David J. Berndt, PhD

Inventories and Scales for Depression

Depressive Disorders

Inventories and Scales for Depression

David J. Berndt, PhD

e-Book 2015 International Psychotherapy Institute

From *Depressive Disorders* edited by Benjamin Wolberg & George Stricker

Table of Contents

Inventories and Scales for Depression

The assessment of depression necessarily involves many levels of focus. An appropriate diagnosis, for example, may draw upon psychopharmacological evaluation, social and family history, structured diagnostic interviews, psychiatric and/or behavioral ratings, and self-report inventories and questionnaires. Since no one has yet isolated an effective phenotypical marker for depression, a combination of several of the above approaches is most appropriate, if diagnosis of major depressive disorder is the goal. However, when lack of resources and time forces clinicians and researchers alike to use only one of the above, inventories and scales, either self-reported or rated by others, may be relied upon with too little caution. These instruments, however, are useful when appropriately evaluated. This chapter discusses several of these instruments, with a focus on their reliability and validity.

Depression inventories and scales are most appropriately used for the assessment of severity and frequency of depressive symptoms and features. The analysis of symptom patterns is another frequent use. Because depressed individuals are relatively more accurate at self-reporting their states than are those with other psychiatric diagnoses (e.g., schizophrenia, behavior disorders), the usefulness of these measures is apparent.

The emergence of structured diagnostic interviews and more reliable

diagnoses from objective standards such as Spitzer, Endicott, and Robins' (1978) Research Diagnostic Criteria (RDC) and DSM-III and DSM-IV (American Psychiatric Association, 1980, 1988) has had an immense impact on American psychiatry and psychology. No longer constrained by low agreement (reliability) between clinicians on diagnosis, researchers and clinicians alike are less hesitant to believe the accuracy of their diagnosis. A consequence of this new confidence is for a similar optimism to lead some researchers to use self-report or observer ratings of depression to actually identify cases. Research on case identification is minimal, and skepticism is necessary until researchers demonstrate that the diagnoses are accurate and valid. Until such research develops sufficient data, scales should be supplemented whenever possible by (minimally) a structured diagnostic interview.

This chapter will briefly discuss the history of self-report and observer rating scales. The four most widely used measures will be evaluated in considerable depth. Several newer instruments will then be more briefly discussed, many of which are still evolving and have yet to be comprehensively tested by the research community, despite their frequent use.

A BRIEF HISTORY OF INVENTORIES AND SCALES

A casual reader of today's literature might well conclude that only a few instruments, such as the Beck (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) Depression Inventory (BDI), are used in the assessment of depression. Four instruments, including the BDI, are currently the most widely used inventories and scales for the assessment of depression; they will be discussed in depth in the next section. However, these instruments emerged as the profession's standards within a historical context. This section briefly reviews the evolution of inventories and scales for the measurement of depression.

Jasper (1930) was the first to propose an instrument, the Depression-Elation Scale (D-E), that attempted to measure depression in a self-report format. Jasper envisioned his scale as tapping a general depression factor (Spearman, 1904). He assumed that not only depression-elation, but optimism-pessimism, and enthusiasm-apathy were all subsumed under a general single dimension. His 40-item trait measure of depression included 20 nonpersonal items focused on such topics as pessimism or apathy for sociopolitical institutions. The other 20 items resembled more closely today's subjective self-report items. All items were rated on a five-point scale and took into account how difficult it was to answer the question.

Chant and Myers (1936) were the first to employ Thurstone-type scaling to the assessment of depression. This instrument contained 22 items with values ranging from .3 for "I wish I had never been born" to 10.7 for "Life could not be better for me" (Chant & Myers, 1936, p. 35). The score was the average score of all items checked "yes."

Guilford and Guilford (1939) developed the third published scale that attempted to isolate depression from other constructs. The Guilfords developed their 17-item factor-D as part of an early study of introversion-extroversion, and were more concerned with personality than with psychopathology. The scale is remarkable in that it was the first instrument that benefited from the emerging reliance upon factor analysis to evaluate scales.

World War II brought psychology out of its academic confines and thrust it prominently into a clinical partnership with psychiatry, especially in the area of assessment. In depression studies, the self-report literature was dominated for two decades by Hathaway and McKinley's (1942) MMPI-D scale. The 60-item scale remains one of the most widely used measures, and will be evaluated in the next section.

Between the 1942 publication of the MMPI-D scale and 1967, the four most widely used measures gained their foothold. In 1961 a scale was

published by a psychoanalytic psychiatrist (Beck et al., 1961), which went on to become the most widely used self-report measure of depression. In addition to its early arrival on the scene, low-cost distribution with no formal publisher, and psychometric improvement over the MMPI-D, the BDI's popularity was clearly tied to the fame of its author, who became a leading cognitive-behavioral theorist.

In 1965 the Zung Self-rating Depression Scale (SDS; Zung, 1965) gained a strong foothold, especially among physicians, who received free copies of the SDS from a manufacturer of antidepressant medication. The primary alternatives to these three measures have been observer-rating scales, the most widely used of which is Hamilton's (1967).

In the following decades, other inventories and scales evolved. Lubin (1965) developed an adjective checklist that permitted a quick state measure with several parallel forms. Wessman and Ricks (1966) developed an instrument specifically for assessment of repeated measures. Their scale was the first to provide the advantage of measuring 16 different affects. The measures, however, were theoretically derived and have not been properly evaluated psychometrically.

