3

Issues in Selecting an Assessment Battery

Larry E. Beutler

Although the preceding chapter has addressed, in general terms, the test characteristics and qualities that should be considered in selecting a set of instruments with which to address defined referral questions, it is important to give greater attention to the availability and selection of specific instruments. Accordingly, this chapter addresses two major contemporary issues that must be considered in this process of selecting assessment procedures. The first of these is the issue of the relative merits of a standard battery of tests, as opposed to those of a battery of instruments selected to address only the specific questions raised by a referring clinician about a given patient.

The second issue is that of the relative merits of qualitative versus quantitative methods of assessing human functioning. This is a particularly salient issue on the contemporary scene, and one that is emerging in many fields of behavioral science. In recent years, a good deal of attention has been given to the development of new paradigms and methods for deepening our understanding of human experience. In clinical practice, the debate between the relative merits of qualitative and quantitative methods arises from a concern that quantitative procedures, arising as they do from group or nomothetic studies, fail to capture the idiosyncrasies that characterize individual motivation and response. Qualitative methods, based on intensive individual analysis of process, promise a new view of individual behavior.

Standard versus Problem-Focused Assessment

Standard Batteries: Content, Advantages, and Disadvantages

Ever since the work of Rapaport, Gill, and Schafer (1946), most clinical treatment programs have advocated and employed a standard set of assessment devices as part of the intake procedure (Sweeney, Clarkin, & Fitzgibbon, 1987). Although the same instruments are administered to all incoming patients, with little or no modification, the interpretation itself is usually modified according to patients' ethnic background, sex, and referral problems. This "standard battery" approach to assessment is designed to provide a broad base of similar and reliable information from which to compare patients, make diagnoses, evaluate areas of patients' strength and weakness, determine prognoses, and plan treatment.

At times, the selection of these standard tests places a premium on brevity. Thus, "screening batteries" are largely comprised of paper-and-pencil instruments that collectively require little clinician time. At other times, especially in long-term inpatient care settings, the standard battery is much more extensive and includes a variety of time-intensive, individual assessment devices (Sweeney et al., 1987). In either case, the instruments selected for a standard battery are chosen in order to ensure the ability to observe a broad array of response domains, and to provide stimulus materials whose demand characteristics represent both simple and complex environments (see Lubin, Larsen, & Matarazzo, 1984; Sweeney et al., 1987).

The kind of test most frequently selected in the modal standard battery is an omnibus personality test of "trait-like" qualities. One or two instruments of this type are often included, in order to obtain behavioral samples from both subjective and objective experience domains. The next most frequently selected instrument type consists of intellectual and cognitive tests designed to determine level of abstract reasoning, problem-solving efficiency, and the nature of cognitive organization. Symptom and other state measures, though high on the list, are less frequently selected in the standard battery than tests of either global personality or general cognitive functioning. When symptom measures are included, however, tests that evaluate several different problem domains and that provide both an estimate of the objective level of social dysfunction and an indication of patient subjective distress are favored.

In terms of specific instruments, the Minnesota Multiphasic Personality Inventory (MMPI), the Wechsler Adult Intelligence Scale–Revised (WAIS-R), and the Rorschach are the most frequently selected devices. The Wechsler Memory Scale (WMS), the Shipley Institute of Living Scale (which yields similar information on intellectual level and efficiency), and the Bender

Visual Motor Gestalt Test (BVMGT) are also frequently included in screening batteries. Specific symptom measures, such as the Beck Depression Inventory (BDI), the Brief Symptom Inventory (BSI), and the BSI's longer counterpart, the Symptom Checklist 90–Revised (SCL-90-R), complete the list of most used instruments (Sweeney et al., 1987).

When an institution or clinical facility places more emphasis on obtaining a comprehensive picture of each patient than on the amount of time required, the intake battery is also likely to include tests that tap the domains of interpersonal functioning. The Thematic Apperception Test (TAT) and some form of sentence completion test are frequently used to assess interpersonal needs and responses in the extended and screening evaluations, respectively. Likewise, projective drawings are among the most frequently selected devices in extended batteries (Sweeney et al., 1987).

Approaching assessment through the use of a general or "core" battery of devices has several advantages over individualized assessment procedures. For example, through the consistent and repeated use of the instruments from a core battery, a clinician may develop a set of explicit expectations about the characteristics of those patients who seek services at a given clinical institution. By referencing setting-specific norms and by observing the patient characteristics that are associated with a good response to the treatment in a particular setting, the clinician develops the ability to extract very individualized interpretations from the test materials. Thus, a core battery may allow highly individualized interpretations because of the increased expertise resulting from an in-depth familiarity with the instruments used.

In addition, a core battery permits the accumulation of a data base that will allow a clinician to review the changes over time in patients applying for service at a given site (and, where applicable, the changes within individuals from one admission to another). Even the overall efficacy of various treatment programs in a facility can be determined if postdischarge follow-up evaluations are included in the standard battery. In contrast, if each entering patient receives a different set of tests based upon his/her particular presentation, it is difficult either to compare patients entering the facility at different times or to estimate the efficacy of the treatment programs established.

On the other hand, there are drawbacks to using a core battery that is applied to everyone. The primary drawback is the lack of flexibility for addressing the unique needs of individual patients. That is, there are questions that a single, all-purpose test battery is simply unable to answer. Because of the insensitivity of omnibus tests to specific neuropathologies, for example, neuropsychological assessments were developed. The modal battery consisting of the MMPI, the Rorschach, the SCL-90-R, the BVMGT, and the WAIS-R simply is ill suited for either identifying the nature of such impairments or localizing neuropathology.

The Problem-Focused Approach

Adapting the neuropsychology model to functional mental health issues, some authors (Sweeney et al., 1987; Clarkin & Hurt, 1988) argue for a more focused or problem-specific form of evaluation as an alternative to the use of a "core" battery. This type of assessment battery is comprised of instruments that are intensely focused on the most salient issues for the patient's diagnosis and treatment; it may be very different for different individuals, depending upon the nature of the questions asked by referring professionals. The advantages of this "individualized" approach lie in its ability to respond specifically to presenting issues and referral questions. Problem-focused assessment allows a more in-depth analysis of a given patient's problems than the usual core battery, because it acknowledges that some tests are better for addressing certain problems than are others.

The prevalence, in practice, of using a core battery approach rather than individualized assessment suggests that many clinicians believe that a problem-focused approach (1) is too expensive or (2) does not provide enough of an increment in knowledge to justify its use over the simpler core battery. This belief is supported by the suggestion that the overwhelmingly large number of patient variables that are important in diagnosis and treatment planning can be reduced to a relatively small number of dimensions (Beutler, 1991; Goldberg, 1992). Under most circumstances, this finite number of patient dimensions is of sufficient specificity to allow extrapolation to the planning needs of most settings and most referral questions.

Of course, the use of a core battery does not by any means preclude the additional use of specific instruments. Indeed, there may be decided advantages to combining these approaches. Two methods are possible. A clinician may use several different core batteries in a given setting, each tailored to particular problems typically presented by patients who come to that setting. For example, many clinics have specialized treatment programs for anxiety disorders, depression, and eating disorders. Depending on a patient's initial complaints, as assessed by the first telephone contact or interview, one of several core batteries may be administered to address these complaints separately.

Alternatively, and perhaps more advantageously, a clinician may use a core battery of a few basic instruments and supplement this battery with individualized tests that reflect the needs of specific patients. Thus, for a person with initial complaints of depression, a standard battery consisting of an omnibus personality test, a symptom checklist, an assessment of social background, and a test of interpersonal relationships may be augmented with tests that are sensitive to mood and affect, memory, and suicidality. The supplemental tests allow desirable individualization in assessing those functional areas that are presented in the referral question, whereas the core tests allow comparisons to be made across patients and time. Using

this approach, a clinician not only can evaluate each of the areas of functioning discussed in Chapter 2, but can tap special needs and deficits that address the specific referral question.

