JOURNAL OF PERSONALITY ASSESSMENT, *78*(3), 451–460

Convergent Validity of the Beck Depression Inventory–II With the Reynolds Adolescent Depression Scale in Psychiatric Inpatients

David G. Krefetz, Robert A. Steer, and Nazli A. Gulab

Department of Psychiatry
School of Osteopathic Medicine
University of Medicine and Dentistry of New Jersey

Aaron T. Beck

Beck Institute for Cognitive Therapy and Research
Bala Cynwyd, Pennsylvania

The Beck Depression Inventory–II (BDI–II; Beck, Steer, & Brown, 1996) and the Reynolds Adolescent Depression Scale (RADS; Reynolds, 1987) were administered to 56 female and 44 male psychiatric inpatients whose ages ranged from 12 to 17 years old. The Cronbach coefficient αs for the BDI–II and RADS were, respectively, .92 and .91 and indicated comparably high levels of internal consistency. The correlation between the BDI–II and RADS total scores was .84, $p < .001$. Binormal receiver-operating-characteristic analyses indicated that both instruments were comparably effective in differentiating inpatients who were and were not diagnosed with a major depressive disorder; the areas under the ROC curves for the BDI–II and RADS were, respectively, .78 and .76. The results (a) indicate that the BDI–II and the RADS have similar psychometric characteristics and (b) support the convergent validity of the BDI–II for assessing self-reported depression in adolescent inpatients.

In describing some of the problems incurred in screening for psychopathology in children and adolescents, the Surgeon General of the United States (U. S. Department of Health and Human Services [DHHS], 1999) stressed that the boundaries "between one diagnosis and another are fluid [and that clinicians] may be greatly

aided by various diagnostic aids such as brief questionnaires that can be completed in a waiting room" (p. 138). One of the diagnostic aids listed by the Surgeon General (DHHS, 1999, p. 139) was the amended Beck Depression Inventory (BDI–IA; Beck & Steer, 1993).

Although the clinical utility and psychometric properties of the BDI–IA have been extensively investigated across a variety of clinical and nonclinical samples of adolescents (for a review of the BDI–IA's use with adolescents, see Steer & Beck, 1988), it should be noted that the BDI–IA was upgraded to the BDI–II (Beck, Steer, & Brown, 1996). The BDI–II now addresses all nine of the symptom criteria listed in the American Psychiatric Association's (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*) for a major depressive episode. However, some investigators might question whether a self-report instrument such as the BDI–II, which was originally developed for assessing the severity of depression in adults who were diagnosed with various psychiatric disorders, can adequately address depressive symptoms in adolescents (for a detailed discussion of the issues involved in assessing depression in children and adolescents with self-report instruments, see Reynolds, 1994). For example, the somatic symptoms of depression and some affective symptoms, such as irritability, have been found to decrease as adolescents grow older (Weller & Weller, 2000).

The psychometric properties of the BDI–II, especially its convergent validity with respect to other measures of depression, have been minimally studied in adolescents. Steer, Kumar, Ranieri, and Beck (1998) administered the BDI–II to 105 male and 105 female psychiatric outpatients between 12 and 18 years old and reported that its internal consistency was high (Cronbach coefficient $\alpha = .92$). They also performed an iterated principal factor analysis with the 21 BDI–II symptom ratings and found the cognitive and noncognitive dimensions of self-reported depression that have been previously identified in samples of adult psychiatric outpatients (for a review of BDI–II research, see Steer & Beck, 2000). However, we have been unable to identify any research that has addressed whether the psychometric properties of the BDI–II are comparable to those afforded by a self-report instrument that was specifically developed to measure the severity of depression in adolescents, such as the Reynolds Adolescent Depression Scale (RADS; Reynolds, 1987).

The purpose of this study was to compare the psychometric characteristics of the BDI–II with those of the RADS for adolescent (12 to 17 years old) psychiatric inpatients. We were also interested in ascertaining whether both instruments were comparable in discriminating adolescents who were diagnosed with a *DSM–IV* major depressive disorder (MDD) from those who were not diagnosed with a MDD. MDD was focused on because this is not only a prevalent mood disorder diagnosed in adolescents (Roberts, Lewinsohn, & Seeley, 1995) but also the mood disorder that the BDI–II was upgraded to address (Beck et al., 1996).