Costello and Comrey (1967) were the first researchers to develop a depression scale that specifically reduces the role of anxiety. A very thorough

research process resulted in two orthogonal measures of trait anxiety and depression. Other scales developed in the late 1960s included a psychodynamically oriented scale by Leckie and Withers (1967), and another attempt at multiple symptom assessment by Hunt, Singer, and Cobb (1967). Internal consistency for the measure by Hunt et al. was unacceptably poor for the individual symptoms. Popoff's (1969) brief test was aimed at measuring depression in medical populations, with several "covert" items derived to detect somatization of depression.

Two of the more innovative measures of depression were developed in the late 1960s and 1970s. The Visual Analogue Scale (Aitken, 1969) is a simple and useful state measure of depression. In its modified form it is simply a 100-mm line. Respondents are asked to indicate their current mood by placing a mark on the line between the anchors "worst" and "best." Another measure, developed by Cohen and Rau (1972), requires respondents to select faces that represent how they feel. Both measures share the advantages of requiring minimal literacy and possibly offering less cultural bias than most other instruments.

Surprisingly few inventories have emerged in the 1970s and 1980s. The 40-item Institute for Personality and Ability Testing (IPAT) Depression Scale was developed by Krug and Laughlin (1976). The most recently developed observer-rating scale, by Montgomery and Asberg (1979), has received little

attention outside of Britain. The Wakefield Scale (Snaith, Ahmed, Mehta, & Hamilton, 1971) evolved into the Leeds Scale (Snaith, Bridge, & Hamilton, 1976), which is, again, used primarily by the Europeans.

Several instruments that emerged from this period have, however, begun to receive wider attention. Perhaps the best known is Radloff's (1977) Center for Epidemiological Studies Depression Scale (CES-D). Its 20 items were a composite of items from other scales. In part because of its use in a national epidemiological study, Radloff's instrument rapidly captured the interest of researchers.

The Carroll Rating Scale for Depression (Carroll, Fielding, & Blashki, 1973) is a self-report measure that translates the Hamilton (1967) Rating Scale for Depression (HRSD) to self-report format. It is widely used in drug studies, and by those who share Hamilton's concern that the Beck and other self-report instruments focus too closely on mood and ignore somatic components.

The Multiscore Depression Inventory (Berndt, Petzel, & Berndt, 1980) was the first depression inventory to develop reliable subscales for 10 features of depression. It is also available in a brief 47-item format that measures nine of the 10 symptoms and features. A comparable children's form is also available (Berndt & Kaiser, in press).

The Depressive Experiences Questionnaire (Blatt, D'Afflitti, & Quinlan, 1976) is derived from psychodynamic literature, and is popular among researchers who are interested in Blatt's et al. (1976) concepts of anaclitic and introjective depression.

EVALUATION OF FOUR WIDELY USED SCALES

Four inventories and scales are clearly the most widely used by most researchers and clinicians. As noted earlier, the Beck Depression Inventory (BDI) (Beck et al., 1961) is used by psychology researchers almost to the exclusion of even the other three major instruments (Berndt et al., 1980). Because these four instruments are so widely used, they will be discussed and evaluated in depth. In the next section, four promising instruments that are emerging as important contributions to assessment of depression will be more briefly evaluated.

The Beck Depression Inventory (BDI)

The most widely used self-report measure of depression is the Beck Depression Inventory. Although initially designed to be administered by trained interviewers, the 21-item scale with a four-alternative multiple-choice format quickly circulated in self-report format throughout academic and clinical settings. Despite its widespread use, the BDI was revised in 1971 and copyrighted in 1978 (Beck, Rush, Shaw, & Emery, 1979). It may be purchased from test publishers and is available for computerized administration and scoring.

The BDI reflects nearly 30 years of research and well over 1,000 studies. Few comprehensive reviews of its psychometric properties have been

published outside of Beck's own group. The most thorough review to date is a paper by Beck, Steer, and Garbin (1988).

Description

Most authors fail to distinguish between the 1961 (Beck et al., 1961) and 1978 (Beck et al., 1979) copyrighted versions of the BDI. Beck et al. (1988) point out that only 6 of the 21 items remain unchanged. Beck and Steer (1984) analyzed internal consistency and item properties of the original and revised versions, and concluded that they were similar, and Lightfoot and Oliver (1985) found high correlation between the two instruments in college students (r=.94), although the 1961 version had slightly lower mean scores than the 1978 version. Future users of the BDI should clearly indicate which of the two versions they use.

A short (13-item) version of the BDI (Beck & Beck, 1972) has also been used. Although the forms correlate well, little research has been done on the short form, and the factor structure appears significantly different (Berndt, 1979; Gould, 1982). Reliability appears adequate from the studies cited, but the short form must be used with caution at this stage, until it has received more comprehensive study.

Reliability

One reason the BDI gained its foothold across the country is that it was the first self-report measure with good reliability. Internal consistency ranges from a low of .73 (Gallagher, Nies, & Thompson, 1982) in a nonclinical population to .95 in two studies (Coleman & Miller, 1975; Steer, McElroy, & Beck, 1982). Internal consistency appears adequate for both clinical and nonclinical samples, with the psychiatric samples generally midway between .80 and .90 and the nonpsychiatric samples at a comparable reading.

Test-retest reliability is complicated to evaluate because the 1961 form's instructions were phrased more as a state measure and the 1978 version's instructions implied desire for a traitlike response. Because few researchers differentiate the versions, it is not a simple task to evaluate. Two of the lowest test-retest reliabilities were published with the state instructions. Bailey and Coppen (1976) reported r = .65 over one week, and May, Urquhart, and Tarran (1969) reported a three-week reliability of .48. Both of these studies used psychiatric samples, which may have contributed more variability.

