

# A PSYCHOMETRIC COMPARISON OF THE BECK DEPRESSION INVENTORY AND THE INVENTORY FOR DIAGNOSING DEPRESSION IN A COLLEGE POPULATION

Terry M. Pace  
Maria del Carmen Trapp  
University of Oklahoma

The relationship between the Beck Depression Inventory (BDI) and the Inventory for Diagnosing Depression (IDD) was evaluated in a college population. The BDI is an established self-report depression instrument. The IDD is a relatively new self-report depression instrument. The IDD was designed to address the BDI's lack of full correspondence with *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.) (*DSM-III*) through *DSM-IV* criteria. The two instruments were found to be highly correlated and Cronbach's alpha was found to be high for each instrument. The diagnostic performance of three BDI cutoff scores was found to vary considerably when compared to IDD diagnostic criteria. Implications for selection and use of self-report depression inventories are discussed.

Myers et al. (1984) found the point-prevalence rate of major unipolar depressive disorder in the general adult community in the United States to be about 3% as measured by the Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliff, 1981). Specific reports about the prevalence of depression in college students have yielded inconsistent conclusions; however, most studies have documented a prevalence rate several times higher than found in the general adult population.

Hammen (1980) studied 400 college freshmen, and using a BDI (Beck & Steer, 1987) cutoff of 16 to indicate moderate depression, found a prevalence rate of 9.25% ( $n = 37$ ). Hammen conducted

clinical interviews with 21 of the 37 students who had BDI scores of 16 or above. Of these 21, 13 met criteria for diagnosis of major or minor depression based on the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robbins, 1978). It is not possible to obtain a precise RDC-prevalence rate for this entire sample, but a reasonable estimate might be made as follows: 13 (62%) of the 21 students with elevated BDI scores who were interviewed met RDC criteria; 62% of the whole group of 37 students in the sample with elevated BDI scores equals 23; and 23 is 5.75% of 400. Thus, an estimated RDC-prevalence rate might be 5.75% for this sample of 400 college freshmen.

Rimmer, Halikas, and Schuckit (1982) prospectively studied a cohort of 158 college students, conducting RDC diagnostic interviews during each of the 4 years of college. This study found that depression rates varied from 15% during the

---

Address correspondence to Terry M. Pace, PhD, University of Oklahoma, Counseling Psychology Program, Department of Educational Psychology, 820 Van Vleet Oval, Norman, OK 73019. Phone: 405-325-1480. Fax: 405-325-6655.

freshman year, 13% during the sophomore year, and 14% during the junior year to 17% during the senior year. These results suggest an average rate of depression to be about 15% for this college sample. The difference between the 15% rate found by Rimmer et al. and the estimated rate of 5.75% found by Hammen (1980) is difficult to account for, but could be due to the demographics of the two samples or the training and behavior of the interviewers in the two studies.

A number of researchers have questioned the use of self-report instruments as a basis for classifying college students as depressed (Gotlib, 1984; Hammen, 1980). The arguments made by these authors are based on the typically high rates of depression found via self-report and include concerns regarding instability of self-reported depression among college students (Hammen). However, these conclusions have been questioned. Specifically, other research indicates that depression rates are actually quite high among college students (Rimmer et al., 1982) and that perceived instability of depression among college students may be a function of sensitization when re-administered the same depression inventory (Hatzenbuehler, Parpal, & Matthews, 1983). In any case, it would appear that the evidence is inconclusive about the appropriate use of self-report depression instruments with college students. Self-report may remain a valid and reliable measure of depression in college students and deserves continued investigation.

Due to the time and skill needed to implement diagnostic clinical interviews, the need for quantifiable assessment data, objective indexes of treatment progress, and useful screening procedures, standardized self-report inventories have gained popularity in usage as an efficient aid for both clinicians and researchers. Self-reports are not a substitute for clinical interviews or diagnostic criteria, but can expedite the diagnostic process and provide valuable data about symptom severity and symptom profiles that may be useful in classifying depression and measuring response to intervention.

The “gold standard” for diagnosis of depression is the use of a clinical interview and the application of standardized criteria as specified in *DSM-IV* (American Psychiatric Association [APA], 1994). Most self-report depression inventories do not include full coverage of *DSM-IV* diagnostic criteria for major depressive disorder. For example, the BDI does not contain questions regarding increased appetite, weight gain, hyposomnia, psychomotor agitation or retardation, nonsuicidal death wishes, and problems with concentration. On the other hand, the IDD (Zimmerman, Coryell, Corenthal, & Wilson, 1986) aligns with the full diagnostic criteria of *DSM-IV*.