METHOD

Sample

The sample was composed of 100 adolescents (12 to 17 years old) who were consecutively admitted to the inpatient Child and Adolescent Psychiatric Unit of Kennedy Memorial Hospital/University Medical Center, a general hospital located in Cherry Hill, New Jersey. Cherry Hill is a suburban, middle class to upper middle class community situated near Camden, New Jersey and Philadelphia, Pennsylvania. There were 56 female and 44 male adolescents.[1] With respect to ethnicity 12 were African American, 68 were White, 11 were Hispanic, and 9 were Asian. The mean age was 14.65 (SD = 1.53) years.

All of the adolescents were independently diagnosed by one of two psychiatrists (David Krefetz and Nazli Gulab) who were (a) board certified in child and adolescent psychiatry and (b) blind to the BDI–II and the RADS results at the time that they had initially recorded their *DSM–IV* diagnoses into the patients' medical charts. To verify the accuracy of a MDD diagnosis, 50 of the adolescents were also randomly asked by psychiatric residents to answer nine yes–no questions from the Mood Module of the Primary Care Evaluation of Mental Disorders (PRIME–MD; Spitzer et al., 1995); the residents were also asked to record whether any of these 50 adolescents were taking antidepressant medications at the time of testing. The principal diagnoses were classified into two broad diagnostic groups based on the type of *DSM–IV* Axis I disorder for analysis purposes; 57 were diagnosed with a MDD (20 single and 37 recurrent episodes) and 43 were diagnosed without a MDD (5 bipolar, 8 dysthymia, 15 depression not otherwise specified, 5 adjustment disorders, and 10 impulse control/disruptive disorders).

Instruments

BDI–II. The 21-item BDI–II measures the severity of self-reported depression in adolescents and adults (Beck et al., 1996). It is scored by summing the highest ratings for each of the 21 symptoms and takes less than 10 min to complete. Each symptom is rated on a 4-point rating scale ranging from 0 to 3, and total scores can range from 0 to 63. The time frame for the BDI–II ratings is for the "Past Two Weeks, Including Today." According to Beck et al. (1996), BDI–II total scores ranging from 0 to 13 represent normal to minimal depression, total scores from 14 to 19 are mild, total scores from 20 to 28 are moderate, and total scores from 29 to 63 are severe.

A number of studies have investigated the psychometric characteristics of the BDI–II with respect to both clinical and nonclinical populations (for a comprehen-

[1]Because the sample is composed of 100 adolescents, percentages will not be reported.

sive review, see Steer & Beck, 2000). These studies have generally found that the BDI–II has high internal consistency (coefficient αs > .90) and moderate to strong convergent validities (*r*s > .50) with other self-report and clinical rating scales of depression in adult and adolescent psychiatric patients, college students, and normal adults.

RADS. The RADS is composed of 30 sentences and was specifically developed to measure the severity of self-reported depressive symptoms in adolescents (13 to 18 years old). However, Reynolds and Mazza (1998) found that the RADS may also be used with children as young as 11 years old. In describing the RADS, Reynolds (1987) stressed that it was not constructed "to provide a diagnosis of a specific and definite depressive disorder, [but to assess] a range of symptomatology associated with depression" (p. 1). Consequently, the RADS contains some statements that do not reflect *DSM–IV* criteria for mood disorders, such as "feeling loved" or "worrying about school."

In completing the RADS, adolescents are asked to use a 4-point rating scale ranging from 1 (*almost never*) to 4 (*most of the time*) to describe how they usually feel with respect to what each sentence states. In scoring the RADS, the ratings for seven positively worded sentences are reversed, and the 30 ratings are then summed to yield a total score that can range from 30 to 120. According to Reynolds (1987), a total cutoff score of 77 and above has been "determined to delineate a level of symptom endorsement associated with clinical depression" (p. 6).