Recommended Instruments for Various Response Domains

As discussed in Chapter 2, within some set of boundary conditions, the behaviors sampled by a given test can generate hypotheses about behaviors that reflect several different response domains. Six such domains are the most central ones to most referral questions: (1) historical background, (2) cognitive functioning, (3) emotional functioning, (4) interpersonal–intrapersonal functioning, (5) diagnostic status, and (6) prognosis and treatment response. (These are the domains covered in Sections IV–VII of the psychological report outline presented in Figure 2.1.) The consultant/clinician who is able to integrate disparate information from multiple sources of information, whether derived from a core or a problem-specific test battery, is in a position to provide accurate information about the meaning and nature of present and future behaviors as they relate to these areas (see Lovitt, 1988). The limits of any test for sampling from and generalizing to these several domains are found in each instrument's reliability, sensitivity, and specificity. Each instrument may be more adept and reliable for assessing some areas of functioning than for assessing others.

Clarkin and Hurt (1988) have identified a number of areas in which reliable and sensitive instruments exist for specific purposes. Adapting their suggestions, Table 3.1 identifies the instruments whose focus and content are most useful for each of the six response domains listed above and presented in Chapter 2. Two points should be noted in reference to Table 3.1. First, the list of tests is only representative, not comprehensive; it does little justice to the very large number of available measures that may be used to assess each area. Indeed, there are instruments that may be better suited for specific purposes than those presented here. This list of instruments represents an effort to balance the adequacy of the information obtained with the time cost of each instrument.

Second, the table does not account for the fact that omnibus, trait-oriented instruments (e.g., the Millon Clinical Multiaxial Inventory [MCMI], the MMPI-2, and the Rorschach) also include special scales and procedures that can be extracted and used for more specific purposes, such as assessing risk for depression, severity of alcohol abuse, and anger control. The reader will find more information about some of these special scales and their uses in the chapters of this book devoted to these tests.

To aid in the selection of instruments to use in either a core or a problem-focused battery, I now describe the instruments presented in Table 3.1. This introduction should provide an initial familiarization for the reader.

TABLE 3.1. Recommended Instruments for Various Response Domains

Domain/instrument(s)	Rater
Historical background	
Life Experiences Survey	Patient
Social Support Questionnaire (SSQ)	Patient
Cognitive functioning	
General functioning	
Mini-Mental State Examination (MMSE)	Clinician
Intellectual functioning	
Wechsler Adult Intelligence Scale–Revised (WAIS-R)	Patient
Shipley Institute of Living Scale	Patient
Memory functions	
Wechsler Memory Scale (WMS)	Patient
Cognitive process/content	
Rorschach	Clinician
Perceptual-motor functioning	
Bender Visual Motor Gestalt Test (BVMGT)	Clinician
Emotional functioning	
General severity and pattern	
Symptom Checklist 90–Revised (SCL-90-R)	Patient
Brief Symptom Inventory (BSI)	Patient
Client Emotional Configuration Scale	Clinician
Depression	
Hamilton Rating Scale for Depression (HRSD)	Clinician
Beck Depression Inventory (BDI)	Patient
Anxiety	
State-Trait Anxiety Inventory (STAI)	Patient
Anger/hostility	
Buss-Durkee Hostility Scale	Patient
Interpersonal-intrapersonal functioning	
Coping style	
Minnesota Multiphasic Personality Inventory–2 (MMPI-2)	Patient
Inventory of Interpersonal Problems	Patient
Structural Analysis of Social Behavior	Clinician
Sexual disturbance	
Derogatis Sexual Functioning Inventory	Patient
Child Abuse Potential Inventory	Patient
Marital/family disturbance	
Dyadic Adjustment Scale	Patient
Family Environment Scale	Patient
Marital Satisfaction Inventory	Patient
Social adjustment	
Social Adjustment Scale–Self-Report	Patient
Michigan Alcoholism Screening Test	Patient
Diagnosis	
Structured Clinical Interview for DSM-III-R (SCID)	Clinician
Structured Interview for DSM-III Personality (SIDP)	Clinician
Anxiety Disorders Interview Schedule (ADIS)	Clinician
Millon Clinical Multiaxial Inventory (MCMI)	Patient

(*continued*)

TABLE 3.1. (Continued)

Domain/instrument(s)	Rater
Prognosis and risk	
Suicide potential	
Scale of Suicide Ideation	Clinician
Beck Hopelessness Scale (BHS)	Patient
Alcohol abuse potential	
Alcohol Use Inventory	Patient
Schizophrenia prognosis	
Camberwell Family Interview	Clinician

Some of these instruments, especially the Structured Clinical Interview for DSM-III-R (SCID), the Mini-Mental State Examination (MMSE), the BVMGT, the MMPI-2, the MCMI, the Rorschach, and the WAIS-R, will be given additional and more intensive consideration in later chapters.

Historical Background

Details about a patient's history can best be obtained with the interview procedures to be discussed in Chapter 4. It is not sufficient simply to know what has happened to an individual, however; a clinician also needs to have an understanding of the impact of these events and the resources that are available to support change.

The objective measurement of life changes and their impacts is very complex. In order to accomplish the task in the most complete fashion, a very extensive, multidimensional assessment procedure is required (Monroe, 1982; Schulz & Tompkins, 1990; Zimmerman, 1983). A less intensive approach to this problem may focus on two related dimensions: life changes and social support systems. The information provided in assessments of these two dimensions will ordinarily be supplemented by the historical information available from diagnostic interviews and procedures, which are discussed somewhat later in this chapter.

The Life Experiences Survey (Sarason, Johnson, & Siegel, 1978) is a 57-item self-report instrument that requires patients to report the subjective impact of change events over the prior year. The scale consists of two parts. The first part refers to life changes that are common to individuals in a variety of situations; to this list, patients are allowed to add events that have been significant and peculiar to them. The second part lists 10 events that are particular to students, and this part is excluded when one is evaluating nonstudents. In both parts, patients first indicate whether the events occurred in the past year and then rate (separately) the desirability and impact of the event along a series of 7-point scales. Scores for positive change, negative change, and total change are obtained.

The Social Support Questionnaire (SSQ; Sarason, Levine, Basham, & Sarason, 1983) is a 27-item self-report inventory designed to assess both perceived number of social supports and satisfaction with these social support systems. Responses to the SSQ have been found to be negatively related to level of subjective discomfort, especially among women; they are also related to subsequent persistence in a difficult or frustrating task. This test provides a relatively efficient method of determining the source and strength of supportive family and social relationships.

Cognitive Functioning

Cognitive functioning is a multidimensional domain. The aspects of functioning that are most salient for most patients include problem-solving level, abstract reasoning abilities, memory, perceptual content and accuracy, and perceptual–motor integration. Cognitive functioning (including these several subareas) is the domain that is given the greatest attention in the assessment of organic and intellectual impairment. The numerous neuropsychological procedures that are used for very specific purposes are not reviewed here. Instead, a few instruments that together provide a range of information within and across the various subareas of cognitive functioning are surveyed.

General Functioning. The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) consists of a brief standardized interview administered by a clinical rater. It has demonstrated good reliability and is sensitive to different pathologies. In particular, it quite adequately distinguishes patients suffering from organic dementia from those with functional disturbances. The MMSE taps the subareas of cognitive control, abstract reasoning, orientation (time, place, and person), memory, and thought processes. It usually serves as a screening device to supplement or replace more intensive and time-consuming assessment procedures that focus on separate aspects of mental state.

Intellectual Functioning. The Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981) is a standardized, individual assessment device that yields three global scores (Performance IQ, Verbal IQ, and Full Scale IQ) and 11 subscales reflecting more specific aspects of cognitive processing. These various aspects of cognitive functioning are most directly represented by two stable factors, expressing performance in the areas of Verbal Comprehension and Perceptual Organization (Silverstein, 1982). The WAIS-R is discussed in detail in Chapter 5.