Using a college-student population, the current study focused on the correlation of the BDI and the IDD to evaluate concurrent validity of the instruments, and Cronbach’s alpha was calculated to measure the internal consistency for both instruments. The diagnostic performance of three commonly used BDI cutoff scores was also evaluated using IDD diagnosis as the criteria.

## Method

### Subjects

Subjects were 220 White college students from a large state university in the Midwest. All students were recruited from lower- and upper-level general education courses to participate in a study on depression. Research bonus points were given by course instructors for participation. The sample was 74.3% female and 25.7% male with an average age of 22.17 years ( $SD = 4.68$ ). Although the exact year of college was not obtained, most of the students were known to be in their sophomore or junior years, with freshmen and seniors fairly well represented in the sample.

### Instruments

#### *BDI*

The BDI is the most widely used self-report instrument for depression screening (Beck, Steer, & Garbin, 1988). The original BDI was developed by Beck, Ward, Mendelson, Mock, and Erbaugh (1961) and was revised by Beck, Rush, Shaw, and Emery (1979). The BDI has 21 items that are rated

on a 4-point scale (0–3), reflecting increasing symptom severity. The BDI is scored simply by totaling the highest responses for all the items. Total scores range from 0 to 63. Guidelines for interpreting scores are generally as follows: 0 to 9—no depression; 10 to 19—mild depression; 20 to 29—moderate depression; and 30 or higher—severe depression (Kendall, Hollon, Beck, Hammen, & Ingram, 1987). In a major review, Beck et al. (1988) concluded the BDI has been shown to have acceptable reliability and validity. Typical reliability results show 1-week test-retest to be in the .70s, and validity results indicate the BDI correlates in the .60s, with clinical diagnosis.

### **IDD**

The IDD is a 22-item self-report scale designed to diagnose major depressive disorder according to *DSM-III* (APA, 1980) criteria. However, the IDD also covers all the symptoms of major depressive disorder specified by *DSM-III-R* (APA, 1987) and *DSM-IV*. In addition, like the BDI, the IDD yields a continuous measure of the severity of depression, with scores marked on a 5-point scale (0–4), reflecting increasing symptom severity. For quantitative purposes, the IDD is scored easily by totaling the highest response for all items. Total scores range from 0 to 88. Empirical cutoff scores reflecting levels of severity have not yet been published for the IDD. Zimmerman and Coryell (1987) provide the diagnostic scoring algorithm for the IDD.

Although the IDD is a relatively new instrument, the initial psychometric research is positive. Zimmerman et al. (1986) found consecutive day test-retest reliability to be .98 for a sample of eight nondepressed and eight depressed psychiatric inpatients. For the same sample, split-half reliability was .93 and Cronbach's alpha was .92. In a study of 398 first-degree relatives of psychiatric inpatients and normal controls, Zimmerman and Coryell (1987) reported split-half reliability of .91 and an alpha of .92.

Validity research on the IDD has also been positive. Zimmerman et al. (1986) reported a positive correlation of .87 between the IDD and BDI. They also reported 78.1% to 81.7% agreement between

an IDD diagnosis of depression and clinical diagnosis of depression by experienced psychiatrists. Additionally, Zimmerman and Coryell (1987) found very similar point-prevalence rates in a general, adult community sample of 3.5% and 2.8%, respectively, for the IDD and the DIS. The overall diagnostic agreement between the IDD and DIS was found to be 97.2%.

### **Procedure**

All subjects were scheduled for an individual session and completed a consent form, the BDI, the IDD, and a demographics questionnaire, with all instruments presented in counterbalanced order.

## **Results**

### **Descriptive Statistics**

The BDI had a mean of 10.99 ( $SD = 9.21$ ) with a range of 0 to 44. The IDD had a mean of 15.04 ( $SD = 11.28$ ) with a range of 0 to 52.

### **Concurrent Validity**

A Pearson correlation was calculated for the total mean score of the BDI and IDD, yielding a highly significant correlation,  $r = .90$ ,  $p < .000$ .

### **Internal Consistency Reliability**

Cronbach's alpha was calculated for both instruments, yielding an alpha of .92 for the BDI and an alpha of .91 for the IDD.