In nonpsychiatric samples, the highest test-retest reliability reported was by Lightfoot and Oliver (1985) over two weeks (.90). More typical coefficients were in the .60s and .70s, still appropriate for periods from one week to four months.

Validity

Because so much research has been done with the BDI, a tremendous amount of validity data exist. Concurrent validity correlation coefficients have been, on average, in the .70s for the HRSD (Davies, Burrows, & Poyton, 1975), although the range is from a low of .41 (Carroll et al., 1973) to a high of .86 (Steer, Beck, & Garrison, 1986). Among self-report measures, the worst concurrent validity is with the MMPI-D scale (at least nine studies below .65), although these results may reflect more upon psychometric properties with the MMPI-D scale than the BDI. Correlations with the Zung have tended to be adequate (Blatt, Quinlan, Chevron, McDonald, & Zuroff, 1982), with the exception of some studies with nonclinical samples. Lubin's (1965) Depressive Adjective Checklist (DACL) is most frequently used in nonclinical samples, where correlations are routinely below .70; however, in a clinical sample (Byerly & Carlson, 1982) the correlation was .73. Correlation with the Multiscore Depression Inventory (Berndt, Berndt, & Byars, 1983) was .76 in a medical sample.

Construct validity is extensive for the BDI, and relevant variables from REM sleep difficulty (Akiskal, Lemmi, Yerevanian, King, & Belluomini, 1982) to suicidal behaviors (Emery, Steer, & Beck, 1981) have demonstrated its usefulness. Factorial structure has been studied in at least 20 studies (see review in Beck et al., 1988).

Final Comments

Several problems with the BDI must be kept in mind. First, in nonclinical populations there is tremendous skew, making sensitivity to subtle changes in mood hard to measure. The Carroll et al. (1973) concerns, that self-report measures like the BDI do not go beyond mood, are still noteworthy, and it is best to combine the BDI with either a structured diagnostic interview or an observer rating. The BDI warrants use for research and clinical purposes as a measure of severity of depressed mood, or as a screening measure to detect possible cases of depression.

MMPI-D

The MMPI-D (depression) scale (Hathaway & McKinley, 1942) is the "grandaddy" of all depression self-report measures, and still is used frequently, either as part of the entire inventory or as a 60-item true-false scale. Hindsight permits easy criticism of the MMPI-D scale for its psychometric failures, but readers should remember that at the time it was developed it was far superior to any similar measure, and most of today's psychometric concerns evolved out of research that attempted to improve the MMPI and its scales. The publishers are revising the MMPI; the long-awaited revision is imminent. Perhaps some of the criticisms raised here will be remedied in the version for the 1990s.

Reliability and Cohesiveness

One troublesome problem for researchers is the lack of homogeneity of the scale. Internal consistency reliabilities are below acceptable standards. Test-retest reliability is comparably low. The problem arises from the lack of a cohesive construct. Factor-analytic studies (Comrey, 1957; O'Conner, Stefic, & Gresock, 1957) have demonstrated that the scale is factorially complex. For example, among 60 items, Comrey found nine factors, and the one appearing to measure depression consisted of only five items that loaded higher than .30. More recently, construct validity of the original scale was disputed on the grounds that it measures personality rather than illness (Snaith et al., 1971).

Several authors attempted to improve on the original scale by refinement. McCall (1958) found 26 items from the original 60 which he determined to be face valid: The 26 items did discriminate better than the other items between a depressed and nondepressed clinical sample. Similarly, Dempsey (1964) developed a short version of the MMPI-D and isolated a single dimension. The item scale had better internal consistency, but it shared many items with Comrey's (1957) largest factor, described as measuring neuroticism.

Cantor's (1960) attempt to develop a short form appears to have been more widely accepted because of adequate internal consistency and some evidence of concurrent validity. Stein (1968) used the entire MMPI item pool

and derived clusters, one of which was labeled depression and apathy (versus positive and optimistic outlook), although the cluster shared only 10 items with the original scale. Rosen's (1962) Depression Reaction Scale was also developed empirically, by choosing items that discriminated a group of neurotic depressives from other psychiatric patients. The 42 items shared only four items with the MMPI-D. The most true to the original scale was the revision by Harris and Lingoes (1955), and the depression measure computed by them is now used frequently.

Not one of these scales, however, was developed with the benefit of the increased reliability and validity of diagnosis gained from structured diagnostic interviews and objective criteria such as DSM-III and DSM-IV (American Psychiatric Association, 1980, 1988). The next revision, it is hoped, will use modern methods to isolate items that are psychometrically sound.

Validity

Lacking construct validity, the complex MMPI-D scale has been difficult to validate, although among hundreds of publications many studies do support its usefulness. Typical of the critical studies, however, is one by McNair (1974), which demonstrated that the MMPI-D is less sensitive to drug intervention than other scales are. Correlations with concurrent measures demonstrate that the 60-item version is clearly below today's psychometric

standards (Beck et al., 1988; Blatt et al., 1982).

Most of the credible evidence for validity comes from the use of the scale in combination with other scales on the MMPI. Profile analyses are widely accepted among users of the MMPI. Dahlstrom, Welsh, and Dahlstrom (1972) concluded that none of the revisions was superior to the original scale, and most researchers who use it rely primarily on profile analyses.

Summary and Evaluation

The MMPI-D scale is still widely used, despite its need for modernization. The forthcoming MMPI-II (Hathaway & McKinley, 1989) will probably remedy some of the problems. Clinicians and researchers continue to use it because of its ease of administration (especially computerized versions) and the continued popularity of the original MMPI in its entirety. There is little support for the use of the 60-item scale, although it may still be useful in profile analyses.