The psychometric characteristics of the RADS have been extensively investigated with respect to both clinical and nonclinical populations (for a comprehensive review of its psychometric properties, see Reynolds, 1987). Briefly, the RADS has high internal consistency (coefficient αs > .90), strong test–retest reliability (*r*s > .60), and moderate to strong convergent validities (*r*s > .50) with other self-report and clinical rating scales of depression. Reynolds (1987, p. 21) listed the Pearson product–moment correlations between the RADS and BDI–IA that were reported in 10 studies, and these correlations ranged from .70 to .76, *p*s < .001.

Mood Module. The PRIME–MD Mood Module is a focused interviewing guide composed of "yes" and "no" questions that are asked by a clinician. The first nine items of the Mood Module are symptoms of a MDD, and a patient is asked whether he or she has had the symptom "nearly every day for the last 2 weeks." If five or more of the nine symptoms are present and one of these five symptoms is sadness/hopelessness or anhedonia, then a diagnosis of a MDD is supported. It is important to note that we used an adaptation of the Mood Module for adolescents that was employed by Winter, Steer, Jones-Hicks, and Beck (1999) in which the *DSM–IV* MDD symptom of irritability is added to Criterion 5 along with feeling down, depressed, and hopeless. Spitzer et al. (1994) reported that the interjudge

agreement between being diagnosed with a MDD by a primary care physician and a mental health clinician using a more detailed psychiatric interview yielded a Cohen's κ coefficient of .61 for 431 patients, a value that was interpreted as indicating satisfactory interjudge agreement for diagnosing a MDD.

Procedure

The order in which the adolescents were administered the BDI–II and the RADS was randomly alternated. As previously mentioned, 50 of the 100 adolescents were also asked by a psychiatric resident to answer nine yes–no questions from the Mood Module before the BDI–II and RADS were administered. The psychiatrists recorded their diagnoses before the Mood Module, the BDI–II, and the RADS were administered. The study was conducted with the approval of our Institutional Review Board.

RESULTS

Order of Administration

Because both the BDI–II and the RADS require less than 10 min to complete and contain similar questions, we first tested whether the order in which the BDI–II and the RADS had been administered might have influenced the patients' responses. The mean BDI–II and RADS total scores for the 47 adolescents who were administered the RADS first were, respectively, 25.57 (SD = 14.96) and 77.64 (SD = 17.75), whereas the mean BDI–II and RADS total scores for the 53 adolescents who were administered the BDI–II first were, respectively, 23.92 (SD = 12.40), and 79.06 (SD = 15.66). There were no significant mean differences with respect to the order in which the BDI–II or the RADS was administered; the respective $t(98)$ statistics for the BDI–II and RADS mean differences were .60 and –.42, ps > .10. Therefore, order of administration was not controlled for in the subsequent analyses.

Correlations With Sample Characteristics

The correlations of the BDI–II total scores with sex (0 = male, 1 = female), being White (0 = no, 1 = yes), age (years), or taking an antidepressant at the time of testing (0 = no, 1 = yes) were, respectively, .22 (p < .05), .11, .18, and .19, and the correlations of the RADS total scores with these same respective variables were .18, .12, .07, and .11. The only significant correlation was between the BDI–II total scores and sex indicating that being female was positively associated with increasing severity of self-reported depression as measured by the

BDI–II. However, sex explained < 5% of the variance in the BDI–II total scores. Because the magnitudes of the eight correlations represented small effect sizes (rs < .30) according to Cohen's (1992) guidelines, none of the characteristics was controlled for.

Severity of Depression

The mean BDI–II total score was 24.70 (SD = 12.50, range = 1 to 59), and the RADS total score was 78.39 (SD = 16.60, range = 41 to 111). These mean values indicated that the overall sample of adolescent inpatients was, respectively, moderately and clinically depressed according to Beck et al.'s (1996) and Reynolds's (1987) cutoff score interpretative guidelines. The correlation between the BDI–II and the RADS total scores was .84, $p < .001$.

Internal Consistencies

The Cronbach coefficient αs for the BDI–II and RADS were respectively, .92 and .91, and indicated high internal consistency. According to Cicchetti's (1994) guidelines, the internal consistencies of both instruments were excellent for clinical purposes.

