Reliability and Validity of the Hamilton Depression Inventory: A Paper-and-Pencil Version of the Hamilton Depression Rating Scale Clinical Interview

William M. Reynolds  
University of British Columbia

Kenneth A. Kobak  
University of Wisconsin—Madison

A self-report, paper-and-pencil version of the Hamilton Depression Rating Scale (HDKS; Hamilton, 1960) was developed. This measure, the Hamilton Depression Inventory (HDI; W. M. Reynolds & K. A. Kobak, 1995) consists of a 23-item full form, a 17-item form, and a 9-item short form. The 17-item HDI form corresponds in content and scoring to the standard 17-item HDRS. With a sample of psychiatric outpatients with major depression \( n = 140 \), anxiety disorders \( n = 99 \), and nonreferred community adults \( n = 118 \), the HDI forms demonstrated high levels of reliability \( r_s = .91 \) to .94, \( r_h = .95 \) to .96. Extensive validity evidence was presented, including content, criterion-related, construct, and clinical efficacy of the HDI cutoff score. Overall, the data support the reliability and validity of the HDI as a self-report measure of severity of depression.

Depression is one of the most prevalent mental health problems in the United States (Kessler et al., 1994; Regier et al., 1988), with 1-month prevalence rates ranging from 2% to 3% for major depression and over 6% for any form of affective disorder (Regier et al., 1993). The fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994) reports a point-prevalence rate for major depression of between 5% and 9% for women and between 2% and 3% for men.

For decades, mental health professionals have relied on semi-structured clinical interviews and self-report measures for the identification of depression in adults. The use of these measures, most of which are considered severity measures of depression, does not provide for the formal diagnosis of depression. However, this does not limit their usefulness for the evaluation of the severity of depressive symptomatology in clinical and research applications (Reynolds, 1994).

The Hamilton Depression Rating Scale

The Hamilton Depression Rating Scale (HDRS; Hamilton, 1960, 1967) was one of the first semistructured interview measures developed for the clinical evaluation of severity of depression in adults. The HDRS is one of the most frequently used clinical interview measures of the severity of depression (e.g., Edwards et al., 1984; Endicott, Cohen, Nee, Fleiss, & Sarantakos, 1981; Fava, Kellner, Munari, & Pavan, 1982; Sayer et al., 1993; Williams, 1988) and is often used as the criterion measure against which self-report measures of depression are validated (e.g., Carroll, Feinberg, Smouse, Rawson, & Greden, 1981; Montgomery & Asberg, 1979).

Although frequently used in research, the relative lack of standardized administration instructions and scoring criteria for the HDRS has been problematic. Cicchetti and Prusoff (1983) in a study of interrater reliability of a 22-item version of the HDRS found low levels of reliability for individual items, with 14 of the 22 items demonstrating intraclass correlation coefficients of less than .40. The lack of scoring guidelines has led a number of investigators (e.g., Endicott et al., 1981; Miller, Bishop, Norman, & Maddever, 1985; Williams, 1988) to develop item modifications and suggest administration and scoring procedures. The issues of training, scoring, and differences in version of the HDRS used in research have been evaluated and discussed by Hooijer et al. (1991), who found small but meaningful differences across HDRS versions and training.

Several self-report versions of the HDRS have been developed by researchers, two of which were based on computer admin-
The Hamilton Depression Inventory

The HDI consists of 23 items (symptoms) that are evaluated by 38 probes or questions. Eleven of the HDI items use multiple questions (2–4 probes) to evaluate the symptom content of that item. For example, on Item 10, which examines the symptom of psychological aspects of anxiety, two questions are presented to the examinee: One question inquires about the frequency of feeling anxious over the past 2 weeks (rated from 0 = not at all or rarely to 4 = almost all of the time), and the second question evaluates the severity of the anxious feelings (rated from mild to very severe). The respondent is instructed to skip the second part (symptom severity rating) if the response to the initial question was 0 (not at all or rarely). For items with multiple questions, questions are summed and weighted to produce an item score consistent with the range (0–2 or 0–4) outlined by Hamilton (1960, 1967). The HDI is designed for use with persons 18 years and older and requires a 5th grade reading level. The HDI takes approximately 10 min to complete, although greater time may be required by elderly persons, individuals with severe psychomotor retardation, or persons who are slow readers. The HDI evaluates the severity of depressive symptoms over the previous 2 weeks.

The presentation and content of items on the HDI differs from most other paper-and-pencil self-report measures of depression. This divergence is in part due to Reynolds and Kobak’s (1995) goal to create a self-report measure that emulates a clinician-administered interview. In a clinical interview for depression, the clinician often evaluates multiple subsymptoms or components of a symptom, as well as determines the duration and frequency of symptom occurrence. For example, to assess the symptom of insomnia, it is useful to determine the length of time required by the client to fall asleep, as well as how often over the past several weeks the client has had difficulty falling asleep. Likewise, some symptoms of depression are multifaceted in their clinical domain or symptom expression. Thus, dysphoric mood may be evaluated by such symptom components as feeling sad or blue, the intractability of such feelings, and behavioral elements of tearfulness or crying. Our earlier work (Kobak et al., 1990) with a 17-item computer-administered form of the HDRS demonstrated the psychometric and clinical usefulness of the multiple question approach for the emulation of clinician-administered clinical interviews. As a function of the multiple questions, the 23-item full-scale HDI includes 38 questions, the HDI–17 consists of 31 questions, and the 9-item HDI Short Form includes 15 questions. By including multiple questions for many symptoms of depression, the HDI provides for a more comprehensive evaluation of individual symptoms than is typically assessed by self-report depression measures.

As noted above, the first 17 items on the HDI evaluate symptoms of depression formulated by Hamilton (1960) for the HDRS. Six additional items were added to the HDI to enhance the content validity of this scale by including symptoms of major depressive disorder and dysthymic disorder delineated by the revised third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–III–R; American Psychiatric Association, 1987) and DSM–IV (American Psychiatric Association, 1994) and including the alternative criterion B for dysthymic disorder presented in DSM–IV. These additional items and their symptom identification in the DSM–IV include hypochondriasis (p. 321), detachment (p. 321), feelings of worthlessness (p. 321), helplessness–pessimism (p. 718), hopelessness (p. 320), and difficulty making decisions (p. 322). In this manner, the HDI consists of all 23 items (the 17 HDRS symptoms plus 6 additional symptoms), and the HDI–17 is composed of 17 items analogous to the 17 HDRS symptoms. Thus, the HDI–17 provides a score based on item content similar to the HDRS, and the HDI provides this coverage with additional DSM–IV symptoms.