One very major concern is the poor normative basis for computation of scores. While some publishers provide up-to-date normative data, the Minnesota group awaits the revision and continues to rely on old norms not appropriately general or stratified by age or geography.[1] This is most problematic with the use of the MMPI for adolescents, who, when compared with the 1940s normal adults, appear depressed. This problem disappears for

published versions that use good adolescent norms.

Computer users especially should be careful to see that clients are evaluated with the appropriate age norms; the failure to do so may well violate ethics of psychology and other responsible professions. Computer versions of the MMPI are useful if supported by sound clinical interpretation by a qualified psychologist. A good discussion of several computerized versions of the MMPI is in a book by Lachar (1974).

Zung Self-Rating Depression Scale

Zung's (1965) Self-Rating Depression Scale (SDS) has been one of the most widely used measuring instruments, especially in psychiatric and medical settings. Its ease of use and of scoring is its most attractive feature. Twenty items require respondents to rate themselves on four-point, Likert-type scales, anchored by the extremes of "none or little of the time" and "most or all of the time." The scale is balanced: half the items are symptomatically negative, the other half are positive. It is particularly easy to score with a plastic overlay, and it is available also in interviewer rating scale form (Zung, 1972).

The SDS, like all well-known instruments, is not without its detractors. Because the items were taken verbatim from interviews of psychiatric patients, the wording is considered objectionable by some nonpsychiatric

patients (Froese, Vasquez, Cassem, & Hackett, 1974). Hamilton (1972) criticized the scale for not including items on hypochondriasis, guilt, and retardation; Hamilton also believed the suicide question was poorly written. Another problem is that the anchor points, representing frequency of occurrence, result in mild, persistent symptoms counting more than severe, infrequent symptoms (Carroll et al., 1973). For chronic patients, the requirement to compare their present state with previous conditions presents difficulties.

Reliability

Surprisingly little research on reliability has been reported for the Zung. Knight, Waal-Manning, and Spears (1983) report an alpha coefficient of .79, which is modest but adequate. Evaluation of test-retest reliability remains a question for future researchers.

Validity

While there is a paucity of information on reliability, its widespread use provides considerable information on its validity. McNair (1974) pointed out that it is one of the more sensitive instruments in studies evaluating drug interventions. Adequate-to-good validity coefficients are obtained with the HRSD, the BDI, and the Multiscore Depression Inventory (Berndt, 1986;

Brown & Zung, 1972; Turner & Romano, 1984; Zung, 1965).

The SDS appears to discriminate depressed patients from nondepressed patients and normal respondents (Zung, 1965). With outpatients or less severely depressed patients, its ability to discriminate severity of depression seems less clear. Although Biggs, Wylie, and Ziegler (1978) reported different scores by level of rated depression in outpatients, Carroll et al. (1973) found the SDS inadequate for severity ratings.

Summary and Evaluation

Zung's SDS, with its ease of use and established validity with inpatients, will continue to be used, especially in inpatient studies on effectiveness of interventions. More reliability data, especially test-retest, are needed, and, because of its extreme wording and still unproven validity in less severely depressed settings, the SDS must be used with caution in non-inpatient or nonpsychiatric settings. As a screening instrument it may have some value, but recommended cutoff scores should not be substituted for diagnosis.

The Hamilton Rating Scale for Depression (HRSD)

Although other rating scales of depression are preferred, the first (Hamilton, 1960) remains the most widely used today. The HRSD can be administered independently, or extracted from the structured SADS

(Schedule for Affective Disorders and Schizophrenia) interview (Endicott & Spitzer, 1978).

The original scale involved 21 items, only 17 of which were to be counted in a total score, following a brief interview (Hamilton, 1960). The current instrument has evolved from two modifications of the original HRSD. Anchor points for each of the items were added. Of the original 17 items, nine are five-point scales (0-4) and eight are three-point scales (0-2). In addition, three items assessing helplessness, hopelessness, and worthlessness were added for the NIMH Treatment of Depression Collaborative Research Project (Elkin, Parloff, Hadley, & Autry, 1985).

Rating scales such as the HRSD have certain advantages over self-report measures. Foremost is the ability to rate noncognitive measures such as retardation. Another significant asset is that lack of literacy or severe disorganization of the patient does not disrupt the assessment.

Reliability

The most thorough evaluation of the psychometric properties of the HRSD may be found in a review by Hedlund and Vieweg (1979). Sparse information on internal consistency exists, and with the authors' intent to assess a heterogeneous set of symptoms, low internal consistency is not surprising. Schwab, Bialow, and Holzer (1967) reported item-total

correlations ranging from .45 to .78; however, Bech,' Bolwig, Kramp, and Refaelsen (1979) reported item correlation with a median of .47 and a range from –.02 to .81. Some of the items may detract from the usefulness of the HRSD as a global measure of severity, and this warrants further research.

More important, however, is interrater reliability. Published materials by the author leave considerable ambiguity for rater training, and it is surprising that interrater reliability has been remarkably good. Hedlund and Vieweg's (1979) review found nine studies with interrater reliability of .84 or above, and only one with a reliability inadequately low. O'Hara and Rehm (1983) found that acceptable reliability (r = .76) could be achieved with undergraduates who had been trained for only five hours. Most researchers employing the SADS have both skill and training, so the version extracted from the structured interview is likely to be reliable. Evidence from Endicott, Cohen, Nee, Fleiss, & Sarantakos, (1981) indicated such interrater reliability may be in the low .90s.

Validity

The review by Hedlund and Vieweg (1979) provided the best discussion. The scale demonstrated adequate concurrent validity, whether with clinicians' ratings, the BDI, the SDS, or the Multiscore Depression Inventory (Berndt, 1986; Hedlund & Vieweg, 1979).