The Shipley Institute of Living Scale (Shipley, 1940) is a self-administered device comprised of two subscales designed to assess verbal recognition/comprehension and abstract reasoning. These scores are combined

to provide an index of cognitive efficiency and an estimated full-scale IQ that is comparable to that obtained with the longer WAIS-R (Paulson & Tien-Tih, 1970; Zachary, Crumpton, & Spiegel, 1985). The Shipley Institute of Living Scale frequently serves as a screening device when the depth of knowledge available from the WAIS-R is not considered necessary to answer the referral question being asked.

Memory Functions. The Wechsler Memory Scale (WMS; Wechsler, 1945, 1987) is a standardized measure of verbal, perceptual, remote, recent, rote, and logical memory. Visual and auditory stimuli are presented to provide a general estimate of information retrieval and storage. The instrument provides a standard comparison to the levels of functioning that can be expected on the basis of general intelligence, and has become a standard device for most assessment batteries that evaluate cognitive impairments (Cattell & Johnson, 1986).

Cognitive Process/Content. The famous Rorschach Inkblot Test (Rorschach, 1921/1942) is comprised of 10 standard cards. The test is administered in two phases—a free association phase and an inquiry phase. In the first phase, respondents are asked to indicate what the inkblots on the cards appear to be. In the second, they describe the characteristics and qualities of the blots themselves that led them to their responses. The Rorschach provides an avenue by which to observe a patient's thought processes. Unusual cognitive organization and mental content are readily observed in the nature of the patient's response. Recent developments in the standardization of scoring have led to improved reliability and to an increasing array of studies on validity (Exner, 1974). The Rorschach is discussed at greater length in Chapter 6.

Perceptual-Motor Functioning. The Bender Visual Motor Gestalt Test (BVMGT; Bender, 1938) is a brief screening device that was originally designed for detecting brain damage. Its use has been extended to the assessment of other cognitive and personality functions, however (Hutt, 1985; Koppitz, 1975). The BVMGT consists of nine designs presented to a patient in a constant order, with instructions to draw the figures "the best you can." Structural inaccuracies and distortions are scored to assess perceptual-motor integrity and problem-solving organization, with some patterns being used as projective indicators for the presence of interpersonal needs, conflicts, and cognitive integration. To facilitate the use of this instrument in this latter way, variations of the standard administration procedure have frequently been used (see Groth-Marnat, 1990). Although these procedures are promising, the greatest strength of the BVMGT continues to be in its assessment of perceptual-motor and organizational ability.

Emotional Functioning

The domain of emotional functioning, as outlined in Chapter 2, includes the assessment of both mood and affect; estimates of the chronicity of dysphoria, when present; evaluation of emotional stability; and a determination of the level of emotional control that the patient exhibits. The instruments described here and listed in Table 3.1 are designed to allow the assessment of general emotional qualities; symptoms of emotional dysphoria and disturbances; and specific aspects of behavior that are related to depression, anxiety, and anger. These latter areas of disturbance are the most likely ones in which mood and affect will be noted.

General Severity and Pattern. The Symptom Checklist 90–Revised (SCL-90-R; Derogatis, 1977; Derogatis, Rickels, & Rock, 1976) is a 90-item self-report instrument that yields nine symptom scores and three global summary scores. The symptom dimensions include somatization, obsessive–compulsive behaviors, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism (this last score is usually interpreted as reflecting social alienation in nonpsychotic populations). The summary scores reflect overall subjective distress, symptom specificity or spread, and the intensity of presenting symptoms.

The Brief Symptom Inventory (BSI; Derogatis, 1992) is a brief form of the SCL-90-R. Like the parent instrument, this 53-item self-report form yields nine symptom scores and three global summary scores. These scales are the same as those on the longer version. The brevity of the BSI is especially useful in situations requiring rapid assessment, and it has been normalized on older populations (Hale, Cochran, & Hedgepeth, 1984).

The Client Emotional Configuration Scale (Daldrup, Beutler, Engle, & Greenberg, 1988) assesses the mode of emotional expression exhibited by the client. These expressive modes are based upon theoretical descriptions of boundary disturbances as derived from gestalt therapy literature, and include retroflection, introjection, confluence, projection, and deflection. Ratings are made by an experienced clinician using a Likert scale.

Depression. The Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967) provides an independent rating of patient dysphoria. The scale taps such areas as sleep disturbances, libido and sexual functioning disturbance, diffuse somatic complaints, suicide ideation, guilt, and anergia. The clinical utility and reliability of the HRSD have been well documented, and it has been used widely in both research and clinical practice (Endicott, Cohen, Nee, Fleiss, & Sarantakos, 1981).

The Beck Depression Inventory (BDI; Beck, 1978) is an easily administered self-report device for assessing severity of depressive symptoms. It is

readily applied in a repeated administrations and reliably assesses depression level (Beutler & Crago, 1983). Its value lies in the ease with which it is administered and the ecological or face validity of patient responses. Because it is not tied to specific diagnostic criteria, it is not easily used either for establishing a diagnosis or for differentiating between depressed and nondepressed individuals. However, it does provide a stable indicator of depressed mood in clinical, medical, and normal populations.

Anxiety. The State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is a 40-item, multiple-choice, self-report inventory. This is a revised version of the original scale by Spielberger, Gorsuch, and Lushene (1970). It is a quickly administered method of assessing anxiety in two domains. "State anxiety" reflects a patient's stress level in specified situations and can often be taken as an index of how well the patient is presently coping with current anxiety-evoking environments. On the other hand, "trait anxiety" reflects stable individual differences and a general propensity to perceive stressful situations as dangerous. Compared to state anxiety, trait anxiety reflects the adequacy with which a patient is likely to cope with distressing events over time.

Anger/Hostility. The Buss–Durkee Hostility Scale (Buss & Durkee, 1957) assesses the degree of interpersonal anger and associated diminished functioning in four areas. The subscale scores indicate level of experienced anger, the nature of its expression, and the degree to which a respondent internalizes or externalizes blame. Of particular importance for predicting a patient's response to external environments, the test provides an estimate of the degree to which the patient can express anger directly to offending persons in the environment.

Interpersonal–Intrapersonal Functioning

Chapter 2 has outlined some of the dimensions of symptomatic and interpersonal functioning that have been proposed as being among the most relevant for making treatment decisions (see also Beutler & Clarkin, 1990). Of particular concern to the present discussion are patient conflict areas, coping styles, and potential for resisting the influence of others. Chapter 2 has also noted that interpersonal–intrapersonal functioning includes both trait-like and state-like qualities, which must be taken into account during assessment. The trait-like aspects are typically described as aspects of personality, whereas the state-like aspects often reflect levels of distress and reactivity to stress. Thus, to some degree, these latter concepts overlap with the more general concepts considered in connection with emotional functioning. The instruments presented here include both state and trait components.

Coping Style. The Minnesota Multiphasic Personality Inventory–2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) is a 567-item revision of the well-known MMPI (Dahlstrom, Welsh, & Dahlstrom, 1972). It is an empirically derived self-report test on which patients' responses are compared to those of individuals presenting with defined symptoms and disorders. The MMPI-2 is scored on 13 scales, 3 of which are designed to assess the validity and conventionality of a person's responses, and 10 of which reflect various clinical patterns and symptoms. In addition, a very large number of special scales and scale combinations have been developed to apply to specific aspects of coping style, symptom manifestation, type and severity of conflict, authority relationships, and response dispositions. Although most scales of the MMPI-2 reflect trait-like qualities, some are also sensitive to transitory changes and disturbances.

The revision of the MMPI has updated the language of the original version, eliminated questionable items, restandardized the responses based on census-based samples, and introduced some new items that may contribute to the development of promising new scales. The MMPI-2 is discussed at length in Chapter 7.

The Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988) uses a self-report format to assess levels of distress arising from interpersonal relationships. Interpersonal problems are organized into the categories that are thought to represent the types of complaints patients most often bring to psychotherapy. The scale yields six subscales: H. Assertive, H. Sociable, H. Intimate, H. Submissive, T. Responsible, and T. Controlling. These subscales reflect two types of problems: those about which a patient indicates that he/she finds it "hard to" do or be (prefixes of "H"), and those about which the respondent indicates that he/she does or is "too much" (prefixes of "T"). The scale is sensitive to changes induced in short-term psychotherapy, and it has been successfully applied both to studies of psychodynamic psychotherapy processes (Horowitz, 1986) and to predicting the likelihood of negative outcomes in psychotherapy (Mohr et al., 1991).

The Structural Analysis of Social Behavior (Benjamin, 1974) assesses a patient's perceptions of interpersonal events along three dimensions: (1) focus on self or other; (2) affiliation (love–hate and friendliness–hostility); and (3) interdependence (control–autonomy giving). This instrument is based upon a circumplex conceptual system and defines interpersonal exchanges among participants as permutations of the three underlying dimensions (e.g., McLemore & Benjamin, 1979). It describes the tendency of the patient to respond in kind to the friendly or hostile behaviors of others, and in a complementary way to behaviors that occur along the control dimension.

Sexual Disturbance. The Derogatis Sexual Functioning Inventory (Derogatis, 1975) is a self-report, multidimensional instrument made up of

247 items. It assesses functioning in seven primary, sex-defined domains: information, experience, drive, attitudes, affects, gender role definition, and fantasy. In addition, however, it both includes a global rating of sexual satisfaction and totally encompasses 53 items from the SCL-90-R. This latter scale allows the screening evaluation of the presence of general psychiatric symptoms. This scale can be separately scored to yield the same dimensions as the SCL-90-R; hence, individual symptoms as well as global distress can be evaluated.

The Child Abuse Potential Inventory (Milner, 1980) is comprised of 10 subscales. The test is in self-report format, and the six clinical subscales are derived from 77 items reflecting aspects of self–other relationships and impulse control. These subscales are designed to evaluate personal rigidity, unhappiness, and general distress. In addition, they tap three domains of interpersonal and family unhappiness. The nonclinical scales are designed to evaluate the patient's consistency of response, random response sets, and willingness to admit to problems (lie scale) in a manner reminiscent of that used with the MMPI. The inventory has been used both descriptively (Milner, 1989) and predictively (Milner & Robertson, 1989), with impressive results.

Marital/Family Disturbance. The Dyadic Adjustment Scale (Spanier, 1976) is the most widely used self-report instrument in the clinical and nonclinical research literature for assessing marital satisfaction and adjustment. The instrument is reported to have excellent psychometric properties both for evaluating the presence of marital disturbance and for highlighting some specific areas of concern, such as sexual, financial, and communication difficulties (Jacob & Tennenbaum, 1988).

The Family Environment Scale (Moos & Moos, 1976, 1981) is a self-report measure that is completed by both the subject and a significant other of the subject's choosing. It is designed to assess 10 domains of family environment: cohesion, expressiveness, conflict, independence, achievement orientation, intellectual/cultural orientation, active/recreational orientation, moral/religious emphasis, organization, and control. Validity and reliability studies have shown the instrument to be statistically sound and meaningful.

The Marital Satisfaction Inventory (Snyder, 1979) is a 280-item self-report instrument presented in a true–false format. The content covers marital relations, with a separate section adapted for use among couples with children. Eleven areas are covered: conventionalization, global distress, affective communication, problem-solving communication, time together, disagreement about finances, sexual dissatisfaction, role orientation, family history of distress, dissatisfaction with children, and conflict over childbearing. Internal-consistency coefficients and test–retest reliabilities average about .90, and research has supported the discriminant validity of the scales (Snyder, Wills, & Keiser, 1981).

Social Adjustment. The Social Adjustment Scale–Self-Report (Weissman & Bothwell, 1976) is comprised of 42 questions relating to everyday adjustment and performance. The questions cover such areas as impairment and adequacy in social role perfomance at work and home, leisure activities, relationships with significant others, integrity of the family unit, and economic self-support. Available norms allow comparisons to be made both to nonpatient and various patient samples.

The Michigan Alcoholism Screening Test (Selzer, 1971) is a brief screening procedure for the detection of excessive alcohol consumption. It is frequently used for the classification of drinking patterns, as well as for determining the clinical significance of changes in alcohol consumption. It is frequently used in clinical practice for obtaining an indication of drinking severity based upon patient self-report.

Diagnosis

Determination of a patient's diagnosis is one of the primary reasons for referral. Although the instruments described up to this point variously assess symptom severity and patterns of interpersonal behaviors, they do not assess diagnostic syndromes. "Syndromes" are clusters of symptoms and signs that are consensually believed to represent distinctive and frequently occurring patterns. Because diagnoses are criteria-based patterns of related symptoms, they can efficiently be determined only with instruments whose structure parallels the current edition of the DSM or ICD.

The unstructured clinical interview is the most frequently used diagnostic device, but has not been found historically to be particularly reliable for this purpose (see Chapter 4). There are instruments that are better suited and more efficient for establishing a diagnosis, however; a few of these are reviewed here.

The Structured Clinical Interview for DSM-III-R (SCID; Spitzer, Williams, & Gibbon, 1986) has modules for assessing both Axis I and Axis II disorders. The interview format is determined strictly by the DSM-III-R criteria (American Psychiatric Association, 1987), and the authors report acceptable reliability scores. Raters are trained on criteria-based videotape samples to ensure comparability.

The Structured Interview for DSM-III Personality (SIDP; Pfohl, Stangl, & Zimmerman, 1983) is a semistructured, independently rated interview assessing the criteria for personality disorders as described in DSM-III. Its format and structure parallel those of the SCID, and its validity is assessed against the criteria of DSM Axis II disorders.

The Anxiety Disorders Interview Schedule (ADIS; DiNardo, O'Brien, Barlow, Waddell, & Blanchard, 1983) is an interview-based assessment of the anxiety symptoms presented in DSM-III. It yields a diagnosis of both type of anxiety disorder and severity; it is thus useful for responding to

diagnostic questions as well as descriptive ones, and is closely linked to the implementation of treatment strategies for specific symptoms as well.

The Millon Clinical Multiaxial Inventory–II (MCMI-II; Millon, 1987) is a revision of the original 175-item MCMI-I (Millon, 1983). It is a multidimensional self-report instrument that yields scales designed to reflect syndromes that parallel DSM-III-R diagnostic criteria. Three basic clusters of scores are obtained–one reflecting patterns that define basic personality styles corresponding with eight of the Axis II disorders, one that assesses more severe personality patterns corresponding with Schizotypal, Borderline, and Paranoid Personality Disorders; and one that assesses patterns corresponding with more circumscribed or transient clinical syndromes. This last cluster includes the spectrum of anxiety syndromes, somatoform disorder, hypomanic disorders, substance abuse disorders, severe depression, and psychotic thinking processes. The MCMI has frequently been proposed as a shorter alternative to the MMPI (Gynther & Gynther, 1983), but current research suggests that the two instruments differ, particularly on the dimension of diagnostic specificity (McCann, 1991). Chapter 8 of this volume discusses the development and use of the MCMI (including that of the latest version, the MCMI-III) in greater detail.

Prognosis and Risk

Most, if not all, of the instruments discussed in the preceding paragraphs have been used to assess prognosis and to explore ways of matching treatments to patient needs. Combinations of scales from the MMPI, for example, have been found to predict differential responses to insight- and symptom-focused treatments and to directive and nondirective treatments (Beutler et al., 1990; Beutler & Mitchell, 1981); the ADIS has been used to determine the probable value of cognitive and exposure therapies (Barlow, 1985); and the SCL-90-R is a frequently used measure of clinical effectiveness (Beutler & Crago, 1983). This section, however, describes several instruments whose purpose is specifically to assess risk and prognosis in selective areas of functioning.