The response format for individual items varies in scoring from 0–2 or 0–4, consistent with the scoring system described...
by Hamilton (1960). For example, insomnia on the HDRS is evaluated by three items that examine initial, middle, and late insomnia, with each item scored on a scale of 0–2. On the HDI, the three insomnia items are each evaluated by two questions that assess the frequency (e.g., number of nights per week that the respondent had difficulty falling asleep) and severity (e.g., length of time it took to fall asleep) of insomnia, with a scoring algorithm designed to emulate the response score of the HDRS.

Method

Participants

The participants were 357 adults (212 women and 145 men) between 18 and 81 years of age, with a mean age of 38.70 years. The sample consisted of 140 outpatients with a DSM-III-R (American Psychiatric Association, 1987) diagnosis of major depression, 99 outpatients with a diagnosis of an anxiety disorder (e.g., generalized anxiety disorder, social phobia, panic disorder, etc.), and 118 nonpsychiatric community controls. Demographic information on the total sample and for each diagnostic group is presented in Table 1. Most of the participants with major depression or an anxiety disorder were relatively pure cases because of criteria for inclusion in pharmacological intervention studies. There was some comorbidity, although in all cases, additional diagnoses were secondary to either major depression or an anxiety disorder. In the sample with major depression, comorbidity was found in 24.3% of the sample and included anxiety disorders (10%), dysthymic disorder (3.6%), substance abuse (6.4%), and personality disorders (4.3%). In the anxiety disorder group, comorbidity was found in 18.2% of the sample and included dysthymic disorder (7.1%), depressive disorder not otherwise specified (7.1%), and personality disorders (4.0%). The difference in age between groups was not significant, nor was the difference in proportions of men and women across diagnostic groups, $\chi^2(2, N = 357) = 2.67$. By ethnicity, the sample was 90% White, 6% African American, 2% Hispanic, and 2% Asian. A significant difference in ethnicity was found, with a higher proportion of non-White participants among community control participants than among participants with anxiety disorders or major depression, $\chi^2(6, N = 348) = 23.38, p < .01$. Overall, the sample's education level was equivalent to approximately two years of college. There was a significant difference in years of education between groups, $F(2, 329) = 5.70, p < .01$. The nonpsychiatric community group had approximately one year more education than did the group of persons with major depression (Scheffe post hoc comparison $p < .05$). The groups did not differ significantly in their reported occupational status, $F(2, 331) = .87$.

Diagnoses were made using the Structured Clinical Interview for DSM-III-R (SCID; Spitzer, Williams, Gibbon, & First, 1988), which was administered by a trained psychologist or social worker, or by a psychiatric nurse. Diagnosticians had extensive training and experience (3–4 years) using the SCID. The exception to this was one clinician who had received extensive training by one of the authors of SCID. Interrater reliability for diagnoses was not available. Community control participants were evaluated as being free from psychopathology by Kenneth A. Kobak using the SCID. Participants with major depression or an anxiety disorder were recruited from screening evaluations for participation in ongoing pharmacological treatment studies being conducted at a major research university and a research section of a health maintenance organization, both located in the midwestern United States, as well as through newspaper advertisements. Participants with psychiatric diagnoses who were not involved in pharmacological studies were paid for completion of the assessment protocol. Control participants were recruited through newspaper advertisements and bulletin boards in the community and paid $10 or $20 for their participation, with the larger amount for the completion of the HDRS clinical interview. Procedures for the recruitment and assessment of participants were approved by the University of Wisconsin Center for Health Sciences Human Subjects Committee.

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Controls</th>
<th>Anxiety disorders</th>
<th>Major depression</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>118</td>
<td>99</td>
<td>140</td>
<td>357</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>37.09</td>
<td>39.11</td>
<td>39.78</td>
<td>38.70</td>
</tr>
<tr>
<td>$SD$</td>
<td>14.40</td>
<td>10.61</td>
<td>13.59</td>
<td>13.36</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% men</td>
<td>38</td>
<td>38</td>
<td>47</td>
<td>41</td>
</tr>
<tr>
<td>% women</td>
<td>62</td>
<td>62</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% White</td>
<td>79</td>
<td>96</td>
<td>94</td>
<td>90</td>
</tr>
<tr>
<td>% African American</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>% Asian</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>15.52</td>
<td>14.75</td>
<td>14.20</td>
<td>14.67</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.05</td>
<td>2.26</td>
<td>2.07</td>
<td>2.15</td>
</tr>
<tr>
<td>Occupation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>4.43</td>
<td>4.10</td>
<td>4.56</td>
<td>4.39</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.85</td>
<td>2.43</td>
<td>2.47</td>
<td>2.59</td>
</tr>
</tbody>
</table>

Note. Analyses are between diagnostic groups.

* Occupation based on a scale from 0 (unemployed) to 9 (executive).

Instruments

In addition to the HDI and the HDRS, a number of other measures were administered to demonstrate convergent and discriminant validity and are described below. These included self-report measures of depression, anxiety, self-esteem, hopelessness, suicidal ideation, and social desirability, with the latter designed to provide information on discriminant validity. Not all participants completed the additional self-report measures listed below.

Beck Depression Inventory. The Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) is a 21-item alternative format measure designed primarily for use with adults. Psychometric characteristics of the BDI have been described elsewhere (e.g., Beck & Steer, 1987; Reynolds & Gould, 1981). In their study of the computer-administered version of the HDRS, Kobak et al. (1990) reported a correlation of .92 between the BDI and the clinician-administered HDRS and a correlation of .93 between the BDI and the 17-item computer-administered version of the HDRS.

Adult Suicidal Ideation Questionnaire. The Adult Suicidal Ideation Questionnaire (ASIQ; Reynolds, 1991a) was used as a convergent validity measure to assess levels of suicidal ideation. The ASIQ is a 25-item adult form of the Suicidal Ideation Questionnaire (Reynolds, 1987), the latter designed for use with adolescents. The ASIQ assesses suicidal thoughts over the past month and uses a 7-point response format ranging from 0 (never had the thought) to 6 (had the thought almost every day). Reynolds (1991b), in a sample of 474 college students, found high reliability ($r_a = .97, r_n = .86$) and significant correlations with measures of depression, hopelessness, and self-esteem, with a cor-
relation of \( r(471) = .60 \) between the ASIQ and BDI. Reynolds, Kobak, and Greist (1990) reported a correlation of \( .81 (p < .001) \) between the ASIQ and the item specific to suicide on the HDRS. Reynolds, Kobak, and Greist (1993) found high levels of reliability (\( r_s = .95 \) to .97) for the ASIQ with a sample of 700 psychiatric outpatients, including 372 persons with major depression, and a test–retest reliability coefficient of .95 with a subsample of 60 psychiatric outpatients.

