Chapter 1

Introduction to Integrative Assessment of Adult Personality

Larry E. Beutler, Rita Rosner, Gary Groth-Marnat, T. Mark Harwood, and Hui-Qi Tong

This book is designed to help clinicians sharpen their skills in attributing useful meaning to discrepant sources of information, formulating valid opinions, and persuading others to comply with treatment recommendations. Specifically, these processes involve four steps: (1) identifying the problem, (2) selecting measurement instruments, (3) integrating sources of information about the problem, and (4) deriving and reporting conclusions, opinions, and recommendations.

In this volume we are concerned with how to address problems that are recurring and enduring. That is, this volume is focused on the assessment of personality and the use of that assessment to make relevant decisions regarding clients, with a major focus on developing helpful treatment plans. Thus, the four steps revolve around issues related to durable aspects of a person's behavior and applying corrective experiences to those behaviors that are problematic. To clarify further, this book is designed to help clinicians assess a patient's or client's personality in a way that makes a difference in how he or she is subsequently treated in a clinic, hospital, other mental health agency, or within his or her everyday context.

Personality is a social construct that refers to behaviors that are socially relevant, stable, have continuity over time, and are assumed to reflect certain motivational properties. Within a description of an individual's "personality" are dimensions and clusters of related dimensions that describe his or her predilections to behave, feel, and interact. All of these dimensions reflect, directly or indirectly, an aspect of social behavior. People are distinguished from one another, not so much by whether certain traits, such as anxiety, are present, but by the pattern of these traits. Most people have some level of most personality traits. Distinctions among people focus on the amount or intensity of the traits and the unique constellations of personality elements. Personality designations describe the relative strengths of different traits when compared to other people as well as to other traits. The pattern of behavioral traits—the relative strengths and weaknesses of the traits—forms a social profile that is used by others to describe, and distinguish among, people.

Clearly, not all traits fall within the domain that we call personality. Personality traits are assumed to evolve, at least partially, through a process of development, social learning, and choice. A person may have a trait of "baldness"; this is a distinctive quality that transcends situations, but it is devoid of the social behavior and choice aspects that characterize a personality trait.

Because choice or behavioral tendencies are implied, the traits of personality are assumed to reflect aspects of an individual's motivations and aspirations; however, these traits vary in the degree and the directness that they signify the presence of both motives and emotions. Some traits directly identify an assumed emotional quality. Thus a person may be characterized as a "happy," "anxious," or "moody" person when a particular temperament or pattern of temperaments is frequently present and manifest in a variety of situations. Other traits are described in less affective ways that more directly reflect social interactions. Thus a person may be said to be "gregarious," "extraverted," "introverted," "ascendant," "motivated," "intelligent," "outgoing," "reclusive," or "passive," reflecting different qualities of his or her social behavior, and by inference, social motives or aspirations.

Thus, critical to a quality being identified as a "personality trait" are both its reference to socially significant behavior and the assumption that the trait transcends any particular social environment. Simply put, traits are enduring qualities of behavior, distinguished from transitory and situation-specific states of being or feeling, that cannot be explained simply by the social or biological pressures of the moment. More specifically, the "behavior" to which these traits refer may involve an external act or an internal motivating process; however, one's perception of these external or internal pressures is, at least partially, dependent upon one's unique constellation of personality traits. When a person is described in the terms of personality, a picture is painted of his or her distinctive and volitional predispositions to act. To illustrate the differences among states and traits, on one hand, and personality and nonpersonality traits, on the other, try a thought game.

How would we describe the personalities of George W. Bush, Bill Clinton, and Barack Obama? In certain ways, these individuals are distinguished from other people. They are all male, and they have been or are a president of the United States. These are attributes and roles that endure, but neither being male nor being a past or current president can be defined as a personality characteristic. Being *male* is an attribute that is independent of learned behavior, and being a *president* reflects a consequence of social behavior but is not a specific social behavior in itself. Neither term, therefore, reflects identifiable social, volitional, or emotional qualities that are attributable to learning and development.