Diagnostic Discrimination

For the 50 adolescents who were administered the Mood Module, 39 (78%) were classified by the Mood Module as meeting criteria for a MDD. With respect to the agreement between the Mood Module and the psychiatrists' medical chart diagnoses, there were 33 (66%) patients for whom both a psychiatrist and the Mood Module agreed that a diagnosis of a MDD was warranted; 11 (22%) patients for whom a psychiatrist and the Mood Module agreed that a diagnosis of a MDD was not warranted; 6 (12%) patients for whom a psychiatrist did not think that a diagnosis of a MDD was warranted, but the Mood Module indicated that the criteria for a MDD were present; and no (0%) patients for whom a psychiatrist thought that a diagnosis of a MDD was warranted, but the Mood Module indicated that the criteria for a MDD were not present. The overall rate of agreement was 88%, Cohen's $\kappa = .71$, $p < .001$. Therefore, we concluded that a MDD was being diagnosed reliably.

The mean BDI–II and RADS total scores of the 57 adolescents who were diagnosed with a MDD were, respectively, 30.09 (SD = 12.80) and 84.81 (SD = 15.23), whereas the mean BDI–II and RADS total scores of the 43 adolescents who were not diagnosed with a MDD were, respectively, 17.56 (SD = 11.27) and 69.88 (SD = 14.51). The mean BDI–II, $t(98) = 5.10$, $p < .001$, and RADS, $t(98) = 4.95$, $p < .001$, total scores of the adolescents who were diagnosed with a MDD were significantly higher than those who were not diagnosed with a MDD. The point-biserial correla-

tions of the BDI–II and RADS total scores with being diagnosed with a MDD (0 = no, 1 = yes) were, respectively, .46 and .45, *p*s < .001, indicating medium effect sizes according to Cohen (1992). The effect sizes of both instruments were obviously comparable.

To determine the overall effectiveness of each instrument to differentiate adolescents who were and were not diagnosed with a MDD by a psychiatrist, binormal receiver-operating-characteristic (ROC) curves (Somoza & Mossman, 1991) were calculated using ROCKIT (Metz, 1998) and are shown in Figure 1. As Figure 1 indicates, the shapes of the ROC curves were similar, and the areas under the ROC (AUR) curves for the BDI–II and the RADS total scores were, respectively, .78 (95% CI, .68 to .86) and .76 (95% CI, .66 to .84). The AUR is an overall index of the accuracy of discrimination provided by a scale, and an AUR of .50 represents chance discrimination. These AUR values indicate that again both instruments afforded comparable levels of differentiation. A BDI–II total cutoff score of 24 and above had the highest clinical efficiency (72%) with a sensitivity rate of 74% and a specificity rate of 70%, a positive predictive value of .76, and a negative predictive value of .67. Similarly, a RADS total cutoff score of 70 and above afforded maximum clinical efficiency (70%) in differentiating between adolescents with and without a MDD with a sensitivity rate of 86%, a specificity rate of 49%, positive predictive value of .69, and a negative predictive value of .72.

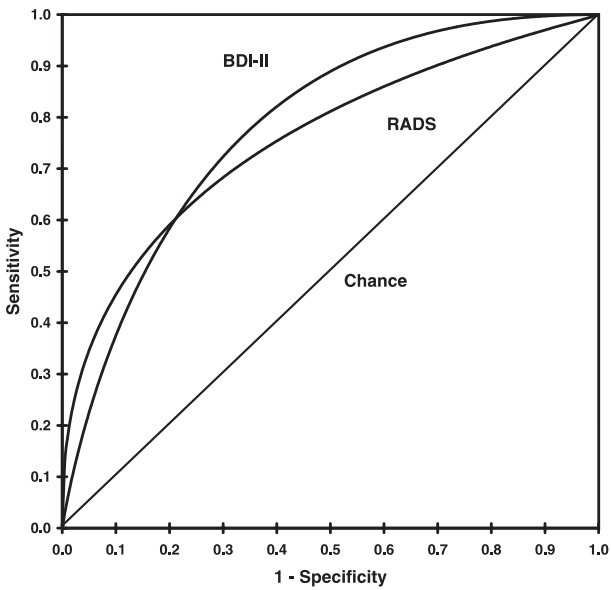

FIGURE 1 Binormal receiver-operating-characteristic (ROC) curves for Beck Depression Inventory–II (BDI–II) and Reynolds Adolescent Depression Scale (RADS).