Suicide Potential. The Scale of Suicide Ideation (Beck, Kovacs, & Weissman, 1979) is an interview-based instrument that is administered by a trained clinician. It is designed to quantify clinical indicators of suicide potential. It focuses on the intensity of current conscious suicidal intent by tapping the presence of self-destructive thoughts and wishes, suicidal threats, overt suicidal plans, and depressive cognitions. It employs a flexible format in order to allow the clinician to elicit as much information as possible in each area, in order to accurately determine the presence of suicidal behaviors and to estimate the probability of future suicidal acts.

The Beck Hopelessness Scale (BHS; Beck, Weissman, Lester, & Trexler,

1974) is a 20-item, self-report, true-false questionnaire that assesses the aspect of clinical depression that most closely relates to suicidal behavior. A patient's sense of futility and hopelessness has been found to be more predictive than depressed mood of risk for suicidality. Accordingly, the BHS has good internal consistency and concurrent validity, and is sensitive to relatively small changes in depression and suicidal thoughts over time.

Alcohol Abuse Potential. The Alcohol Use Inventory (Horn, Wanberg, & Foster, 1974; Wanberg, Horn, & Foster, 1977) was developed to provide an assessment of the nature and range of alcohol-related problems. The 147 items of the instrument are grouped into 22 scales and organized around a three-factor structure: styles of alcohol use, symptoms and consequences of alcohol use, and perceived benefits of drinking. It appears to yield quite reliable and consensually valid responses (Wanberg et al., 1977; Wanberg & Horn, 1983).

Schizophrenia Prognosis. The Camberwell Family Interview (Brown & Rutter, 1966) is a semistructured interview designed for administration to a significant family member of a patient with schizophrenia. It yields information about "expressed emotion," based upon ratings of criticality, hostility, and overinvolvement. This interview requires extensive training and experience for reliable adminstration, but does appear to be a significant predictor of prognosis. For example, not only does the level of expressed emotion predict rehospitalization, but treatments that modify patterns of expressed emotion are successful in reducing rehospitalization rates.

Comment

It should be reiterateed that the list and descriptions of psychological tests provided in the foregoing pages and in Table 3.1 are far from complete; the most notable omissions are instruments specifically assessing neuropsychological functions. Nonetheless, the clinician who is armed with an understanding of these tests will be able to address most of the questions that form the basis for referral. With this in mind, let us now turn to a consideration of the other major issue to be addressed in this chapter—seeking a balance between qualitative and quantitative assessment and interpretative methods.

Qualitative versus Quantitative Assessment

Dissatisfaction with Quantitative Methods

In Chapter 1, the descriptions of measurement procedures have emphasized the role of normative comparisons, reliability estimates, and validity

demonstrations. All of these properties of measurement involve numbers and numerical concepts. Sensitivity and specificity are expressed as percentages; validity and reliability are expressed as correlations between numbers; norms reflect means and standard deviations of numbers. Within this quantitative perspective, important information about the value of different measures is obtained when observations can be transformed into numbers, compared through numerical manipulations, and then translated back to descriptive language.

Preferences for quantitative measurement and methodologies have characterized the fields of psychology and measurement theory for several decades, especially in academic circles. This preference for numbers—which are explicit in meaning, replicable, and comparable from person to person—may account for the rise and success of "empirical" tests like the MMPI and MCMI, as well as for the relative demise of "rational" tests like the Rorschach and TAT in academic circles.

"Empirical" tests are those based upon the demonstration that the scores (numbers) are different among patients with different, known characteristics (i.e., normative and criteria-group comparisons). These empirical demonstrations are at the very foundation of quantitative assessment, and, of necessity, rely on the demonstration of *group* differences in numerical scores. However, some professionals in the field have become disillusioned with quantitative methods and have criticized academic psychology and measurement theorists for the failure to attend to individual idiosyncrasies. These individuals attach far less importance to subgroup norms as the basis for assessing the value of clinical methods. They maintain that comparing a given individual to a standard based upon small criteria groups, as a means of determining the meaning of that individual's behavior, obscures clinically relevant uniqueness. They favor, instead, an "ipsative" description of the person, in which each individual serves as his/her own reference point for describing relative strengths and weaknesses.

The latter approach has been particularly favored by clinicians. Not surprisingly, practitioners who work daily with people are often less persuaded by demonstrations that an individual's test scores are either different from or similar to those of various reference groups than are academic psychologists, who are more familiar and comfortable with numerical concepts. Hence, although academic psychologists frequently criticize and even eschew tests like the Rorschach, clinicians continue to use such tests as a basis for developing clinical impressions. Many clinicians hold the belief that in comparison to empirically derived tests, "rationally" constructed and interpreted ones capture more of the complexity of human behavior. Tests like the Rorschach and projective figure drawings, for example, derive from theoretical concepts rather than empirical demonstrations, and their interpretation involves a process of synthisizing abstract concepts from nonnumerical productions. These procedures purport to offer a method for

conceptualizing and assessing the complexity of intrapsychic needs and conflicts.

Efforts to Bridge the Gap

Exner's (1974) and others' translations of Rorschach narratives into numbers represent efforts to bridge the gap between the empirical and rational viewpoints by reconstructing qualitative narratives into a numerical scoring system to which quantitative methods can be empirically applied. As these efforts demonstrate, nothing in narrative productions inherently precludes clinicians from describing individuals by the use of quantitative scores; respondents can easily be described by reference to deviations from group mean values. Yet the emergence of non-numerical methods in contemporary psychological research belies the assertion that quantitative measurement procedures are well adapted to the verbal narrations and artistic productions widely used by clinicians. The complex and multidimensional relationships described in clinical formulations of personality functioning are difficult to distill into numbers. Even if procedures such as those developed by Exner are used, and numbers that capture some degree of the complexity represented in these clinical formulations (e.g., ratios of Rorschach determinants) can be constructed, many clinicians are concerned that these numbers fail to preserve the character of the phenomena being observed. They ask such questions as these: Do ratios and combinations of numbers adequately capture the essence of love? Do they adequately distinguish among different kinds of nonobservable experiences (love, anger, lust, etc.)? Do numbers adequately allow us to compare the amount that people love their wives or husbands with the amount that they love their mothers? Can numbers capture the variations in love-driven behaviors that occur when a child's life is threatened or when a spouse or lover is unfaithful?

It is equally hard to place other concepts—for instance, "conflicts," "ego," "anger," and "impulse"—within a numerical framework. These concepts often emerge in narrative form in the rationally derived clinical methods. These narratives are thought to represent the interplay of numerous complex forces and to allow a unique picture of individual, rather than group, behavior to emerge. Consider, for example, a patient's narrative response to a Rorschach card (card VIII):

> This is the face of a rooster. He has been killed. Oops! He's wearing sunglasses and has an extra eye on top of his comb. His insides are rotting, and here is where his spine is breaking through the skin and poking out.

Does a score that identifies the location of the percept, the use of form, and the content—"rooster," "sunglasses," and "anatomy"—adequately cap-

ture the essence of this narrative response? Advocates of the methods of "narrative assessment" and "hermeneutics" represent increasingly persuasive forces within contemporary measurement theory. These qualitative methods, attending as they do to the wholistic structure and content of natural language, have a particular affinity for clinicians for whom quantitative methods do not appear to be adaptable to describing the complexities and color of individual differences.

My colleagues and I believe that qualitative methods do offer an additional perspective in the measurement of human experience. However, we also believe that quantitative and qualitative methods are not inherently in opposition to each other; in fact, they are potentially synergistic (i.e., they can complement and add to each other). Qualitative methods of interpretation emphasize idiographic (i.e., idiosyncratic) patterns, whereas quantitative methods are distinguished by their nomothetic (i.e., normative) basis of deriving meanings from patient productions. The former methods rely on an ipsative comparison, in which various qualities of the patient himself/herself serve as a standard of relative comparison; the latter methods emphasize a normative or group comparison, in which the patient is compared to an outside norm reflective of others' responses. Narrative descriptions can enliven and deepen an understanding of test scores, while test scores can be used both to ensure the objectivity of narratives and allow a normative interpretation.