What personality qualities distinguish these three men? What are the enduring aspects of their social behavior, their dispositions to respond in certain patterned ways? How do these patterned behaviors distinguish one person from another? There are many differences among these three men, but which ones are relevant enough to select as "distinctive," and which are volitional and enduring enough to be called "personality"? To which one would we be most likely to apply the descriptors *nice, honest, bright, sneaky, dominant, strong, tender,* or *resourceful*?

We've heard Mr. Clinton described as "lecherous," Mr. Bush as "dumb," and Mr. Obama as "bigoted." Are these labels justified? Does it make a difference that their application is more a function of the party of the individuals using these terms than of specific

behavioral descriptions or scores on personality measures? One may see in these descriptors evidence that the terms that are used to describe people's personality frequently can mask strong opinions about attributes beyond their social behaviors.

In deciding that some particularly praiseworthy or derogatory terms describe presidents' personalities, we make some important, implicit assumptions. We assume that honesty, sneakiness, niceness, and bigotedness are qualities that these men carry with them from one situation to another. They are honest in all or most situations; they are sneaky under many conditions; they are nice in most things they do; and so on. That is, these behaviors are not situational; they are recognizable behaviors and behavioral dispositions that will be seen in many situations. We also assume that the behaviors referenced by these terms are meaningful and relevant, and that they relate to other characteristics, the totality of which represents the individual's personality. For what roles and tasks are these traits relevant? For what roles and tasks are they irrelevant? Which of these characteristics are associated with one another? Which are likely to change when the person is not in the role of president?

Psychological assessment is designed to answer such questions, as well as other questions, that have implications for clinical decisions. If an individual's personality describes a relatively unique pattern of interrelated response tendencies that cut across a large number of situations, there must be some way of describing these tendencies that is more specific than simply applying a categorical label, such as *weak, honest, extraverted*, and the like. Psychological "assessment" or "measurement" is the application of a system of classification or numbering to these qualities, in order to describe the patterns that exist among the qualities and the individual differences in people.

The objectives of psychological assessment in clinical settings involve answering questions that pertain to six clinically relevant domains of behavior: (1) the individual's diagnosis or disorder; (2) the etiology or causes of the disordered behavior; (3) the prognosis or anticipated course of this problem; (4) the nature of the treatments that may ameliorate or alter that course or prognosis; (5) the degree of functional impairment in both routine and specialized life functions; and (6) the person's pattern of strengths and adaptive capacities.

To answer questions relevant to these domains, we must be able to differentiate between behaviors that are a product of single situations (i.e., states) and those that are more general products of personalities (i.e., traits). We also must be able to distinguish between normal or usual behavior and abnormal or pathological behavior. Only by knowing what constitutes "usual" behaviors and "normal" responses to life's situations will a clinician be able to identify the nature and severity of behavioral disturbance and to assess the relevance of the measured behaviors for the questions asked.

WHAT ARE THE VALUE AND NATURE OF PSYCHOLOGICAL ASSESSMENT?

All kinds of people seek, or are referred for, psychological assessment. A troubled person may seek information about him- or herself in order to help him or her to solve a problem or change his or her feelings; a physician may seek information about a patient in order to identify what is causing a symptom; an adult child may seek information about a parent in order to understand a difficult behavior. Psychological assessment is the pro-

cess of discovering the social patterns in another person's behavior by (1) distinguishing situational from enduring behaviors, (2) predicting how the person will respond to different situations, and (3) finding an environmental event or treatment that will help or allow a person to change a troubling self-related aspect.

Unfortunately, the already complex process of constructing a helpful picture of a complex person is made all the more difficult by several common practices that are associated with psychological assessment. One of these practices is derived from the custom of Western societies to view various professionals as experts. This Westernized view of health and mental health professionals tends to perpetuate a fractionated view of patients and contradictory recommendations.