DISCUSSION

The overall pattern of results indicated that the BDI–II and the RADS demonstrated comparable psychometric characteristics. Both instruments had high internal consistency (coefficient αs > .90), and inpatient adolescents who were diagnosed with a MDD described more severe depression than the inpatient adolescents who were not diagnosed with a MDD on both instruments, *p*s < .001. The present coefficient α of .92 for the BDI–II is the same as that found by Steer et al. (1998) with adolescent psychiatric outpatients, and the present coefficient α of .91 for the RADS is typical of that usually found with the RADS (Reynolds, 1987). The strong positive correlation ($r = .84$, $p < .001$) of the BDI–II with the RADS also supported the convergent validity of the BDI–II with respect to measuring the severity of depression in adolescent psychiatric inpatients. However, the magnitude of this correlation is higher than that found in the range of correlations (.70 to .76) between the RADS and BDI–IA given by Reynolds (1987, p. 21). Perhaps this strong correlation is partially attributable to the diagnostic composition of the sample; 85% of the sample were diagnosed with mood disorders. The correlations that Reynolds (1987) reported between the BDI–IA and the RADS were calculated for either nonclinical samples or clinical samples that contained a much smaller proportion of adolescents who were diagnosed with mood disorders. Of course, the strong correlation might also be attributable to different versions of the BDI having been used.

The RADS was not significantly associated with any of the adolescents' background characteristics, but the BDI–II was positively correlated with being female ($r = .22$, $p < .05$). This correlation is only slightly higher than that ($r = .19$, $p < .01$) found by Steer et al. (1998) between the BDI–II and sex in 210 adolescent psychiatric outpatients. The mean BDI–II score of their female outpatients was approximately 5 points higher than that of their male adolescents. Reynolds (1987, p. 14) also reported that female adolescents' mean RADS total score tends to be 5 to 7 points higher than that of male adolescents. However, Beck et al. (1996) found that BDI–II total scores were not meaningfully associated with being White and age in their predominately adult psychiatric sample, and Reynolds (1987) concluded that the impact of ethnicity and age on RADS total scores was minimal. Age was not significantly correlated ($r = .18$) with the BDI–II in this study, but Steer et al. (1998) found the same correlation between the BDI–II and age in their outpatient sample. In any event, the magnitudes of all of the correlations of the BDI–II and the RADS with the present adolescents' characteristics reflected small effect sizes.

Because the BDI–II specifically addresses *DSM–IV* MDD criteria, we would not have hypothesized that the RADS would afford as much differentiation between adolescent inpatients who were and were not diagnosed with a MDD as the BDI–II would have. Reynolds (1987) did not construct the RADS for diagnostic purposes, and the RADS contains several sentences that do not reflect *DSM–IV* diagnostic criteria for depression. However, the magnitudes of the AURs for both instruments only differed by two hundredths of a point. Conversely, it is interesting

that an instrument that was originally developed to assess the severity of self-reported depression in adult psychiatric patients should differentiate adolescent psychiatric outpatients as well as an instrument that was developed to measure self-reported depression in adolescents.

The ROC analyses, however, indicate that neither instrument was highly effective in screening for a MDD. For example, a BDI–II cutoff score of 24 and above had the highest clinical efficiency, but the sensitivity and specificity rates were in the mid-70% range; a RADS cutoff score of 70 and above achieved a sensitivity rate of 86%, but its specificity rate was only 49% or approximately chance. Such lack of differentiation, especially the low rates of specificity, might again be attributable to the sample being moderately depressed and predominately (85%) composed of adolescents who were diagnosed with mood disorders. Adolescents who are admitted to an inpatient unit, especially those who are involuntarily admitted, might be expected to describe at least some symptoms of dysphoria and anxiety, regardless of whether they meet criteria for a mood disorder. The BDI–II and RADS might have been able to discriminate more effectively between adolescents who were and were not diagnosed with a MDD if the sample had been drawn from a school or outpatient service. Finally, it should be remembered that these sensitivity and specificity rates for the BDI–II and RADS are based on the 57% prevalence of a MDD in this sample and that the cutoff scores found in this study may not be the most clinically efficient ones in screening for a MDD in samples with different underlying prevalences of a MDD.