The HRSD has been widely used to assess the impact of therapeutic interventions, whether psychopharmacological or psychotherapeutic, as in the NIMH Collaborative Research Project. One problematic tendency is to use a specified criterion for defining a case, most typically a cutoff score of 17 for inpatients (see Endicott et al., 1981) or 14 for outpatients (Sotsky & Glass, 1983). When such a criterion is combined with a structured diagnostic interview and preferably other criteria as well, then such a cutoff may be useful. Even in these circumstances, further research must clarify the optimum level for specificity and sensitivity.

Summary and Evaluation

The Hamilton (1960, 1967) Rating Scale for Depression is the most widely used rating scale for depression. It has impressive interrater reliability and considerable evidence of validity, but further research on reliability (internal consistency and test-retest) may improve the scale. With further psychometric study, this long-time standard might be improved by enhancement of internal consistency, if validity is not sacrificed.

Used with self-report and/or interviews like the SADS, the HRSD makes an excellent addition to an assessment battery when identifying "cases." As a measure of severity of depression it also seems an excellent choice as a complement to a more mood-oriented self-report measure.

OTHER INVENTORIES AND SCALES

While the BDI, SDS, HRSD, and MMPI-D are the best known instruments, several others have received increasing attention in recent years and fill needs not met directly by these four instruments. This section focuses on instruments that have made important contributions to the literature and show promise of having greater impact in ensuing years.

CES-D

Radloff's (1977) Center for Epidemiological Studies Depression Scale may be the best-normed trait measure of self-reported depression. It was used in a large national epidemiological study and while the results are not available to the public, the eventual publication of complete norms, stratified by a variety of demographic variables, will increase the usefulness of the scale. A 20-item composite of several other scales, the CES-D is brief to administer; individuals are asked to respond to a scale (0-3) on the basis of frequency of occurrence during the prior week, with 0 for no days, and 3 for five to seven days.

Internal consistency reliability of the CES-D is good. For patient groups, coefficient alpha and Spearman-Brown coefficients were .90 and .92; for normal respondents, they were in the mid .80s. Test-retest reliability after six months was an adequate .54 (Radloff, 1977). Correlations with concurrent

measures are favorable, ranging from .70 with Lubin's DACL (1965) to .81 with the BDI and .90 with the Zung SDS (1965) in recovered depressed patients (Weissman, Prusoff, & Newberry, 1975).

The biggest problem with self-report measures is reliance on cutoff scores from one instrument. The author recommends a cutoff of 16, and while this has sufficed for some research with adults, it has been problematic, especially with false negatives, for adolescents (Boyd, Weissman, Thompson, & Myers, 1982; Lewinsohn & Teri, 1982). Lack of specificity and sensitivity may be problematic for researchers, but clinicians tempted to use cutoff scores with this or other self-report measures should consider the consequences of missing 34-36 percent of depressed patients, or overdiagnosing the disorder by a 2:1 ratio.

In fairness, this problem exists throughout assessment, where a simple cutoff score facilitates research design or provides easy "rules of thumb" for students. Even in the "hard" sciences, sensitivity and specificity are not better with the dexamethasone suppression test. As mentioned at the beginning of the chapter, a combination of two or three levels of assessment is preferable.

The Multiscore Depression Inventory

Only one of the current depression scales in wide use provides reliable scores for a variety of subscales assessing "symptoms and features" of

depression. The Multiscore Depression Inventory (MDI; Berndt, 1986; Berndt et al., 1980) is too recently published to have been held to the close scrutiny of many of the instruments reviewed, but the initial evidence is encouraging. Lanyon (1984) described the MDI as being developed with "extensive care," and with potential for good validity.

The MDI measures a global "severity" of depression in a trait format and provides 10 nonoverlapping subscales that measure guilt, lack of energy, irritability, learned and instrumental helplessness, sad mood, pessimism, low self-esteem, social introversion, and cognitive difficulty. A brief (47-item) version (Berndt, Petzel, & Kaiser, 1983) takes only 10 minutes and provides reliable scores on nine of the ten subscales. The full-scale MDI consists of 118 items with a yes/no forced-choice format. It is easily scored using a plastic overlay, and a computerized testing report is also available from the publisher.

Internal consistency of the global scale is an impressive .97 with a medical outpatient sample (Berndt, Berndt, & Byars, 1983). The same study reported subscale reliabilities in the high .80s and low .90s for most of the 12-item subscales. The briefer scale (Berndt, Berndt, & Byars, 1983) sacrifices very little internal consistency. Test-retest reliability for this instrument is the highest of all measures surveyed (Berndt & Kaiser, 1980) and most of the subscales also remain stable over a 3-4 week period, with the exception of

instrumental helplessness.

Concurrent validity with the BDI, SDS, DACL, Popoff, and Zung are all above .70 (Berndt, 1986; Berndt, Petzel, & Kaiser, 1983). Some initial validity evidence supports the MDI's usefulness with clinical subgroups, but much of the research using profile analyses, such as the study comparing inpatient depressed and conduct-disordered patients (Berndt & Zinn, 1982), awaits replication by independent researchers.

The MDI demonstrates a consistent factor structure, and both factor and cluster analyses support the validity of the separate subscales (Berndt, 1981). The short form's factor structure is consistent with results from the long form, which led the authors to eliminate one subscale that did not hold up under analysis (Berndt, Berndt, & Kaiser, 1984).