Cautions about Qualitative Interpretations

At their current level of development, qualitative interpretations of test materials are subject to several sources of error. The interpretations may not be accurate; they may not be replicable or constant; they may reflect a rater's mood or diet rather than actual characteristics of the patient; and they may have no heuristic value for predicting and planning treatment. Thus, even qualitative interpretive methods must come to grips with issues of reliability and validity. In order to be useful, non-numerical concepts (such as those complex verbal ones that characterize narrative descriptions) must be capable of reliable classification, and each category must be distinguishable from others. That is, a clinician must be able to assert that a conflict with a mother is manifestly different from a conflict with a wife; that two ego states are different; that aggressive impulses are different from sexual ones; or that two dynamic intrapsychic patterns differ from each other.

Fundamentally, qualitative interpretations encounter the same problems of measurement as traditional quantitative methods, because they must be assured of at least construct validity (Hogan & Nicholson, 1988), if not of sensitivity, specificity, and reliability. Measurement, at least at the nomi-

nal level, is needed in order to establish the value of these procedures. Nominal measurement allows narratives to be subject to assessments of interrater reliability, and thereby helps to assure that the qualities observed are not simply a reflection of irrelevant qualities of the clinical interpreter.

The "Scrud Test" as an Example

Let us further consider the interrelationships of normative and ipsative interpretation by reflecting on some of the examples used to illustrate the concepts of response variability, reliability, and validity in previous chapters. Contrast, for example, the limited response variabilities to both the "President's Test" and the "Chalk Test" (see Chapter 1) with what would happen if we asked students to draw a "scrud," a meaningless term in the English language (see Chapter 2).[1] In the first two tests, some consistency is expected among the responses of different people; in the "Scrud Test," no two individuals' responses will be the same, because everyone will have his/her own idea of a "scrud." Since the test environment that we have constructed with our instructions is held constant—it is the same for everyone—we must assume that the variability among responses reveals something unique about each of the respondents. Thus, each response is a unique production that should provide us with an avenue into the idiosyncratic nature of each individual's internal experience.

But what does each unique response mean? Here is where the test's "normative value" is necessary for an accurate interpretation. We must keep in mind that all responses are unique, but that they also have qualities similar to those of others' responses. Hence, we can begin our analysis by first looking at ways in which the responses of different people are similar. This will allow us to define what constitutes a "usual" response to the instructions. If we know what constitute the usual and the unusual aspects of an individual's response to the test environment we have created, we can begin to assess what qualities of people are associated with making these various responses. Our "normative" reasoning follows the logic that normatively unusual responses, which are nonetheless similar to one another, may indicate that the people who produced them are also similar to one another in some ways.

This point can can be illustrated by reference to Figure 3.1. This figure presents four drawings made by a college student who was asked to draw, in a randomized order, the following things: (1) two lines that love each other, (2) two lines that hate each other, (3) a happy line, and (4) an angry

[1]In spite of some obvious similarities, the "Scrud Test" described here and the "Blivet Test" described by McIvor (1979) as a parody on projective methods were independently developed.

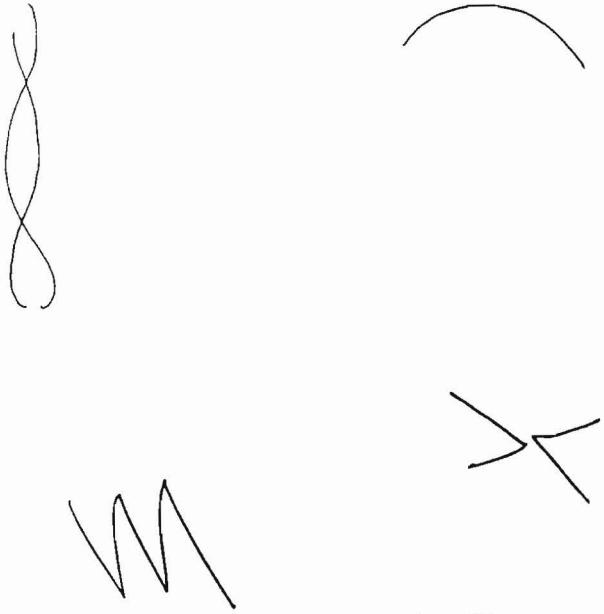

FIGURE 3.1. Emotional qualities of lines.

line. Most of us can probably classify which lines are which at a rate of accuracy that exceeds chance (chance = 25%). We can do this because most people respond in a similar way. The two entwined and curving lines represent "loving lines"; the two angulated and heavy lines represent "hateful lines"; the soft, curved line is a "happy line"; and the jagged, heavy line is an "angry line." When we know these expected responses, we can determine whether a particular person's responses are unusual.

To extrapolate this example to our "Scrud Test," if we observe that there are striking similarities among all of those who draw "scruds," we may infer that "scrudness" is a frequently occurring quality among people. Likewise, we may inspect the internal consistency of the "scrud" by assessing the similarities of different aspects of the same drawing. For example, we may measure the degree to which each quadrant of a respondent's "scrud" is composed of the same types of "loving," "hateful," "happy," or "angry" lines. Furthermore, we may infer qualities of the person by comparing a given drawing to some standard of "usual" shape, pencil pressure, and organization. We may even augment our drawing task by giving the person an equivalent form of the test. We may ask him/her, for example, to draw a "brump." If the person's "brump" is "scrud"-like, we can conclude that both drawings reflect the same qualities of the person. If a similar figure is produced at a much later time, when the same person is asked to draw a

"brump," we may conclude that "scrudness" is a stable quality that does not erode with time. All of these conclusions require a normative basis of comparison.

Alternatively, using a rational approach to interpreting a person's artistic productions, we can apply the normative qualities of line drawing in order to look at the idiosyncratic or unique aspects of the person. For example, armed with a theoretical formulation that people "project" their unwanted inner feelings onto unstructured environments, we *may* be justified in inferring that if the person has a propensity for drawing a "scrud" whose lines resemble "angry" ones, he/she may be angry, or that if the person has a predilection for drawing lines that resemble "loving" ones, he/she may be affiliative and caring. Thus, we may look at the structure of a given person's "scrud" and conclude that the uniqueness of this person is captured in (1) intolerance for ambiguity—the person has a need to impose order on his/her environment (the "scrud" is an object with a defined form, rather than an abstract figure), (2) a tendency to avoid self-exposure (the "scrud" is a smaller figure than usual), and (3) hidden anger and resentment (the "scrud" is drawn with heavy lines and has many sharp corners).

However, caution is necessary. Although it may be justified on theoretical grounds to infer that because a person's "scrud" and "brump" are composed entirely of soft, curved lines, he/she has the characteristics of being happy and loving, we can only do so if we have demonstrated the validity of these conclusions on some conceptual basis. A quantitative standard would argue that such an interpretation is warranted only if those who draw such lines exhibit behaviors or score on other tests that are known to indicate these other qualities. This is "construct validity." The logic of such an approach is that if a person makes drawings like those of loving and contented people, he/she is probably also loving and contented; however, this quantitative approach requires a reduction of drawings to a comparison of numbers. Because, as duly noted above, number reduction is not always in keeping with the effort to retain the richness of global productions, contemporary qualitative measurement emphasizes other criteria of "validity." These criteria are used both in selecting procedures that rely on qualitative interpretation and in the actual use of interpretive analytic procedures.

Criteria for the Use of Qualitative Methods

The following is a tentative list of evaluative criteria for judging the validity of procedures that are reliant on the analysis of narrative productions. This list represents an adaptation of guidelines provided to the review board members of the *Journal of Consulting and Clinical Psychology*, in order to assist

them in evaluating the adequacy of qualitative research.[2] In the current context, these guidelines are considered to be useful when a clinician either selects instruments that are based on qualitative methods or undertakes to depart from or supplement a quantitative interpretation of patient productions.

