test was apparently not developed to assess specific dimensions of depression. Factor analytic studies of the BDI have historically produced inconsistent findings (Beck et al., 1988), and preliminary research on the BDI-II suggests some variations in factor structure within both clinical and student samples (Dozois et al., 1998; Steer & Clark, 1997; Steer, Kumar, Ranieri, & Beck, 1998). Furthermore, one of the authors of the BDI-II (Steer & Clark, 1997) has recently advised that the measure not be scored as separate subscales.

SUMMARY. The BDI-II is presented as a user-friendly self-report measure of depression severity. Strengths of the BDI-II include the very strong empirical foundation on which it was built, namely almost 40 years of research that demonstrates the effectiveness of earlier versions. In the development of the BDI-II, innovative methods were employed to determine optimum cut scores (ROC curves) and evaluate item performance and weighting (item-option curves). The present edition demonstrates very good reliability and impressive test item characteristics. Preliminary evidence of the BDI-II's validity in clinical samples is also encouraging. Despite the many impressive features of this measure, one may wonder why the test developers were not even more thorough in their presentation of the development of the BDI-II and more rigorous in the evaluation of its effectiveness. The test manual is too concise, and often omits important details involving the test development process. The clinical sample used to generate cut scores and evaluate the psychometric properties of the measure seems unrepresentative in many respects (e.g., racial make-up, patient setting, geographic distribution), and other aspects of this sample (e.g., education level, family income) go unmentioned. The student sample is relatively small and, unfortunately, drawn from a single university. Opportunities to address important questions regarding the measure were also missed, such as whether the BDI-II effectively assesses or screens the DSM-IV concept of major depression, and the extent to which it may accomplish this better than earlier versions. This seems to be a particularly important question given that the BDI was originally developed as a measure of the depressive syndrome, not as a screening measure for a nosologic category (Kendall, Hollon, Beck, Hammen, & Ingram, 1987), a distinction that appears to have become somewhat blurred in this most recent edition. Also, not reported in the manual are analyses to examine possible sex biases among the BDI-II item set. Santor et al. (1994) reported that the BDI-IA items were relatively free of sex bias, and given the omission of the most sex-biased item in the BDI-IA (body image change) from the BDI-II, it is possible that this most recent edition may contain even less bias. Similarly absent in the manual is any report on the item-option characteristic curves for nonclinical samples. Santor et al. (1994) reported that for most of the BDI-IA items, response option weights were less discriminating across the range of depression severity among their college sample relative to their clinical sample, an anticipated finding given that students would be less likely to endorse response options hypothesized to be consistent with more severe forms of depression. Also, given that previous editions of the BDI have shown inconsistent associations with social undesirability (e.g., Tanaka-Matsumi & Kameoka, 1986), an opportunity was missed to evaluate the extent to which the BDI-II measures something different than this response set. Despite these relative weaknesses in the development and presentation of the BDI-II, existent evidence suggests that the BDI-II is just as sound if not more so than its earlier versions.

REVIEWER'S REFERENCES

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: The Depression Inventory. In P. Pichot & R. Oliver-Martin (Eds.), Psychological measurements in psychopharmacology: Modern problems in pharmacopsychiatry (vol. 7, pp. 151-169). Basel: Karger.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). Cognitive therapy of depression. New York: Guilford.

Berndt, D. J., Schwartz, S., & Kaiser, C. F. (1983). Readability of self-report depression inventories. Journal of Consulting and Clinical Psychology, 51, 627-628.

Hatzenbuehler, L. C., Parpal, M., & Matthews, L. (1983). Classifying college students as depressed or nondepressed using the Beck Depression Inventory: An empirical analysis. Journal of Consulting and Clinical Psychology, 51, 360-366.

Steer, R. A., Beck, A. T., & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T. A. Ban (Eds.), Assessment of depression (pp. 123-142). New York: Springer-Verlag.

Tanaka-Matsumi, J., & Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. Journal of Consulting and Clinical Psychology, 54, 328-333.

Beck, A. T., & Steer, R. A. (1987). Beck Depression Inventory manual. San Antonio, TX: The Psychological Corporation.

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding the use of the Beck Depression Inventory. Cognitive Therapy and Research, 11, 289-299.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical Psychology Review, 8, 77-100.

Kendall, P. C., & Watson, D. (Eds.). (1989). Anxiety and depression: Distinctive and overlapping features. San Diego, CA: Academic Press.

Beck, A. T., & Steer, R. A. (1993). Beck Depression Inventory manual. San Antonio, TX: Psychological Corporation.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. (1994). Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. Psychological Assessment, 6, 255-270.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1997). Further evidence for the construct validity of the Beck Depression Inventory-II with psychiatric outpatients. Psychological Reports, 80, 443-446.

Steer, R. A., & Clark, D. A. (1997). Psychometric characteristics of the Beck Depression Inventory-II with college students. Measurement and Evaluation in Counseling and Development, 30, 128-136.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. Psychological Assessment, 10, 83-89.

Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. Journal of Psychopathology and Behavioral Assessment, 20, 127-137.

Waller, N. G. (1998). [Review of the Beck Depression Inventory-1993 Revised]. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook (pp. 120-121). Lincoln, NE: The Buros Institute of Mental Measurements.