People who want to discover something about themselves or someone else have different ways of formulating the questions that they want answered. The way that they express a question is heavily determined by how the society in which they live has come to think about people who need help and those who are designated to serve them. Thus the questions raised about a person who is having difficulty are embedded within a cultural network that makes certain assumptions about people and about the nature of human behavior. Sometimes these cultural formulations are misleading, as can be seen in the way that psychological assessment is often used in medical or mental health settings by different professionals. The culture of modern, Westernized health care contains many different professions. These professions all claim to have some expertise about a portion of an individual's behavior. Thus, when a health care provider seeks to know something, he or she must decide which mental health professional has expertise in that particular area. Often this decision is tendered in such a way that it results in a fragmented view of people. It is not unlikely that a referring professional who wants a broadly based picture of a person will send a patient to a psychiatrist to get a "mental status examination"; to a psychologist to get a "personality," "mood," or "intellectual" evaluation; to a social worker to get a "social history"; and to a family practitioner to get a "medical history." Yet, each of these practitioners can provide a pretty comprehensive evaluation that includes more than these referral questions would suggest.

The terms used when professionals or others request consultation and assistance reflect the ways that people subdivide and categorize human experience and functioning. The labels that we use to describe the domain(s) of behavior we want assessed (1) reveal those aspects of behavior we consider to be relevant, and (2) imply that this domain is different and distinct from other aspects of psychological performance. Such distinctions among different domains of assessment are usually quite arbitrary, however. We believe that the implication that intelligence, mental state, and personality (among other constructs) are independent of one another creates an inaccurate, fractionated picture of a person. In the course of human functioning, cognitive and intellectual skills are not independent of personality; mental status is neither different nor legitimately dissociated from emotional or behavioral status; and so on. And although different professionals use different techniques and procedures, the actual evaluations that they conduct have many more elements in common than do elements that are restricted by the terms often used to define the referral question.

Nevertheless, because the different terms that have been used to describe assessment have been associated with different professional groups, it is not unusual for a client or patient to be referred to three or more different types of mental health professionals in order to obtain essentially the same service. By their unquestioned acceptance

of such referrals, referents or consultants implicitly endorse the view that social history can really be considered independently of personality; that mental state can really be considered separately from emotional life; and that personality can exist independently of intelligence or thought.

But, neither is it surprising that the recommendations and opinions of one professional may contradict that of other professionals. Indeed, such disparity among consultants is inevitable when members of each discipline frame their opinions from distinct methods, use different language expressions (jargon), and rely on theories to which the other professionals are neither privy nor considered to hold expertise. In reading different reports from different professionals, it is often not apparent that they are describing the same person. This variability of results perpetuates the myth that different domains of functioning are really being assessed by these different methods.

Kopta, Newman, McGovern, and Sandrock (1986) have demonstrated the profound differences that various theoretical frameworks and procedures can make in the nature and cost of recommended treatment. We need not invoke the story of the blind men and the elephant to see the problems in this picture. Reducing the fragmented and contradictory nature of our descriptions of people promises to lead us to improved efficiency and accuracy when we plan and implement treatments.

A fragmented picture of the client also frequently emerges because of the way in which psychologists choose to report their findings—typically by reporting the conclusions from one psychological test or one type of procedure after another, as if each were of equal value and provided an equally comprehensive and valid view of the person. When reports are written in this "test-oriented" manner, the inevitable contradictions in the findings of different tests are presented and ignored, omitted, or hastily excused in a summary paragraph. The failure to integrate and explain discrepancies in terms of person functions, rather than test functions, leaves the reader with a confused picture of the patient.

This book is written with the specific goal of eliminating, or at least reducing, this fragmented view of patients who undergo evaluation for behavioral and emotional problems. It is designed to help clinicians integrate information from different domains of patient experience and to make sense of the discrepant findings that exist among the sources of available information. We have undertaken three tasks in this book:

1. Provide an integrated model of personality and functional concepts that can be used to describe adult functioning efficiently and effectively within the context of responding to a wide variety of referral questions.
2. Outline a general method for resolving conflicts among data sources, while integrating and summarizing clinical information.
3. Provide representative and specific guidelines for extracting and integrating information from a few of the most frequently used and empirically defensible clinical assessment procedures.