1. *Method appropriateness.* A preliminary test of the usefulness and appropriateness of selecting qualitative assessment procedures is based on an initial determination that more economical procedures are either unavailable or inappropriate to the referral question. This concern with the appropriateness of the method accepts the proposition that quantitative methods are more clearly developed than qualitative ones and should be given preference whenever possible. The use of qualitative analysis must be clearly more appropriate than quantitative analytic methods to the subject matter, questions, and goals of the referral question. Moreover, when used, qualitative methods should be supplemented by quantitative ones whenever possible.

2. *Openness.* Qualitative analyses of test productions should clearly be framed within an explicit statement of the theoretical orientation that underlies it. Where appropriate, the internal processes and relevant reactions of those who interpret the data should be made explicit.

3. *Theoretical sensitivity.* Accepted theory, rather than personal adaptations, should be used to inform and guide the selection and interpretation of qualitative methods. This requires that a clinician be well informed as to current theoretical developments and protect, as much as possible, the interpretation from his/her idiosyncratic adaptations of theory.

4. *Bracketing of expectations.* If a clinician's interpretations of test productions depart from consensual theoretical formulations, this should be stated explicitly, in order to make the referring professional aware of the potential influence of the clinician's implicit expectations or biases. Where possible, the clinican should make use of checks against possible bias (e.g., having another clinician review his/her analysis).

5. *Responsibility.* In order to allow others to judge their conclusions, clinicians who select and interpret qualitative assessment procedures should clearly describe the nature of the procedures used and the conditions under which they are administered, as well as any efforts to achieve a consensual interpretation.

6. *Saturation/generalizability.* Where interpretations are intended to

[2]These criteria do not represent my own original ideas. They are adapted from the suggestions of Robert Elliot, who produced them upon the recommendation of Fredrick Newman, an associate editor of the journal. They also reflect the input of Judith Green, who served as a reviewer of the guidelines before they were submitted to the review board. I thank Drs. Elliot, Newman, and Green for their contributions; at the same time, I would like to acknowledge that these latter authors are not responsible for my own interpretation of their work.

indicate the presence of trait-like qualities, clinicians should make efforts to ensure that they have sampled an appropriate number and range of situations and methods to provide a thorough description of the phenomenon they have targeted. Where interpretations are meant to suggest the presence of situational responses, on the other hand, clinicians should ensure that these criteria samples of behavior have been studied thoroughly and comprehensively. Clinicians should be aware of the nature and limitations of their behavior samples.

7. *Verification methods.* The strongest test of the validity of qualitative procedures is the assurance that the interpretive categories, descriptions, themes, and interpretations are cross-validated. The most systematic methods for verification include the following:

a. *External verification.* The categories and themes that are extracted from the presentations are demonstrably related to some other variables (e.g., outcome).

b. *Testimonial/informant validity.* The interpretations are consistent with the reports of informants.

c. *Analytic "auditing" procedures.* Multiple qualitative analyses are undertaken. This may include either using an outside "auditor," or adding a same-analyst "verification step" for checking the interpretations for discrepancies.

d. *Triangulation.* Evidence of agreement is checked among multiple and varied perspectives, in order to identify the common or fundamental processes that underlie these different perspectives.

8. *Grounding.* The clinician who reports qualitative data does well to provide some examples of responses in the resulting report to illustrate the interpretations. The clinician should also take care to avoid departing substantially from the data in making the accompanying interpretations.

9. *Coherence.* The categories and constructs that arise in the interpretation of qualitative productions should be checked for how well they fit together to form a coherent story, narrative, "map," or framework for the phenomena or domain.

10. *Believability/helpfulness to readers.* When described in the report to the referring professional, the interpretations must be integrated with other data and presented in a believable narrative that enables the referrer or other readers to understand the patient's experience and presentation more explicitly and more fully than they would without this report.

11. *Intelligibility.* The resulting written report should be presented in a clear, accessible fashion, free of unnecessary jargon.

Conclusions

This chapter has addressed two dimensions on which the clinician must seek balance in the selection of psychological assessment methods. The first of

these is the balance between the advantages of using an all-purpose core battery and those of using procedures that vary from individual to individual, depending upon the nature of the referral question presented. I have argued in favor of a combination of standard instruments, which allow the accumulation both of clinician familiarity and of setting-specific normative data, and individualized instruments, which are selected because they have the power to extract information that is relevant to the referral questions being asked.

This recommendation is based upon my colleagues' and my observations that different procedures are more or less effective for deriving valid information, depending upon the domain of experience and functioning being targeted for assessment. Hence, the use of specialized instruments makes sense, and I have discussed a number of instruments in terms of the areas of their special strengths. However, all instruments contain information concerning a common core of functional abilities, and most questions asked of clinicians can be answered by addressing a relatively small number of functional domains to which patients may be exposed in everyday experience. Thus, applying a small number of well-established core instruments, which tap the most common domains of functioning in a reliable and valid way, may allow the clinician to develop a high level of skill and to acquire a setting-specific internal norm with which to refine the interpretation of findings.

The second type of balancing required is that between the use of qualitative and quantitative data collection methods. As I have noted, reducing observations to numbers for quantitative analysis may fail to capture some subtle and representative patterns that exist in complex behaviors—patterns that may be maximally helpful in responding to referral questions about cognitive processing and personality. Accordingly, qualitative methods are recommended if quantitative ones are either unavailable or inappropriate to the question(s) raised. However, interpretations and life-changing decisions based upon qualitative methods must be protected against clinician bias. Thus, cross-validation with quantitative methods, reliance on theory-consistent interpretations, explicit descriptions of how procedures are used, and the establishment of consensual checks are often necessary to enhance and ensure the validity of these procedures. Ultimately, a combination of procedures may serve to be most useful in responding to the needs of most referring professionals.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Barlow, D. H. (Ed.). (1985). *Clinical handbook of psychological disorders: A step-by-step treatment manual.* New York: Guilford Press.

Beck, A. T. (1978). *Depression Inventory*. Philadelphia: Center for Cognitive Therapy.

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale of Suicide Ideation. *Journal of Consulting and Clinical Psychology, 47*, 343–352.

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42*, 861–865.

Bender, L. (1938). *A visual motor gestalt test and its clinical uses* (Research Monograph No. 3). New York: American Orthopsychiatric Association.

Benjamin, L. S. (1974). Structural Analysis of Social Behavior. *Psychological Review, 81*, 392–445.

Beutler, L. E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology, 59*, 226–232.

Beutler, L. E., & Clarkin, J. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions*. New York: Brunner/Mazel.

Beutler, L. E., & Crago, M. (1983). Self-report instruments. In M. J. Lambert, E. R. Christensen, & S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 453–497). New York: Brunner/Mazel.

Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1990, June). *Predictors of differential response to therapeutic procedures*. Paper presented at the annual meeting of the Society for Psychotherapy Research, Toronto.

Beutler, L. E., & Mitchell, R. (1981). Psychotherapy outcome in depressed and impulsive patients as a function of analytic and experiential treatment procedures. *Psychiatry, 44*, 297–306.

Brown, G. W., & Rutter, M. L. (1966). The measurement of family activities and relationships. *Human Relations, 19*, 241–263.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21*, 343–349.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory–2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Cattell, R. B., & Johnson, R. C. (1986). *Functional psychological testing*. New York: Brunner/Mazel.

Clarkin, J. F., & Hurt, S. W. (1988). Psychological assessment: Tests and rating scales. In J. A. Talbott, R. E. Hales, & S. C. Yodofsky (Eds.), *The American Psychiatric Press textbook of psychiatry* (pp. 225–246). Washington, DC: American Psychiatric Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: University of Minnesota Press.