With respect to the study's limitations, in this study we used the psychiatrists' medical chart diagnoses as the "gold standard." Although the PRIME–MD Mood Module was employed to confirm that a reliable diagnosis of a MDD had been made, we did not employ structured clinical interviews, such as the *ChIPS—Children's Interview for Psychiatric Syndromes* (Weller, Weller, Rooney, & Fristad, 1999), to establish any of the adolescents' other diagnoses. Therefore, the validity of the other diagnoses is suspect. Furthermore, this sample was predominately composed of White adolescents being evaluated at an inpatient psychiatric unit in a general hospital located in a middle class suburban community. Future research needs to investigate the psychometric properties of the BDI–II with different ethnic and socioeconomic samples of adolescents drawn from diverse clinical populations.

REFERENCES

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

Beck, A. T., & Steer, R. A. (1993). *Manual for the Beck Depression Inventory.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Metz, C. E. (1998). ROCKIT, Windows 95 (Version 0.9.1 Beta) [Computer software]. Retrieved from http://www-radiology.uchicago.edu/toopage11.htm

Reynolds, W. W. (1987). *Reynolds Adolescent Depression Scale: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M. (1994). Assessment of depression in children and adolescents by self-report questionnaires. In W. M. Reynolds & H. F. Johnston (Eds.), *Handbook of depression in children and adolescents: Issues in clinical child psychology* (pp. 209–234). New York: Plenum.

Reynolds, W. M., & Mazza, J. J. (1998). Reliability and validity of the Reynolds Adolescent Depression Scale with young adolescents. *Journal of School Psychology, 36,* 295–312.

Roberts, R. E., Lewinsohn, P. M, & Seeley, J. R. (1995). Symptoms of *DSM–III–R* major depression in adolescence: Evidence from an epidemiological survey. *Journal of American Academy of Child and Adolescent Psychiatry, 34,* 1608–1617.

Somoza, E., & Mossman, D. (1991). ROC curves and the binormal assumption. *Journal of Neuropsychiatry and Neuroscience, 3,* 436–439.

Spitzer, R. L., Williams, J. B. W., Kroenke, K., Linzer, M., deGruy, F. V., III, Hahn, S. R., et al. (1994). Utility of new procedure for diagnosing mental disorders in primary care: The PRIME–MD 1000 study. *Journal of the American Medical Association, 272,* 1749–1756.

Spitzer, R. L., Williams, J. B. W., Kroenke, K., Linzer, M., deGruy, F. V., III, Hahn, S. R., et al. (1995). *Prime–MD instruction manual updated for* DSM–IV. New York: Biometrics Research Department, New York State Psychiatric Institute.

Steer, R. A., & Beck, A. T. (1988). Use of the Beck Depression Inventory, Hopelessness Scale, Scale for Suicide Ideation, and Suicidal Intent Scale with adolescents. *Advances in Adolescent Mental Health: Depression and Suicide, 3,* 219–231.

Steer, R. A., & Beck, A. T. (2000). The Beck Depression Inventory–II. In W. E. Craighead & C. B. Nemeroff (Eds.), *The Corsini encyclopedia of psychology and behavioral science* (3rd ed., Vol. 1, pp. 178–179). New York: Wiley.

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.

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the surgeon general.* Rockville, MD: U.S. Department of Heath and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

Weller, E. B., & Weller, R. A. (2000). Depression in adolescents: Growing pains or true morbidity? *Journal of Affective Disorders, 61,* S9–S13.

Weller, E. B., Weller, R. A., Rooney, M. T., & Fristad, M. A. (1999). *ChIPS—Children's Interview for Psychiatric Syndromes.* Washington, DC: American Psychiatric Press.

Winter, L. B., Steer, R. A., Jones-Hicks, L., & Beck, A. T. (1999). Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for primary care. *Journal of Adolescent Health, 24,* 389–394.

Robert A. Steer
UMDNJ–SOM
Department of Psychiatry
42 East Laurel Road
SCC 1600
Stratford, NJ 08084–1391
E-mail: steer@umdnj.edu

Received May 15, 2001
Revised October 10, 2001