The method we present for organizing and integrating sources of information is sufficiently general and flexible as to be applicable to many types of referral questions and assessment methods. Concomitantly, we selected the specific procedures illustrated not only because of their frequency of use among clinicians, but because they can be adapted to address many of the questions that referents are likely to have in mind when

they request cognitive assessment, personality assessment, mental status evaluation, and diagnostic testing.

SYSTEMATIC TREATMENT SELECTION: A MODEL OF FUNCTIONAL ELEMENTS

As clinicians who attempt to respond to requests for psychological assessment, we are faced with the dilemma of deciding which implicit or explicit theoretical perspective we will use when defining qualities of each person and transmitting conclusions. We might ask ourselves a number of questions:

> "Should I conceptualize and communicate my opinions using the language of unconscious processes or speak only descriptively, of interpersonal ones?"
> "Can or should I speak in a theory-free language?"
> "Should I implicate biological correlates of personality as possible etiological factors in symptom development or stick with situational and environmental ones?"

The professionals who refer patients come from many different theoretical perspectives. As behavioral consultants, we must be able to transmit findings that will make sense to referents or clients, from within their own theoretical perspectives; however, we often do not know what their perspectives are, and even if we do, we may not share their belief in those views. Resolving the dilemma of how to conceptualize and transmit our opinions is not as hard as it may initially appear if we make a simple decision to keep our observations (1) as close as possible to the actual, observable data; (2) largely descriptive rather than inferential; and (3) based on concepts whose relationship to the questions being addressed are supported by contemporary research findings.

Beutler and Clarkin (1990) originally developed a cross-theory model for assessing patient behavior that was specifically aimed at addressing the major questions typically encountered by psychological consultants and providing answers unfettered by theoretical jargon and concepts. They identified descriptive patient variables that are implicated in subsequent decisions about treatment, course, prognosis, and the like. This system has become known as *Systematic Treatment Selection*™ (STS™) because of its focus on treatment-related questions as a central perspective that binds together all of the usual referral questions. The STS conceptual system has been updated, refined, and empirically tested (Beutler, Clarkin, & Bongar, 2000; Beutler & Harwood, 2000; Beutler, Goodrich, Fisher, & Williams, 1999), a process that has resulted in the identification of a relatively small number of constructs and patient variables that can be assessed reliably and used in treatment decision making. The STS model spawned an assessment procedure, based on clinician ratings (Clinician Rating Form, CRF; Fisher, Beutler, & Williams, 1999a), which was refined through several iterations and finally adapted to administration through a patient-driven computer-interactive platform. More recently, this system has been revised and transferred to a patient self-report format, with substantial additions and changes made to the computer-based system. The changes include significant contributions by Prof. John Norcross, who provided self-help materials and references for the patient (*www.innerlife.com*; Beutler, Williams, & Norcross, 2008). This latter assessment system is described in Chapter 3; however, it is advantageous if the

reader becomes acquainted with the STS conceptual system at this point, because it is used throughout this volume as the foundational structure for interpreting psychological assessment procedures.

Levels of Decision Making

The process of making decisions in psychological assessment can be conceptualized as proceeding from very general to very specific decisions. At each level, different types of information about the patient are used to make decisions, to respond to referral questions, and to construct treatment plans and recommendations. In formulating the STS model, Beutler and Clarkin (1990) identified four basic levels of information that are implicated in these decisions: (1) patient predisposing factors, (2) treatment-related contextual factors, (3) therapy procedures and therapist relationship qualities, and (4) fit of patient and therapy. Table 1.1 presents a brief outline of these four levels and the informational variables that are considered at each level.