Daldrup, R. J., Beutler, L. E., Engle, D., & Greenberg, L. S. (1988). *Focused expressive psychotherapy: Freeing the overcontrolled patient*. New York: Guilford Press.

Derogatis, L. R. (1975). *Derogatis Sexual Functioning Inventory*. Baltimore: Johns Hopkins University Press.

Derogatis, L. R. (1977). *The SCL-90 manual I: Scoring, administration and procedures*. Baltimore: Johns Hopkins University School of Medicine, Clinical Psychometrics Unit.

Derogatis, L. R. (1992). *BSI: Administration, scoring and procedures manual–II* (2nd ed.). Baltimore: Clinical Psychometric Research.

Derogatis, L. R., Rickels, K., & Rock, A. F. (1976). The new SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry, 128*, 280–289.

DiNardo, P. A., O'Brien, G. T., Barlow, D. H., Waddell, M. T., & Blanchard, E. B. (1983). Reliability of DSM-III anxiety disorder categories using a new structured interview. *Archives of General Psychiatry, 40*, 1070–1075.

Endicott, J., Cohen, J., Nee, J., Fleiss, J., & Sarantakos, S. (1981). Hamilton Depression Rating Scale. *Archives of General Psychiatry, 38*, 98–103.

Exner, J. E., Jr. (1974). *The Rorschach: A comprehensive system.* New York: Wiley.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.

Groth-Marnat, G. (1990). *Handbook of psychological assessment* (2nd ed.). New York: Wiley.

Gynther, M. D., & Gynther, R. A. (1983). Personality inventories. In I. B. Weiner (Ed.), *Clinical methods in psychology* (2nd ed., pp. 152–232). New York: Wiley.

Hale, W. D., Cochran, C. D., & Hedgepeth, B. E. (1984). Norms for the elderly on the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology, 52*, 231–232.

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

Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621–626.

Horn, J. L., Wanberg, K. W., Foster, F. M. (1974). *The Alcohol Use Inventory.* Denver: Center for Alcohol Abuse Research and Evaluation.

Horowitz, L. M. (1986, June). *Interpersonal problems and dynamic case formulations.* Paper presented at the annual meeting of the Society for Psychotherapy Research, Wellesley, MA.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G. & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.

Hutt, M. L. (1985). *The Hutt adaptation of the Bender-Gestalt test* (4th ed.). New York: Grune & Stratton.

Jacob, T., & Tennenbaum, D. (1988). *Family assessment: Rationale, methods and future directions.* New York: Plenum Press.

Koppitz, E. M. (1975). *The Bender-Gestalt Test for young children: Vol. 2. Research applications 1963–1973.* New York: Grune & Stratton.

Lovitt, R. (1988). Current practice of psychological assessment: Response to Sweeney, Clarkin, and Fitzgibbon. *Professional Psychology: Research and Practice, 19*, 516–521.

Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935–1982. *American Psychologist, 39*, 451–454.

McCann, J. T. (1991). Convergent and discriminant validity of the MCMI-II and MMPI personality disorder scales. *Psychological Assessment, 3*, 9–18.

McIvor, D. (1979). The Rorschach-Bender-McIvor Blivet Test. *Journal of Irreproducible Results, 25, 3-5.*

McLemore, C., & Benjamin, L. S. (1979). Whatever happened to interpersonal diagnosis?: A psychosocial alternative to DSM-III. *American Psychologist, 34,* 17-34.

Millon, T. (1983). *Millon Clinical Multiaxial Inventory manual* (3rd ed.). Minneapolis: National Computer Systems.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory–II (MCMI-II) manual.* Minneapolis: National Computer Systems.

Milner, J. S. (1980). *The Child Abuse Potential Inventory: Manual.* Webster, NC: Psytec.

Milner, J. S. (1989). Additional cross-validation of the Child Abuse Potential Inventory. *Psychological Assessment, 1,* 219-223.

Milner, J. S., & Robertson, K. R. (1989). Inconsistent response patterns and the prediction of child maltreatment. *Child Abuse and Neglect, 13,* 59-64.

Mohr, D. C., Beutler, L. E., Engle, D., Shoham-Salomon, V., Bergan, J., Kaszniak, A. W., & Yost, E. (1990). Identification of patients at risk for non-response and negative outcome in psychotherapy. *Journal of Consulting and Clinical Psychology, 58,* 622-628.

Monroe, S. M. (1982). Life events assessment: Current practices, emerging trends. *Clinical Psychology Review, 2,* 435-453.

Moos, R., & Moos, B. (1976). A typology of family social environments. *Family Process, 15,* 357-372.

Moos, R., & Moos, B. (1981). *The process of recovery from alcoholism: III. Comparing family functioning in alcohol and matched control families.* Palo Alto, CA: Social Ecology Laboratory, Stanford University, and Veterans Administration Medical Center.

Paulson, M. J., & Tien-Teh, L. (1970). Predicting WAIS IQ from Shipley-Hartford scores. *Journal of Clinical Psychology, 26,* 453-461.

Pfohl, B., Stangl, D., & Zimmerman, M. (1983). The implications of the DSM-III personality disorders for patients with major depressions. *Journal of Affective Disorders, 7,* 309-318.

Rapaport, D., Gill, M., & Schafer, R. (1946). *Diagnostic psychological testing* (Vol. 1). Chicago: Year Book Medical.

Rorschach, H. (1942). *Psychodiagnostics.* Bern: Bircher. (Original work published 1921)

Sarason, I. G., Johnson, J. H., & Siegel, J. M. (1978). Assessing the impact of life changes: Development of the Life Experiences Survey. *Journal of Consulting and Clinical Psychology, 46,* 932-946.

Sarason, I. G., Levine, H. M., Basham, R. B., & Sarason, B. R. (1983). Assessing social support: The Social Support Questionnaire. *Journal of Personality and Social Psychology, 44,* 127-139.

Schulz, R., & Tompkins, C. A. (1990). Life events and changes in social relationships: Examples, mechanisms, and measurement. *Journal of Social and Clinical Psychology, 9,* 69-77.

Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127,* 1653-1658.

Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371-377.

Silverstein, A. B. (1982). Factor structure of the Wechsler Adult Intelligence Scale–Revised. *Journal of Consulting and Clinical Psychology, 50*, 661–664.

Snyder, D. (1979). Multidimensional assessment of marital satisfaction. *Journal of Marital and Family Therapy, 11*, 813–823.

Snyder, D., Wills, R. M., & Keiser, T. W. (1981). Empirical validation of the Marital Satisfaction Inventory: An actuarial approach. *Journal of Consulting and clinical Psychology, 49*, 262–268.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38*(1), 15–30.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-Trait Anxiety Inventory (STAI) test manual for form X*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1986). *The Structured Clinical Interview for DSM III-R–Patient version*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Sweeney, J. A., Clarkin, J. F., & Fitzgibbon, M. L. (1987). Current practice of psychological assessment. *Professional Psychology: Research and Practice, 18*, 377–380.

Wanberg, K. W., & Horn, J. L. (1983). Assessment of alcohol use with multidimensional concepts and measures. *American Psychologist, 38*, 1055–1069.

Wanberg, K. W., Horn, J. & Foster, F. (1977). A differential assessment model of alcoholism: The scales of the Alcohol Use Inventory. *Journal of Studies on Alcohol, 38*, 512–543.

Wechsler, D. A. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*, 87–96.

Wechsler, D. A. (1981). *Manual for the Wechsler Adult Intelligence Scale–Revised*. New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale–Revised manual*. New York: Psychological Corporation.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry, 33*, 111–115.

Zachary, R. A., Crumpton, E., & Spiegel, D. E. (1985). Estimating WAIS-R IQ from the Shipley Institute of Living Scale. *Journal of Clinical Psychology, 41*, 532–540.

Zimmerman, M. (1983). Methodological issues in the assessment of life events: A review of issues and research. *Clinical Psychology Review, 3*, 339–370.