Normative and validity data for adolescents are available for the MDI (Berndt, 1986), but the children's version (Berndt & Kaiser, in press) may prove more useful with younger patients. A unique aspect of the children's adaptation is the use by Berndt and Kaiser of children to generate the items, which were subsequently refined empirically. Normative data are available for ages eight through eighteen. Computerized reports can be obtained from the test publisher, but as always should be interpreted by a trained psychologist.

Depression Adjective Checklist

Lubin's (1965) Depression Adjective Checklist (DACL), the first state measure of depression, is useful in research projects looking for transient mood changes. The checklist format takes only 2-3 minutes for 32-34 adjectives, and the adjectives have high face validity. The multiple forms permit repeated assessments without the contamination of using the same instrument more than once.

Internal consistency is quite good, from .80 to .93 (Lubin, 1965); test-retest reliability over short intervals is similarly good. While concurrent validity evidence is not high, the instrument is different from most other measures in its focus on mood and its purely state format. It is also the least confounded with social desirability, of all the depression scales discussed (Christenfeld, Lubin, & Satin, 1978).

Carroll Rating Scale for Depression

Another measure (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981) that gained quick recognition is the Carroll Rating Scale for Depression (CRS). The CRS follows the more heterogeneous item content of the HRSD. As such, the CRS attempts to go beyond depressive mood items, and more somatic items are included.

The 52-item scale has a yes/no forced-choice format. While the correspondence between the two measures is not one for one, 13 of the 17 CRS items correlated more strongly with their item counterpart on the HRSD than with other items. Also in common with the HRSD is some problem with heterogeneity. Although initial reliability (split-half) was .87, the individual item-total correlations ranged from .05 to .78. Again, the new scale will inevitably be refined and either removing or revising some items will improve the scale.

Concurrent validity data indicate that the CRS and HRSD correlate .71 to .80, and .76 with the BDI (Carroll et al., 1981; Feinberg, Carroll, Smouse, & Rawson, 1981). There is some question whether the CRS is comparable with the HRSD for assessment of severity with moderate to high severity depression (Feinberg et al., 1981).

Factor-analytic and construct validation of the CRS await further research. Nevertheless, because of Carroll's deserved reputation in the psychiatric research community, the scale will see continued use in the future.

SUMMARY

Inventories and scales for the assessment of depression are useful tools for researchers and clinicians who use them with appropriate awareness of their strengths and weaknesses.

Although the Beck Depression Inventory is the most popular, several other instruments have usefulness for different purposes, and no one instrument is particularly useful in isolation. In combination with the structured diagnostic interviews developed in the past two decades and objective diagnostic criteria, these instruments can be useful as measures of severity, as screening devices, or for assessment of the course and symptomatology of depression.

REFERENCES

Aitken, R. C. B. (1969). Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine, 62,* 989-993.

Akiskal, H. S., Lemmi, H., Yerevanian, B., King, D., & Belluomini, J. (1982). The utility of the REM latency in psychiatric diagnosis: A study of 81 depressed outpatients. *Psychiatry Research, 7,* 101-110.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1988). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bailey, J., & Coppen, A. (1976). A comparison between the Hamilton Rating Scale and the Beck Inventory in the measurement of depression. *British Journal of Psychiatry, 128,* 486-489.

Bech, P., Bolwig, T., Kramp, P., & Refaelsen, O. (1979). The Beck-Refaelsen Mania Scale and the Hamilton Depression Scale. *Acta Psychiatrica Scandinavia, 59,* 420-430.

Beck, A. T., & Beck, R. W. (1972). Screening depressed patients in family practice. A rapid technique. *Postgraduate Medicine, 52,* 81-85.

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

Beck, A. T., & Steer, R. A. (1984). Internal consistencies of the original and revised Beck Depression Inventories. *Journal of Clinical Psychology, 40,* 1365-1367.

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.

Berndt, D. J. (1979). Taking items out of context: Dimensional shifts with the short form of the Beck Depression Inventory. *Psychological Reports, 45,* 569-570.

Berndt, D. J. (1981). How valid are the subscales of the Multiscore Depression Inventory? *Journal of Clinical Psychology, 37,* 564-570.

Berndt, D. J. (1986) *Multiscore Depression Inventory Manual.* Los Angeles: Western Psychological Services.

Berndt, S. M., Berndt, D. J., & Byars, W. D. (1983). A multiinstitutional study of depression in family practice. *Journal of Family Practice, 16,* 83-87.

Berndt, D. J., Berndt, S. M., & Kaiser, C. F. (1984). Multidimensional assessment of depression. *Journal of Personality Assessment, 48,* 489-494.

Berndt, D. J., & Kaiser, C. F. (1980). An exploration of some reliability results for the Multiscore Depression Inventory. *Psychological Reports, 47,* 823-826.

Berndt, D. J., & Kaiser, C. F. (in press). *Manual for children's version of the Multiscore Depression Inventory.* Los Angeles: Western Psychological Services.

Berndt, D. J., Petzel, T., & Berndt, S. M. (1980). Development and initial evaluation of a Multiscore Depression Inventory. *Journal of Personality Assessment, 44,* 396-403.

Berndt, D. J., Petzel, T. P., & Kaiser, C. F. (1983). Evaluation of a short form of the Multiscore Depression Inventory. *Journal of Consulting and Clinical Psychology, 51,* 790-791.

Berndt, D. J., & Zinn, D. (1982). Prominent features of depression in affective and conduct-disordered inpatients. In D. Offer, K. Howard, & E. Ostrov (Eds.), *Patterns of adolescent self-image* (pp. 45-56). San Francisco: Jossey-Bass.