**Beck Anxiety Inventory.** To examine the relationship between anxiety and the HDI, we used the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988). The BAI consists of 21 items and uses a 0-4 point 0 (not at all) to 3 (severely—I could barely stand it) response format to assess symptom severity over the past week. Items are keyed in a high anxiety direction. For the development sample of psychiatric outpatients, Beck et al. (1988) reported an internal consistency reliability of .92, and a correlation of .51 with the Hamilton Anxiety Rating Scale and of .48 with the BDI. In a large sample investigation with college students, Reynolds (1991b) reported an internal consistency reliability coefficient of .89 for the BAI and a correlation of .53 with the BDI.

**Rosenberg Self-Esteem Scale.** The Rosenberg (1965) Self-Esteem Scale (RSES) was used as a measure of general self-esteem. The RSES is a 10-item inventory designed to measure overall or general self-esteem. Each item is answered along a 4-point scale from 1 (strongly agree) to 4 (strongly disagree). Although originally scored as a Guttman scale, it has been scored by many researchers as a Likert-type scale (Crandall, 1975; Reynolds, 1988). Items on the RSES are keyed in the positive direction, with a high score indicating positive self-esteem. Adequate internal consistency reliability \( (r_s, .82 \) and .83) has been reported (e.g., Reynolds, 1988; Zorich & Reynolds, 1988) with college students.

**Beck Hopelessness Scale.** Hopelessness, or a pessimistic view of the future, has also been formulated as a psychological construct and operationalized by Beck and colleagues (Beck, Weissman, Lester, & Trexler, 1974) in the form of the Beck Hopelessness Scale (BHS). The BHS consists of 20 items and uses a true–false response format with items keyed in a hopelessness direction. Reynolds (1991b) reported an internal consistency reliability coefficient of .81 for the BHS and a correlation of .55 between the BHS and the BDI.

**Marlowe-Crowne Social Desirability Scale.** Social desirability was assessed using a short form of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) (MCSDS-SF). This measure was used as a methodological check to assess the influence of response bias on self-reported depression on the HDI as well as to provide evidence of discriminant validity. The original form of the Marlowe-Crowne (Crowne & Marlowe, 1960) scale contains 33 items assessing the individual's social desirability response set. Items use a true–false response format and are keyed in a socially desirable direction. The 13-item short form of the Marlowe-Crowne scale developed by Reynolds (1982) was used in the current study. Reynolds (1982) reported internal consistency reliability of .76 for this short form and a high correlation \( (r = .93) \) between the short form and the standard 33-item form.

**Procedure**

All participants were administered the HDI and a majority (92%) were also administered the HDRS. For those participants who were administered the HDRS clinical interview, administration was counter-balanced with the administration of the HDI. Only participants who had never been evaluated with the HDRS were included in this study. Nine interviewers administered the clinician version of the HDRS, although the majority (60%) of HDRS interviews were conducted by Kenneth A. Kobak. The remaining eight interviewers conducted between 7 and 33 interviews, with a median of 21 interviews. Prior to conducting the interviews, interviewers received extensive training in the administration and scoring of the HDRS and were evaluated for scoring agreement using videotaped case interviews. The HDRS training protocol was developed by William M. Reynolds and has been used in clinical research applications for the previous 12 years. This training protocol specifies the scoring of the HDRS items to the half-point. The scoring of HDRS items in this manner (i.e., .5, 1.0, 1.5, 2.0, 2.5, etc.) provides for greater fidelity in the clinical evaluation of depressive symptoms. Interviewers included psychologists, social workers, and advanced graduate students, most of whom had previous experience with this interview. In cases in which the clinician HDRS version was the second measure administered, clinicians were blind to the results of the HDI. The subsample of 329 participants who completed the HDRS included 135 men and 194 women. By diagnostic group, the subsample was composed of 135 persons with a diagnosis of major depression, 77 persons with an anxiety disorder, and 117 nonpsychiatric community controls.

The majority of participants were evaluated with the research protocol in a research room provided by the Department of Counseling Psychology at the University of Wisconsin—Madison. Other evaluations were conducted at an outpatient mental health setting and at the University of Wisconsin—Madison Center for Health Sciences. Because of work schedules for participants, many of the evaluations were conducted in the evenings and on weekends.

Participants completed the HDI either immediately before or after the HDRS. Most of the participants also completed the ASIQ \( (n = 345) \) and the BDI \( (n = 329) \). The number of participants who completed the other self-report measures described above ranged from 91 (BAI) to 106 (RSES). Participants who were recruited by newspaper and bulletin board advertisements and received a psychiatric diagnosis and who were not subsequently enrolled in a pharmacological treatment study were referred for additional services and provided a list of community mental health service providers.

**Results**

**Preliminary Analysis**

Statistical analyses were initially conducted on a 25-item preliminary form of the HDI. On the basis of low item-total scale correlations \( (r < .20) \) and exploratory factor analysis (low factor loadings), two items on the preliminary form of the HDI were dropped from the final version. These items were specific to social isolation (an associated feature in DSM-IV) and weight gain (a component of weight–appetite disturbance). As noted earlier, multiple questions are used by many of the HDI items, with between one and four questions per HDI item. Thus, the final version of the HDI consisted of 23 items that are based on responses to 38 questions. This 23-item form constituted the basic HDI.

In addition to final item selection for the basic HDI, preliminary statistical analyses were conducted for the selection of items that would constitute a short form of the HDI. Criteria for the selection of items for the HDI Short Form (HDI-SF) included high item-total scale correlations and the clinical efficacy of items to discriminate between persons with major depression and nonpsychiatric community controls. A discriminant function analysis of the 23 HDI items between persons with major depression and community controls was computed. One discriminant function was found with eight items demonstrating Wilks's lambda values of .200 to .466 with \( F(1, 256) \) values ranging from 1024.8 to 293.8. In addition to these eight items, one additional item specific to suicidal behavior was added to the short form. Although this item was somewhat less
discriminating between groups, \( \Lambda = .594, F(1, 256) = 174.3 \), it was chosen for inclusion on the short form because of its clinical importance as a symptom of depression and a potential marker of self-destructive behavior. Thus, nine items were selected for evaluation as a short form of the HDI. Items included on the HDI-SF demonstrated item-total scale correlations between .63 and .87 with the total 23-item HDI.