### **Diagnostic Performance**

The diagnostic performance of the BDI cutoff scores of 10, 20, and 30 were examined using IDD diagnosis as the criterion. All calculations were conducted according to the guidelines provided by Kessel and Zimmerman (1993). Calculations for each cutoff score included the following calculations: Point-Prevalence Rate (current rate of depression as identified by each BDI cutoff and IDD diagnosis); Hit Rate (observed percentage of agreement in classifying depressed and nondepressed cases for each BDI cutoff and IDD diagnosis); Kappa (chance corrected percentage of agreement for each BDI cutoff score with the IDD); Sensitivity (the true-positive rate or the percentage of depressed cases according to the IDD accurately

identified as depressed by each BDI cutoff); Specificity (the true-negative rate or the percentage of nondepressed cases according to the IDD accurately classified as nondepressed by each BDI cutoff); Positive-Predictive Power (the percentage of cases classified as depressed by each BDI cutoff that are also identified as depressed by the IDD); Negative-Predictive Power (the percentage of cases classified as nondepressed by each BDI cutoff that are also identified as nondepressed by the IDD); False-Positive Rate (percentage of nondepressed cases according to the IDD that are incorrectly classified as depressed by each BDI cutoff); and False-Negative Rate (percentage of depressed cases according to the IDD that are incorrectly classified as nondepressed by each BDI cutoff). Raw diagnostic results for each cutoff score are presented

in Table 1. Results of all diagnostic performance calculations are presented in Table 2.

## Discussion

The IDD is a relatively new self-report inventory for diagnosing depression, whereas the BDI is a well established self-report instrument for assessing the severity of depressive symptoms. Using a college sample, our results found that the two instruments are highly correlated ( $r = .90$ ), which may be interpreted as supporting the concurrent validity of both. Despite differences in the symptom coverage of the two instruments, the differences do not appear to globally differentiate the results of the instruments. At least globally, the instruments appear to be measuring the same

Table 1  
*Raw Diagnostic Performance for Three BDI Cutoff Scores Using IDD Diagnosis as the Criteria*

BDI	BDI cutoff scores					
	IDD diagnosis present			IDD diagnosis absent		
	10	20	30	10	20	30
Positive	47	30	10	59	10	0
Negative	1	18	38	113	162	172

*Note.* BDI = Beck Depression Inventory; IDD = Inventory for Diagnosing Depression; Positive = meets or exceeds cutoff; Negative = below cutoff.

Table 2  
*Diagnostic Performance of Three BDI Cutoff Scores Using IDD Diagnosis as the Criteria*

Diagnostic performance	BDI cutoff scores		
	10	20	30
Point-Prevalence Rate	.21	.14	.05
Hit Rate	.73	.87	.83
Kappa	.45	.59	.29
Sensitivity	.98	.63	.21
Specificity	.66	.94	1.00
Positive-Predictive Power	.44	.75	1.00
Negative-Predictive Power	.99	.90	.82
False-Positive Rate	.34	.06	.00
False-Negative Rate	.02	.38	.79

*Note.* BDI = Beck Depression Inventory; IDD = Inventory for Diagnosing Depression.

construct. Cronbach's alpha was found to be very high for both instruments.

The point-prevalence rate of depression has been established at about 3% in the general, adult community population (Myers et al., 1984). Past research with the IDD found a very similar rate of depression in a general adult community sample (Zimmerman & Coryell, 1987). In the current study of college students, the rate of IDD diagnosed depression was 22%. This large increase over the 3.5% rate reported by Zimmerman and Coryell was unexpected. The sample in this study consisted of college students and was distributed across all 4 years of college, with an average age of 22 years. Also, the sample was predominately female (74%). Perhaps, the demographic features of this college sample, being mostly female, may have inflated the rate of depression. However, even so, the IDD may indicate increased rates of depression for college students over what would be expected, based on general adult population rates. While still somewhat high, these results are not unbelievable, based on the 15% depression rate among college students reported by Rimmer et al. (1982). However, it may also be that the IDD overdiagnoses depression in college students and, in spite of following complete *DSM-IV* criteria, the use of the term diagnosis may be misleading for this self-report instrument.

Perhaps most importantly, using IDD diagnosis of major depressive disorder as the criterion, this study empirically evaluated the diagnostic performance of three widely used BDI cutoff scores. Compared to the 22% point-prevalence rate found by the IDD, the BDI yielded point-prevalence rates of 21% for a cutoff of 10, 14% for a cutoff of 20, and 5% for a cutoff of 30. These results demonstrate how the use of different BDI cutoff scores impact classification rates. The most liberal cutoff of 10 yielded about the same results as diagnosis by the IDD. A cutoff of 20 produced a more moderate 14% rate—that is very close to the 15% rate reported by Rimmer et al. (1982). Finally, a cutoff of 30 yielded a rate of 5%—that is much closer to the 3% rate found in the general adult population.