Biggs, J. T., Wylie, L. T., & Ziegler, V. E. (1978). Validity of the Zung Self-Rating Depression Scale. *British Journal of Psychiatry, 132,* 381-385.

Blatt, S. J., D'Afflitti, J., & Quinlan, D. M. (1976). Experiences of depression in normal young adults. *Journal of Abnormal Psychology, 85,* 383-389.

Blatt, S. J., Quinlan, D. M., Chevron, E. S., McDonald, C., & Zuroff, D. (1982). Dependency and self-criticism: Psychological dimensions of depression. *Journal of Consulting and Clinical Psychology, 50,* 113-115.

Boyd, J. H., Weissman, M. M., Thompson, W. D., & Myers, J. K. (1982). Screening for depression in a community sample. Understanding the discrepancies between depression symptom and diagnostic scales. *Archives of General Psychiatry, 39,* 1195.

Brown, G. L., & Zung, W. W. K. (1972). Depression scales: Self-physician-rating? A validation of certain clinincally observable phenomena. *Comprehensive Psychiatry, 13,* 361-367.

Byerly, F. C., & Carlson, W. A. (1982). Comparison among inpatients, outpatients, and normals on three self-report depression inventories. *Journal of Clinical Psychology, 38,* 797-804.

Cantor, A. (1960). The efficacy of a short form of the MMPI to evaluate depression and morale loss. *Journal of Consulting Psychology, 24,* 14.

Carroll, B. J., Feinberg, M., Smouse, P. E., Rawson, S. G., & Greden, J. F. (1981). The Carroll Rating Scale for Depression I. Development, reliability and validation. *British Journal of Psychiatry, 138,* 194-200.

Carroll B. J., Fielding, J. M., & Blashki, T. G. (1973). Depression rating scales: A critical review. *Archives of General Psychiatry, 28,* 361-366.

Chant, S. N. F., & Myers, C. R. (1936). An approach to the measurement of mental health. *American Journal of Orthopsychiatry, 6,* 134-140.

Christenfeld, R., Lubin, B., & Satin, M. (1978). Concurrent validity of the depression adjective checklist in a normal population. *American Journal of Psychiatry, 135,* 582-584.

Cohen, B. D., & Rau, J. H. (1972). Nonverbal technique for measuring affect using facial expression photographs as stimuli. *Journal of Consulting and Clinical Psychology, 38,* 449-451.

Coleman, R. E., & Miller, A. J. (1975). The relationship between depression and marital maladjustment in a clinic population: A multitrait-multimethod study. *Journal of Consulting and Clinical Psychology, 43,* 647-651.

Comrey, A. L. (1957). A factor analysis of items on the MMPI depression scale. *Educational and Psychological Measurement, 17*, 578-585.

Costello, C. G., & Comrey, A. L. (1967). Scales for measuring depression and anxiety. *Journal of Psychology, 66,* 303-313.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook, vol. 1: Clinical interpretation.* Minneapolis: University of Minnesota Press.

Davies, B., Burrows, G., & Poyton, C. (1975). A comparative study of four depression rating scales. *Australian and New Zealand Journal of Psychiatry, 9,* 21-24.

Dempsey, P. (1964). A unidimensional depression scale for the MMPI. *Journal of Consulting Psychology, 28,* 364-370.

Elkin, I., Parloff, M., Hadley, S., & Autry, I. (1985). The NIMH Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry, 42,* 305-316.

Emery, G. D., Steer, R. A., & Beck, A. T. (1981). Depression, hopelessness, and suicidal intent among heroin addicts. *International Journal of the Addictions, 16,* 425-429.

Endicott, J., Cohen, J., Nee, J., Fleiss, J., & Sarantakos, S. (1981). Hamilton Depression Rating Scale, extracted from regular and change versions of the Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 38,* 98-103.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35,* 837-844.

Feinberg, M., Carroll, B. J., Smouse, P. E., & Rawson, S. G. (1981). The Carroll Rating Scale for Depression III. Comparison with other rating instruments. *British Journal of Psychiatry, 138,* 205-209.

Froese, A., Vasquez, E., Cassem, N. H., & Hackett, T. P. (1974). Validation of anxiety, depression, and denial scales in a coronary care unit. *Journal of Psychosomatic Research, 18,* 137-141.

Gallagher, D., Nies, G., and Thompson, L. W. (1982). Reliability of the Beck Depression Inventory with older adults. *Journal of Consulting and Clinical Psychology, 50,* 152-153.

Gould, J. (1982). A psychometric investigation of the standard and short form Beck Depression Inventory. *Psychological Reports, 51,* 1167-1170.

Guilford, J. P., & Guilford, R. B. (1939). Personality factors D, R, T, and A. *Journal of Abnormal and Social Psychology, 34,* 23-36.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56-62.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6,* 278-296.

Hamilton, M. (1972). Rating scales and depression. In P. Kielholz (Ed.), *Depressive illness: Diagnosis, assessment, treatment.* Baltimore: Williams & Wilkins.

Harris, R. E., & Lingoes, J. C. (1955). Subscales for the MMPI: An aid to profile interpretation. Mimeographed materials, Dept, of Psychiatry, University of California at San Francisco.

Hathaway, S. R., & McKinley, J. C. (1942). A multiphasic personality schedule. *Journal of Psychology, 14,* 73-84.

Hathaway, S. R., & McKinley, J. C. (1989). *Minnesota Multiphasic Personality Inventory—2.* Minneapolis: University of Minnesota Press.

Hedlund, J., & Vieweg, B. (1979). The Hamilton Rating Scale for Depression: A comprehensive review. *Journal of Operational Psychiatry, 10,* 149-162.