**Descriptive Statistics**

Similar levels of depressive symptomatology were found on each form of the HDI and the HDRS for men and women, with results of t tests between gender indicating nonsignificant gender differences on all scales (t values ranged from .22 to .31). Given the potential for gender differences within groups, a series of two-way analyses of variance (ANOVAs; gender by diagnostic group) were computed. All main effects for gender were nonsignificant (p > .10), as were the interaction terms between gender and diagnostic group (p > .10), suggesting nonsignificant differences between men and women within diagnostic groups on all HDI forms. Correlation coefficients between participants' age and scores on the HDI, HDI-17, HDI-SF, and HDRS were low, ranging from ~.02 to .02, all nonsignificant.

To further examine possible gender differences, we examined HDI item scores between men and women. To control for multiple comparisons, we used a Bonferroni procedure (Dunn, 1961), with the familywise alpha level set at .05, resulting in an experimentwise alpha of .002 (i.e., .05/23) for testing the statistical significance of score differences on HDI items between men and women. Gender differences on HDI items were not statistically significant.

**Internal Consistency Reliability**

The reliability of the HDI was examined from the perspectives of internal consistency reliability using Cronbach's (1951) coefficient alpha (\( \alpha \)) and test-retest reliability (\( r_{tt} \)) using a 1-week retest interval. The internal consistency (coefficient alpha) reliability, mean inter-item correlation coefficient, median item-with-total scale correlations, and standard errors of measurement of each form of the HDI and the HDRS were computed for the total sample and are presented in Table 2. The internal consistency reliability of all forms of the HDI was high and ranged from \( \alpha = .91 \) to .94 for the total sample. These coefficients are of sufficient magnitude to suggest HDI score accuracy for clinical as well as research applications and are similar to that obtained for the HDRS. Of particular note is the relatively high internal consistency reliability of the HDI-SF, which although consisting of only nine items, demonstrated a high level of item homogeneity, with a total sample \( r_{tt} \) of .93 and a median item-total scale correlation coefficient of .76. Internal consistency reliability estimates were similar for men and women, and were of the same magnitude as those reported for the total sample.

As a further examination of item homogeneity and partial evidence of content validity (e.g., Guilford, 1954), the item-with-total scale correlations corrected for part–whole redundancy (\( r_{uw} \)) for the 23-item HDI for the total sample and for men and women were computed. Item-total scale correlation coefficients were high, with 20 of the 23 items demonstrating coefficients of .40 and higher. The two lowest correlations were found on items related to loss of insight (\( r_{uw} = .31 \)) and weight loss (\( r_{uw} = .27 \)). As shown by the median \( r_{uw} \) values reported in Table 2, item-total scale correlations were moderate to high across all forms of the HDI as well as the HDRS.

**Test–Retest Reliability**

The test–retest reliability (\( r_{tt} \)) of the HDI was examined in a sample of 129 participants who were retested approximately one week after an initial assessment. Participants included persons from the community (n = 83) and psychiatric (n = 46) samples. There were 50 men and 79 women in the test–retest sample (37 and 65, respectively, for the HDRS test–retest sample). The mean retest interval was 6.4 days with a range of 3–9 days and a mode of 7 days. All retesting was completed prior to any active treatment.

Table 3 provides the results of the test–retest reliability of the HDI for the total retest sample. As shown, the test–retest reliability coefficient of .96 for the HDI found in this sample indicates a very high degree of rank-order stability of HDI scores. Similarly high levels of test–retest reliability were found for the HDI-17 (\( r_{tt} = .96 \)) and the HDI-SF (\( r_{tt} = .95 \)). HDRS clinical interviews were administered on both occasions to 102 of the 129 (93%) participants with a test–retest reliability of .96 and a mean score difference of .85 points between assessments. Test–retest reliability coefficients computed separately for men and women were of similar magnitude to those found for the total retest sample.

Test–retest reliability coefficients were also computed for clinical and nonclinical participants. For participants with anxiety or depressive disorders, the test–retest reliability coefficients were .89, .90, and .87 for the HDI, HDI-17, and HDI-SF, respectively. For nonclinical participants, test–retest coefficients were .82, .82, and .67 for the HDI, HDI-17, and HDI-SF, respectively. These latter coefficients are somewhat attenuated because of the restricted range of HDI scores found in the nonclinical sample.

Because the test–retest reliability coefficient was computed
using the traditional Pearson product-moment correlation between Time 1 and Time 2, it does not provide an index of raw score stability over time. To examine the magnitude of raw score change over time, we computed dependent sample *t* tests between HDI scores for the two occasions for each scale version. These are presented in Table 3. As shown, the mean difference in HDI scores between the two testings was small, with a less than 1-point (Δ = .90) decrease in HDI score at Time 2. Although small, this decrease was statistically significant, *t*(128) = 3.10, *p* < .01. The mean difference between testings on the HDI was somewhat higher for men (*M*diff = 1.28) than for women (*M*diff = .65). A similar level of score change was found on the HDRS for the total sample (Δ = .85), and for men (Δ = 1.12) and women (Δ = .70). Overall, these reliability coefficients are high and are associated with minimal changes in raw scores between the two assessments.

**Equivalence Between the HDI and the HDRS**

A critical aspect of validity for the HDI is its equivalence with the HDRS clinical interview. The HDRS provides the most appropriate form of the HDI for this comparison because of its design and content correspondence to the HDRS. Means and standard deviations for the clinical interview HDRS and the paper-and-pencil HDRS—17 for each of the diagnostic groups, the control group, and the total sample are presented in Table 4. Similar levels of depressive symptomatology were found on the HDRS and HDI—17 across diagnostic groups. Mean differences between the HDI and HDRS were small, ranging from .17 to .56, suggesting a high degree of score similarity. For the total sample, the mean difference between forms was .17. Table 4 also presents the results of *t* tests for dependent samples that indicate the mean differences between the HDRS and HDI—17 forms for the various groups of participants as well as for the total sample were not significantly different. Between-group differences on the HDI and HDRS are discussed below in the section on contrasted groups validity.

As further evidence for the equivalence and validity of the HDI as a self-report paper-and-pencil-administered version of the HDRS clinical interview, a strong relationship should be found between these two scales. It was expected that individuals would maintain a similar rank order across forms of the Hamilton measures. The correlation between participants' scores on the HDRS and the HDI—17 was .95, *p* < .001, which is high and suggests a strong relationship between clinical interview and paper-and-pencil versions. Correlation coefficients between the HDI and HDI—SF and the HDRS are presented in Table 5. These correlations were also high, with *r* (327) of .95 and .93 for the HDI and HDI—SF, respectively. These data provide strong support for the equivalence and validity of the HDI scales as paper-and-pencil-administered versions of the clinical interview HDRS.