Although a BDI cutoff of 10 produced the closest point-prevalence rate to the IDD, diagnostic agreement, as measured both by absolute hit rate and kappa, was best for a cutoff of 20. The cutoff of 20 tended to balance out diagnostic performance, primarily by yielding strong sensitivity and moderate specificity (i.e., at 20 there were very few false-positives and a modest amount of false-negatives). For uses in a general, college-student population, diagnostic accuracy of the BDI is likely to be maximized by the use of a cutoff of 20. This conclusion is further supported by the close correspondence between the 14% rate for a cutoff of 20 on the BDI and the 15% rate previously established by using RDC criteria assessed by diagnostic interview (Rimmer et al., 1982).

Using either a low-end cutoff of 10 or a high-end cutoff of 30 produced more skewed results compared to the IDD. If the BDI is used solely as a screening instrument, then 10 might be a good cutoff score to select. At 10, the BDI has very high sensitivity (low false-negative rate), thus, very few potential cases will be missed, but the false-positive rate increases substantially, lowering specificity. At 30, the BDI has near-perfect specificity (near zero false-positive rate), but has very poor sensitivity (high false-negative rate), indicating that almost all clear-cut or severe cases will be included, but that many cases will be inappropriately excluded. If the goal is to only identify the most severe cases of depression, then using a cutoff of 30 will be most efficient.

On the one hand, our results indicate the IDD is a valid measure of depression, based on its high correlation with the BDI. On the other hand, the rate of depression in this college sample may be somewhat inflated by the IDD, and use of the BDI with a cutoff of 20 might be the best self-report assessment of depression in a college population. Despite improved symptom coverage, the IDD does not appear to improve upon classification of depression in college students, as compared to the BDI using a cutoff of 20. Thus, while the IDD appears to be a promising self-report depression instrument, more research is needed to determine its comparative value to other instruments. Both



the IDD and BDI need to be evaluated simultaneously in several representative samples (e.g., college, general community, outpatient, inpatient, medical) using a *DSM-IV*-based diagnostic interview as a gold standard for diagnostic performance.

Limitations of the current study include the lack of a gold standard diagnostic interview to establish the true point-prevalence rate in this college sample and to use for evaluating the performance of each instrument. However, given past performance of the IDD, it might be expected that a diagnostic interview would yield similar results. Further, the sample in this study consisted of all White, mostly female, college students recruited from undergraduate general education classes; thus, caution should be used in generalizing these results pending replication with other samples.

### References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.
- Beck, A. T., & Steer, R. A. (1987). *Manual: Beck Depression Inventory*. San Antonio, TX: The Psychological Corporation.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: 25 years of evaluation. *Clinical Psychology Review*, 8, 77-100.
- 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.
- Gotlib, I. H. (1984). Depression and general psychopathology in university students. *Journal of Abnormal Psychology*, 93, 19-30.
- Hammen, C. L. (1980). Depression in college students: Beyond the Beck Depression Inventory. *Journal of Consulting and Clinical Psychology*, 48, 126-128.
- Hatzenbuehler, L. C., Parpel, 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.
- Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research*, 11, 280-299.
- Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment*, 5, 395-399.
- Myers, J. K., Weissman, M. M., Tischler, G. H., Holzer, C. E., Leaf, P. J., Orvaschel, H., Anthony, J. C., Boyd, J. H., Burke, J. D., Kramer, M., & Stoltzman, Z. (1984). Six-month prevalence of disorders in three communities. *Archives of General Psychiatry*, 41, 959-967.
- Rimmer, J., Halikas, J. A., & Schuckit, M. A. (1982). Prevalence and incidence of psychiatric illness in college students: A four year prospective study. *College Health*, 30, 207-211.
- Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics and validity. *Archives of General Psychiatry*, 38, 381-389.
- Spitzer, R. L., Endicott, J., & Robbins, E. (1978). Research diagnostic criteria. *Archives of General Psychiatry*, 35, 773-782.
- Zimmerman, M., & Coryell, C. (1987). The Inventory to Diagnose Depression (IDD): A self-report scale to diagnose major depressive disorder. *Journal of Consulting and Clinical Psychology*, 55, 55-59.
- Zimmerman, M., Coryell, C., Corenthal, C., & Wilson, S. W. (1986). A self-report scale to diagnose major depressive disorder. *Archives of general Psychiatry*, 43, 1076-1081.