Hunt, S. M., Jr., Singer, K., & Cobb, S. (1967). Components of depression. *Archives of General Psychiatry, 16,* 441-448.

Jasper, H. H. (1930). The measurement of depression-elation and its relation to a measure of extroversion-introversion. *Journal of Abnormal and Social Psychology, 25,* 307-318.

Knight, R. G., Waal-Manning, H. J., & Spears, G. F. (1983). Some norms and reliability data for the State Trait Anxiety Inventory and the Zung SRDS. *British Journal of Clinical Psychology, 22,* 245-249.

Krug, S. E., & Laughlin, J. E. (1976). *Handbook for the IPAT depression scale.* Champaign, II: Institute for Personality and Ability Testing.

Lachar, D. (1974). *The MMPI: Clinical assessment and automated interpretation.* Los Angeles: Western Psychological Services.

Lanyon, R. I. (1984). Personality assessment. *Annual Review of Psychology, 35,* 667-701. p. 684

Leckie, E. V., & Withers, R. F. J. (1967). A test of liability to depressive illness. *British Journal of Medical Psychology, 40,* 273-282.

Lewinsohn, P., & Teri, L. (1982). Selection of depressed and nondepressed subjects on the basis of self-report data. *Journal of Consulting and Clinical Psychology, 50,* 590-591.

Lightfoot, S. L., & Oliver, J. M. (1985). The Beck inventory: Psychometric properties in university students. *Journal of Personality Assessment, 49,* 434-436.

Lubin, B. (1965). Adjective checklists for measurement of depression. *Archives of General Psychiatry, 12,* 57-62.

May, A. E., Urquhart, A., & Tarran, J. (1969). Self-evaluation of depression in various diagnostic and therapeutic groups. *Archives of General Psychiatry, 21,* 191-194.

McCall, R. J. (1958). Face validity of the D scale of the MMPI. *Journal of Clinical Psychology, 14,* 77-80.

McNair, D. M. (1974). Self-evaluations of antidepressants. *Psychopharmacologia, 37,* 77-80.

Montgomery, S. A., & Asberg, M. C. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry, 134,* 382-389.

O'Connor, J., Stefic, E., & Gresock, C. (1957). Some patterns of depression. *Journal of Clinical*

Psychology, 13, 122-125.

O'Hara, M., & Rehm, L. (1983). Hamilton rating scale for depression: Reliability and validity of judgements of novice rates. *Journal of Consulting & Clinical Psychology,* 57, 318.

Popoff, L. M. (1969). A simple method for diagnosis of depression by the family physician. *Clinical Medicine, 76,* 24-30.

Radloff, L. S. (1977). The CED-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 358-401.

Rosen, A. (1962). Development of MMPI scale based on a reference group of psychiatric patients. *Psychological Monographs, 76,* 527 (Whole No.), 1-25.

Schwab, J., Bialow, M., & Holzer, C. (1967). A comparison of two rating scales for depression. *Journal of Clinical Psychology, 23,* 94-96.

Snaith, R. P., Ahmed, S. N., Mehta, S., & Hamilton, M. (1971). Assessment of the severity of primary depressive illness: Wakefield Self-Assessment Depression Inventory. *Psychological Medicine, 1,* 143-149.

Snaith, R. P., Bridge, G. W. K., & Hamilton, M. (1976). The Leeds scales for the self-assessment of anxiety and depression. *British Journal of Psychiatry, 128,* 156-165.

Sotsky, S., & Glass, D. (1983). *The Hamilton Rating Scale. A critical appraisal and modification for psychotherapy research.* Paper presented at the annual convention of the Society for Psychotherapy Research, Sheffield, England.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology, 15,* 201-293.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria. *Archives of General Psychiatry, 35,* 773-782.

Steer, R. A., Beck, A. T., & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 121-142). Geneva,

Switzerland: World Health Organization.

Steer, R. A., McElroy, M. G., & Beck, A. T. (1982). Structure of depression in alcoholic men: A partial replication. *Psychological Reports, 50,* 723-728.

Stein, K. B. (1968). The TSC scales: The outcome of a cluster analysis of 550 MMPI items. In P. McReynold (Ed.), *Advances in Psychological Assessment, vol. 1.* Palo Alto, CA: Science and Behavior Books.

Turner, J. A., & Romano, J. M. (1984). Self-report screening measures for depression in chronic pain patients. *Journal of Clinical Psychology, 40,* 909-913.

Weissman, M. M., Prusoff, B., & Newberry, P. B. (1975). *Comparison of CES-D, Zung, Beck self-report depression scales* (Technical Report ADM 42-47-83). Rockville, MD: Center for Epidemiologic Studies, National Institute of Mental Health.

Wessman, A. E., & Ricks, D. F. (1966). *Mood and personality.* New York: Holt. Zung, W. W. K. (1965). a self-rating depression scale. *Archives of General Psychiatry, 12,* 63-70.

Zung, W. W. K. (1972). The depression status inventory: An adjunct to the self-rating depression scale. *Journal of Clinical Psychology, 28,* 539-543.

Notes

[1] The MMPI-II manual was received just prior to this volume's press date. Few changes have been made to the MMPI-D scale. Three items were affected. The major problems discussed above appear to have been carried into the revised version. The most obvious deficiency is the continued lack of adolescent norms. The publisher had reassured this author that the standard and computerized versions of the revised version would have adequate adolescent norms, unlike the current version. The manual (Hathaway & McKinley, 1989) instead refers readers back to the original norms, which were inappropriate at best. Other publishers do have well-developed adolescent norms, including some computerized versions, but the user must be alert to this major omission in the version distributed by the owners of the copyrighted version.