**Convergent Validity: Relationships With Measures of Related Constructs**

Convergent validity evidence in the form of correlations between the HDI and measures of self-esteem, anxiety, suicidal ideation, and hopelessness are presented in Table 5. As shown, the correlation between the HDI and the RSES was moderately high, *r*(105) = -.72. The negative relationship is due to the positive direction of item keying on the self-esteem measure. This is consistent with the nature of depression and feelings of low self-worth and detachment. The correlation between the HDI—SF and the RSES was -.80 with a correlation of -.67 found between the HDI—17 and RSES. The higher correlation coefficient found with the HDI—SF is due in part to the inclusion of the item of low self-worth on the HDI—SF, which becomes a significant proportion of the total score (1/9th of the scale) as compared to the HDI—17, which does not include an item specific to self-worth, and to the HDI, in which the self-worth item constitutes 1/23rd of the scale.

The relationship between the HDI and BAI was moderate, *r*(90) = .69, and lower than that found between the HDI and measures of depression. Similar relationships were found between the BAI and the HDI—17 and HDI—SF. The correlation coefficient between the HDI and BHS was .79, suggesting a moderate to strong relationship. It is not surprising to find the highest relationship between the BHS and the HDI—SF of *r*(91) = .83, given that a significant proportion of the HDI—SF content (2 of the 9 items) focuses on helplessness and hopelessness. The lowest relationship, *r*(91) = .76, was found with the HDI—17, which does not include content specific to hopelessness. The relationship between the HDI and the ASIQ was moderate, as expected, given that not all depressed persons are suicidal and not all persons thinking of suicide are depressed. A correlation...
between these two variables.

2 termination of r = .14. The correlations found between the HDI is of relatively low magnitude, with a coefficient of de-

The correlation coefficient of -.37 found with the MCSDS-SF. The correlation coefficient of -.37 found with the HDI forms and the MCSDS-SF suggest minimal relationship.

coefficients were found between the HDI forms and the MCSDS-SF. This is consistent with the procedure used for select-

Desirability has been operationalized as a response style variable as well as a social psychology construct of need for approval (Crowne, 1979). As shown in Table 5, low correlation coefficients were found between the HDI forms and the MCSDS-SF. The correlation coefficient of -.37 found with the HDI is of relatively low magnitude, with a coefficient of determination of $r^2 = .14$. The correlations found between the HDI forms and the MCSDS-SF suggest minimal relationship between these two variables.

**Contrasted Groups Validity**

A basic procedure for the examination of the validity of clinical measures is the determination of how well the measure differ-

**Table 4**

Means and Standard Deviations for the HDI-17 and HDRS for the Total Interview Sample by Diagnostic Groups, and Mean Differences and $t$ Tests Between Clinician Interview and Self-Report Forms

<table>
<thead>
<tr>
<th>Diagnostic group</th>
<th>Hamilton version</th>
<th>Controls ($n = 117$)</th>
<th>Anxiety disorder ($n = 77$)</th>
<th>Major depression ($n = 135$)</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDRS $M(SD)$</td>
<td>3.67 (3.11)</td>
<td>9.63 (5.11)</td>
<td>22.39 (3.69)</td>
<td>12.74 (9.22)</td>
<td></td>
</tr>
<tr>
<td>HDI-17 $M(SD)$</td>
<td>3.99 (2.99)</td>
<td>10.18 (5.91)</td>
<td>22.22 (5.12)</td>
<td>12.92 (9.35)</td>
<td></td>
</tr>
<tr>
<td>Mean difference</td>
<td>0.32</td>
<td>0.56</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>$t(df)$</td>
<td>1.79 (116)</td>
<td>1.78 (76)</td>
<td>.56 (134)</td>
<td>1.09 (328)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Analyses are based on those participants who completed both the Hamilton Depression Inventory (HDI) and Hamilton Depression Rating Scale (HDRS). Total sample for these analyses is 329.

tiates between diagnostic or contrasted groups (Wiggins, 1973), also known as criterion group validity (Edwards, 1970). In this manner, the contrasted group validity of a measure of depression may be evaluated by the examination of differences in scores between depressed and nondepressed persons. A more rigorous test of contrasted group validity for a measure of depression is the differentiation between persons with depression, nondepressed nonpsychiatric persons, and persons with other psychiatric disorders. The latter provides for the examination of the ability of a test to be specific to depression rather than generalized psychological distress or, in the case of comparison to persons with anxiety disorders, negative affectivity. This data may also be seen as a test of discriminant validity.

To examine the extent to which forms of the HDI differentiated between diagnostic groups, an ANOVA between the HDI scores of the three (nonpsychiatric controls, anxiety disorders, and major depression) groups was computed for each HDI form and is presented in Table 6. Prior to univariate analyses, a multivariate analysis of variance (MANOVA) of the three HDI forms between the three diagnostic groups was computed. Given that multiple options exist for statistical tests of MANOVA procedures with no best overall procedure (Marascuilo & Levin, 1983), we found that results of Wilks's lambda ($\Lambda = .222$) and Hotelling's trace criterion ($T = 3.491$) procedures were both significant, $F(6, 704) = 131.87, p < .0001$ and $F(6, 702) = 204.23, p < .0001$, respectively. Between-group differences were examined using Scheffe comparisons with an alpha level of $p < .01$. As shown in Table 6, relatively large differences were found between diagnostic groups, with univariate $F(2, 354)$ values ranging from 455 to 566 for the various HDI versions. The largest $F$ value was found for the HDI-SF. This is consistent with the procedure used for selecting items for inclusion on the scale. The Scheffe comparisons indicated that the major depression group scored significantly higher on the HDI, HDI-17, HDI-SF, and the HDRS than either of the other two groups. In nearly all instances, the mean score of the group of persons with a diagnosis of major depression was more than twice the score of the group of persons with anxiety disorders.

An examination of the differences between mean scores of...
Table 6
Hamilton Scale Means and Standard Deviations for the Total Sample and by Diagnostic Group and Comparisons Between Groups

<table>
<thead>
<tr>
<th>Hamilton scale</th>
<th>Total sample</th>
<th>Controls</th>
<th>Anxiety disorders</th>
<th>Major depression</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDI</td>
<td>17.85</td>
<td>5.08</td>
<td>14.57</td>
<td>30.93</td>
<td>508.98***</td>
</tr>
<tr>
<td>SD</td>
<td>12.94</td>
<td>3.61</td>
<td>8.32</td>
<td>7.13</td>
<td></td>
</tr>
<tr>
<td>HDI-17</td>
<td>13.05</td>
<td>4.04</td>
<td>10.94</td>
<td>22.13</td>
<td>455.00***</td>
</tr>
<tr>
<td>SD</td>
<td>9.17</td>
<td>2.89</td>
<td>6.22</td>
<td>5.10</td>
<td></td>
</tr>
<tr>
<td>HDI-SF</td>
<td>9.22</td>
<td>2.05</td>
<td>7.19</td>
<td>16.70</td>
<td>566.44***</td>
</tr>
<tr>
<td>SD</td>
<td>7.26</td>
<td>1.92</td>
<td>4.45</td>
<td>3.88</td>
<td></td>
</tr>
<tr>
<td>HDRS</td>
<td>12.74</td>
<td>3.67</td>
<td>9.63</td>
<td>22.39</td>
<td>757.61***</td>
</tr>
<tr>
<td>SD</td>
<td>9.22</td>
<td>3.11</td>
<td>5.11</td>
<td>3.69</td>
<td></td>
</tr>
</tbody>
</table>

Note. Total sample size is 357 for the Hamilton Depression Inventory (HDI), the 17-item version of the HDI (HDI-17), and the Hamilton Depression Inventory—Short Form (HDI-SF), and 329 for the Hamilton Depression Rating Scale (HDRS); sample size for the control group is 118 for the HDI, HDI-17, and HDI-SF; and 117 for the HDRS; sample size for the anxiety disorders group is 99 for the HDI, HDI-17, and HDI-SF; and 77 for the HDRS; sample size for the major depression group is 140 for the HDI, HDI-17, and HDI-SF; and 135 for the HDRS. The degrees of freedom associated with the F tests for differences between groups on the HDI, HDI-17, and HDI-SF are (2, 354) and (2, 326) for the F value associated with group differences on the HDRS. Scheffé post hoc comparisons computed with $p < .01$ between groups indicated significant differences in mean scores for all group comparisons on all of the Hamilton scales. $F$ tests are between diagnostic groups.

persons with major depression and nonpsychiatric controls suggests that the depression group had HDI scores well above those of the normal and anxiety disorders groups. The mean score of 30.93 on the HDI for the major depression group was approximately 2 SD above the HDI mean of the group with anxiety disorders ($M = 14.57$) and 4 SD above the nonreferred community sample mean of 5.08. These results provide strong evidence for contrasted group validity for the HDI scales. For the clinician-administered HDRS, an overall $F(2, 326)$ of 757.61, $p < .0001$, was found with significant differences (Scheffé comparisons, $p < .01$) between all group comparisons. Significant differences were found between diagnostic and nonpsychiatric control groups on all forms of the HDI and the HDRS.

In addition to examining total score differences between depressed and nondepressed adults, the diagnostic classification data allows for the investigation of differences in item endorsement between depressed and nondepressed adults. It was expected that relatively large differences would be found between persons with major depression and nonreferred community adults. As noted, a somewhat more rigorous test is the differentiation between these two groups, along with the group of adults with other psychiatric diagnoses. Research has shown a strong link between anxiety disorders and depressive disorders in adults that may account for similar responses to measures of depression. Furthermore, from a phenomenological perspective, there are a number of symptoms of these disorders that overlap. Thus, a test of validity is the power of the HDI items in discriminating between nonreferred adults, adults with major depression, and adults with anxiety disorders.

Significant differences ($p < .001$) were found between groups for all HDI items, with $F$ test values ranging from 9.85 to 425.61. Scheffé comparisons between groups were computed with alpha set at .01. Statistically significant differences between groups were found on all items, with depressed adults demonstrating higher scores than nonpsychiatric controls and persons with anxiety disorder. On six items (12, 15, 16, 17, 18, and 20), the mean score differences between the controls and persons with anxiety disorders were not significantly different at an alpha level of .01. Overall, HDI scales and HDI items demonstrated strong statistical and clinical significance as indication of contrasted groups validity in differentiating between depressed and nondepressed samples.

**HDI Factor Structure**

The results of factor analysis of the HDI should be viewed as descriptive of the general underlying structure of this measure with the study sample. Because the measurement of depression is more phenomenological than theoretical in orientation, the current analysis may be considered exploratory rather than confirmatory. The factor structure of the HDI was examined using the 23-item basic form of the HDI. A principal component analysis procedure was used, and orthogonal (Varimax; Kaiser, 1958) and oblique rotations were computed. Both analyses resulted in highly similar factor structures and loadings. The results (item factor loadings, eigenvalues, etc.) of the oblique factor solution are presented. Given the expectation that factors of depression would be somewhat related, along with the descriptive nature of the factor analysis, the oblique factor rotation is presented and provides for a reasonably parsimonious fit with...
the nature of the HDI content. The oblique factor rotation was computed with \( k = 0 \) as recommended by Harman (1976). A factor loading of .40 or greater for an item was viewed as meaningful for item-with-factor placement.

Four factors with eigenvalues lambda (\( \lambda \)) of 9.95, 1.54, 1.30, and 1.03 were obtained, accounting for 60.1% of the total variance. As was expected to be found with a homogeneous content measure such as the HDI, the majority of items loaded on the first unrotated factor, with factor loadings ranging from .30 to .89, with a median item loading of .63. A scree test (Cattell, 1966) for examining the meaningfulness of the factors suggests a strong first factor, consistent with the overall symptom specification of depression as a clinical disorder. The rotated four-factor oblique solution for the total sample is presented in Table 7.

The first (rotated) factor can be characterized as a depressed mood–demoralization dimension, with items reflecting cognitive and motivational symptoms of depression such as feelings of loss of interest and pleasure, worthlessness, hopelessness, helplessness, dysphoric mood, suicidal ideation, guilt, and fatigue. All items on this factor demonstrated strong factor loadings with minimal shared variance with the other three factors. The second factor clearly represents a dimension of sleep difficulty with four items specific to insomnia and hypersonnia. As shown in Table 7, the item specific to hypersonnia also loaded on the first factor. The third factor taps a somatic–vegetative component of depression as demonstrated by symptoms of weight loss and loss of appetite. This factor was clearly defined with little item overlap with other factors. Factor 4 consisted of five items that suggest an anxiety–disorientation factor. Items on this factor include those of loss of insight, hypocondriasis, somatic and psychic anxiety, and agitation. The item related to psychological aspects of anxiety (Item 10) also demonstrated a meaningful loading on Factor 1.

The factor score intercorrelations were computed to examine relationships among factors. Factor score correlations were Factor 1 with Factor 2 = .19; Factor 1 with Factor 3 = .31; Factor 1 with Factor 4 = .47; Factor 2 with Factor 3 = .09; Factor 2 with Factor 4 = .17; Factor 3 with Factor 4 = .26. The data suggest a low to moderate degree of interrelationship among the HDI factors. To some extent this can be seen in Table 7 by the low item-factor loadings with factors other than the primary item-factor placement.

**Discussion**

The assessment of depression, whether by self-report or clinical interview, is critical for the clinical evaluation of depression and for use in research designed to enhance our understanding of depression. The measures we use to identify individuals who demonstrate clinical levels of depression need to be psychometrically sound. Psychometric characteristics including internal consistency reliability, test–retest reliability, score equivalence with the clinical interview HDRS, content validity, construct validity, and contrasted groups validity are all psychometric features that are of particular importance and need to be examined when developing a measure of depression.

The characteristics of depression support the use of self-report assessment procedures. Depression is a disorder that includes many symptoms that are internal to the individual and are not readily observable (Reynolds, 1992). Cognitive symptoms of guilt, self-deprecation, suicidal ideation, hopelessness, helplessness, and feelings of worthlessness, are among the symptoms of depression that are subjective to the individual and are often obscure to others unless a formal evaluation is conducted. Likewise, some somatic symptoms such as insomnia, appetite loss and other problems may be difficult for others to observe or may be attributed to physical or other causes, particularly if a larger constellation of depressive symptoms are not observed or identified.

This investigation focused on the development and validation of a self-report, paper-and-pencil version of the interview form of the Hamilton Depression Rating Scale. In this investigation, scores on the HDI demonstrated a very close correspondence to scores on the HDRS in our sample of depressed, anxious, and nondepressed adults, with a mean difference of less than one point. Three versions of the HDI were developed, consisting of 23, 17, and 9 items. All versions of the HDI demonstrated high internal consistency and test–retest reliability, a necessary condition for the establishment of validity. The coefficient alpha internal consistency reliability coefficients suggest a high degree of item homogeneity and, in a manner, support the usefulness of the Hamilton as a severity measure of depression, given that items are summed to a total score representing the depth of depression. The results of the item-total scale correlations provided further support for the homogeneity of HDI content.

---

**Table 7**

Rotated (Oblique) Factor Loadings of HDI Items

<table>
<thead>
<tr>
<th>HDI item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Loss of interest</td>
<td>.858</td>
<td>.021</td>
<td>.048</td>
<td>.035</td>
</tr>
<tr>
<td>22 Hopelessness</td>
<td>.825</td>
<td>.094</td>
<td>.044</td>
<td>.070</td>
</tr>
<tr>
<td>21 Worthlessness</td>
<td>.823</td>
<td>.133</td>
<td>.002</td>
<td>.063</td>
</tr>
<tr>
<td>1 Dyphoric mood</td>
<td>.803</td>
<td>.113</td>
<td>.112</td>
<td>.059</td>
</tr>
<tr>
<td>23 Indecision</td>
<td>.793</td>
<td>.047</td>
<td>.002</td>
<td>.032</td>
</tr>
<tr>
<td>3 Suicidal ideation</td>
<td>.779</td>
<td>.023</td>
<td>.127</td>
<td>.214</td>
</tr>
<tr>
<td>19 Helplessness</td>
<td>.752</td>
<td>-.041</td>
<td>.073</td>
<td>.062</td>
</tr>
<tr>
<td>2 Guilt</td>
<td>.715</td>
<td>.086</td>
<td>-.181</td>
<td>.012</td>
</tr>
<tr>
<td>13 Fatigue</td>
<td>.614</td>
<td>.009</td>
<td>.058</td>
<td>.271</td>
</tr>
<tr>
<td>8 Retardation</td>
<td>.614</td>
<td>-.015</td>
<td>.112</td>
<td>.186</td>
</tr>
<tr>
<td>14 Libido</td>
<td>.593</td>
<td>-.150</td>
<td>-.161</td>
<td>.165</td>
</tr>
<tr>
<td>20 Detachment</td>
<td>.426</td>
<td>-.028</td>
<td>.327</td>
<td>.107</td>
</tr>
<tr>
<td>6 Insomnia (late)</td>
<td>.251</td>
<td>.619</td>
<td>.147</td>
<td>.066</td>
</tr>
<tr>
<td>18 Hypersonnia</td>
<td>.419</td>
<td>-.616</td>
<td>.255</td>
<td>.129</td>
</tr>
<tr>
<td>5 Insomnia (middle)</td>
<td>.227</td>
<td>.562</td>
<td>.057</td>
<td>.247</td>
</tr>
<tr>
<td>4 Insomnia (early)</td>
<td>.206</td>
<td>.436</td>
<td>.225</td>
<td>.188</td>
</tr>
<tr>
<td>17 Weight loss</td>
<td>-.094</td>
<td>.113</td>
<td>.864</td>
<td>-.061</td>
</tr>
<tr>
<td>12 Appetite loss</td>
<td>.085</td>
<td>-.010</td>
<td>.704</td>
<td>.094</td>
</tr>
<tr>
<td>16 Loss of insight</td>
<td>-.213</td>
<td>-.000</td>
<td>.133</td>
<td>.726</td>
</tr>
<tr>
<td>15 Hypocondriasis</td>
<td>.133</td>
<td>.036</td>
<td>-.147</td>
<td>.661</td>
</tr>
<tr>
<td>11 Anxiety–somatic</td>
<td>.317</td>
<td>-.007</td>
<td>.068</td>
<td>.581</td>
</tr>
<tr>
<td>9 Agitation</td>
<td>.207</td>
<td>.267</td>
<td>.025</td>
<td>.498</td>
</tr>
<tr>
<td>10 Anxiety–psychic</td>
<td>.413</td>
<td>.163</td>
<td>.072</td>
<td>.424</td>
</tr>
</tbody>
</table>

Eigenvalue: 9.95, Variance (%): 43.3
In addition to internal consistency reliability, the test-retest reliability of a severity measure of depression is important. The assessment of depression in adults typically focuses on an individual's level of distress and the potential need for intervention. It is important to know whether subsequent changes in an individual's score on a depression measure are due to the intervention, an artifact associated with repeated assessment, or other considerations (e.g., changes in stressors, etc.). The HDI was designed to evaluate depressive symptoms experienced over the previous 2 weeks. The 1-week test-retest reliability coefficients found in this investigation for the three forms of the HDI were high with minimal change in raw scores over this time period. Overall, the results of the test-retest reliability analyses are consistent across forms and suggest high test-retest reliability for the HDI as a measure of depression in adults. The determination of test-retest reliability is critical for a measure of depression, especially if such a measure is to be used in treatment outcome research or clinically to evaluate symptom change due to therapy. Mean score changes over time on the HDI forms were small and, in most cases, were roughly equivalent to about a tenth of a standard deviation. Mean score changes tended to be somewhat larger for men than women, although changes were small. The test-retest data supports the stability of the HDI and is particularly noteworthy for applications that require repeated assessments.

A high degree of correspondence was found between the HDI and the HDRS, providing evidence for validity of the HDI as a measure of depression and, in the case of the HDI-17, equivalence with the HDRS. Differences between the HDI-17 and the HDRS in a relatively large sample (n = 329) of persons with major depression, anxiety disorders, and nonpsychiatric community controls were minimal, with mean differences of less than one point. For the total interview sample the difference in means between the HDI-17 and HDRS was less than .20. These findings, along with the correlation of .95 between these two forms, provide evidence for the validity of the HDI, particularly given that the HDI is a brief 10-min paper-and-pencil measure that can be group or individually administered and the HDRS is a semistructured clinical interview that requires between 20 and 30 min to administer.

Evidence for the validity of the HDI was also examined from the perspectives of convergent and discriminant validity. The relationship between the HDI and measures of related constructs of anxiety, self-esteem, hopelessness, and suicidal ideation were moderately high, with absolute values of correlation coefficients ranging from .65 to .79. Differences in correlations between the various HDI forms and the measures of related constructs were found and are in part a function of the symptom content differences between forms. For example, the BAI demonstrated a correlation of .72 with the HDI-17 and .62 with the HDI-SF. The difference, although not tested for statistical significance, was most likely due to the inclusion of multiple anxiety-related symptoms on the HDI-17 (Items 10 and 11) and a general lack of anxiety-oriented content on the HDI-SF. Similarly, the higher correlation with self-esteem found with the HDI-SF (r = .80) compared to the HDI-17 (r = .67) may be a function of the inclusion of items related to self-worth on the HDI-SF but not included on the HDI-17. The difference between these two forms in their relationship with the BHS reflects the inclusion of items dealing with hopelessness and helplessness on the HDI-SF that are not present on the HDI-17. Thus, the pattern of correlations found between the HDI forms and the measures of related constructs is consistent with the formulation of the HDI and provides evidence for the convergent validity of the various HDI forms. In addition, the HDI also demonstrated a high correlation with the BDI, with similar correlations found across HDI forms.

As a method for establishing validity, contrasted groups analysis has been used for many years in personality assessment. Significant differences in scores between diagnostic and control groups were found on the self-report and clinician administered versions of the Hamilton. Although the F value was somewhat larger for the clinician version, all versions demonstrated large differences between groups of depressed and nondepressed participants.

Construction of the HDI was based in large part on the content formulated by Hamilton (1960) for the HDRS with items added to enhance content validity with symptoms of depression delineated by DSM-III-R. Thus, items were developed on the basis of their congruence with specified clinical symptomatology rather than hypothesized factors of depression. In this manner the factor analysis of the HDI items provides insight as to the underlying factor structure of the scale and allows for post hoc examination of item relationships. Validity information may be inferred from the finding of factors that are reasonably descriptive of clinical aspects of depression. Results of the factor analysis suggest that the HDI measures realistic domains of depression. These descriptive components of depression include dysphoric mood-demoralization, somatic-vegetative, and anxiety-disorientation dimensions and are presented here as descriptive of the underlying dimensions of the HDI, and not as discrete subscales of depression. The resultant factors represent relatively parsimonious dimensions or domains of depressive symptomatology and may be of interest to researchers and clinicians.

The results of this investigation provide strong support for the reliability and validity of the HDI as a measure of the severity of depression and as a self-report version of the HDRS clinical interview. The HDI was developed to provide a paper-and-pencil measure of depression consistent with that obtained from the administration of the HDRS, with the added goals of increasing the coverage of symptoms of depression as formulated by contemporary systems of classification. The development of a short form of the HDI with psychometric characteristics roughly equivalent to the full-scale version provides researchers and clinicians with a brief, reliable, and valid screening measure. The HDI is noteworthy in its design and evaluation of symptoms of depression. By inquiring into multiple components of individual symptoms the HDI, as compared to other self-report measures of depression, allows for greater fidelity of symptom assessment. In this manner, the respondent is able to more accurately describe the level, intensity, and characteristics of specific symptoms of depression, providing clinicians and researchers with a more accurate and valid presentation of their affective status.
References


Received July 14, 1994
Revision received March 14, 1995
Accepted March 28, 1995

Call for Nominations

The Publications and Communications Board has opened nominations for the editorships of the Journal of Experimental Psychology: Animal Behavior Processes, the “Personality Processes and Individual Differences” section of the Journal of Personality and Social Psychology, the Journal of Family Psychology, Psychological Assessment, and Psychology and Aging for the years 1998–2003. Stewart H. Hulse, PhD; Russell G. Geen, PhD; Ronald F. Levant, EdD; James N. Butcher, PhD; and Timothy A. Salthouse, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 1997 to prepare for issues published in 1998. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees.

To nominate candidates, prepare a statement of one page or less in support of each candidate and send to the attention of the chair of the appropriate search committee. Search committee chairs are

- Joe L. Martinez, PhD, for JEP: Animal Behavior Processes. Members of the search committee are Russell M. Church, PhD; Michael Domjan, PhD; Michael S. Fanselow, PhD; and William D. Timberlake, PhD.
- David L. Rosenhan, PhD, for the “Personality Processes and Individual Differences” section of the Journal of Personality and Social Psychology. Members of the search committee are Nancy E. Cantor, PhD; Susan Fiske, PhD; Oliver John, PhD; and Mark Snyder, PhD.
- Carl E. Thoresen, PhD, for the Journal of Family Psychology. Members of the search committee are Arthur M. Bodin, PhD; James H. Bray, PhD; Lucia Gilbert, PhD; John M. Gottman, PhD; and Howard A. Liddle, EdD.
- Hans H. Strupp, PhD, for Psychological Assessment. Members of the search committee are Lee Anna Clark, PhD; Ken Pope, PhD; M. Tracie Shea, PhD; and Auke Tellegen, PhD.
- Lyle E. Bourne, PhD, for Psychology and Aging. Members of the search committee are Carol Barnes, PhD; Alfred Kasznia, PhD; M. Powell Lawton, PhD; and Michael A. Smyer, PhD.

Address all nominations to the appropriate search committee at the following address:

Lee Cron
P&C Board Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

First review of nominations will begin December 11, 1995.