These different levels of information interact with one another as increasingly fine-tuned and complex decisions are addressed. Information from each of the four levels is useful only in relation to certain questions. It is the cascading influence of early levels on subsequent levels that allows us to address highly specific referral questions and refine specific treatment decisions (Beutler et al., 2000). For example, in order to address referral questions pertaining to patient diagnosis, level of functioning, or probable prognosis (level 1 questions), only information from the first level of information is needed. The therapist will need to know something about the patient's current and premorbid level of functioning: level of current and past social support; the complexity and chronicity of the problem; preferred way of coping or interacting with others; how and under what

TABLE 1.1. Levels of Systematic Treatment Selection

Patient predisposing factors

- Problem characteristics: symptoms, functional impairment, complexity/chronicity
- Personality traits: coping style, defensive traits, subjective distress, self-esteem, assets and strengths
- Environment: social support, social history, breadth of positive functioning

Treatment contexts (level of care)

- Setting: restrictiveness of care
- Intensity: frequency and length of treatment
- Mode: pharmacological, psychosocial, or both
- Format: medication or psychosocial treatment class (i.e., individual, group, family, couple therapy)

Therapy procedures and therapist relationship qualities

- Match of therapist and patient demographics and beliefs
- Therapeutic actions: directiveness, insight versus skill and symptom focus, cathartic versus supportive, therapist skill and experience
- Therapeutic alliance quality: collaboration and relationship strength

Fit of patient and therapy

- Fit of functional impairment and complexity with setting, modality, format, and treatment intensity
- Fit of coping style with insight orientation versus symptom/task focus of treatment
- Fit of trait-like resistance to therapist directiveness
- Fit of subjective distress with therapist support versus cathartic evocation

conditions he or she resists the influence of others; and the nature of current symptoms, abilities, and strengths, including level of current distress and history of resilience after crises. This information can be used to respond to many different questions as well as to predict the course of symptom development and resolution.

The second level addresses questions about the optimal level of care (intensity, setting, mode of treatment, etc.) needed by the patient. To respond to such questions, clinicians must make decisions about whether the patient's predisposing qualities fit the available treatment environments. Questions about the desirable length and frequency of treatment, the degree of protective control needed, the advantages of medication and multiple-person treatments, and the selection of the treatment setting, for example, can be answered only by knowing which qualities of the patient serve as indicators for these decisions, what resources and assets are available, and the particular patient's standing on these various factors.

Questions pertaining to the third level of the STS model include such areas as how the patient is likely to interact with health care providers, how to enhance patient cooperation, what treatment model is the best fit for this patient, and how best to develop a good working treatment relationship. Responding to these questions about which particular therapist to use, what model of therapy to apply, or how to alter the patient's prognosis requires information from prior levels as well as knowledge of available treatment options. Information about therapist skill and training, methods that facilitate motivation, the use of relationship enhancement procedures, and the like, is used at this level of decision making.

Finally, the formulation of fine-tuned recommendations about which specific procedures best fit this patient requires an understanding of the patient's predisposing qualities (level 1), the context of the treatment received (level 2), the nature of the available therapist's training and experience in using different treatments (level 3), and the way in which different patient qualities moderate the influence of therapeutic procedures (level 4). At this latter level, questions about differential treatment effects can be addressed by describing the nature of specific applications of procedure to fit patient variables.

PATIENT PREDISPOSING CHARACTERISTICS

Beutler et al. (2000) found that most of the questions posed to consulting psychologists can be addressed by beginning with the assessment of six general but fundamental qualities of patients, then adding to this information all knowledge about environmental demands and the nature of different treatment settings, treatment models, and treatment modes and formats. The clinician can construct responses to referral questions by reference to how these qualities fit certain treatment demands, availabilities, and functions.

Assessment of one's standing on six cross-theory qualities of patients serves as a beginning point for assessment. These to-be-assessed qualities include aspects both of the person and his or her social environment: (1) the areas of greatest impairment and strength; (2) levels of current and past social support; (3) the chronicity and complexity of the presenting problems; (4) typical and usual ways of coping with stress or resolving problems; (5) the degree of trust and compliance invested in interpersonal helping relationships; and (6) the level of stress and discomfort currently experienced. The

foregoing patient dimensions have all received empirical support in a large number of randomized clinical trials (Harwood & Beutler, 2008; Harwood & Beutler 2009a, 2009b; Beutler, Harwood, Kimpara, Verdirame, & Blau, 2011; Beutler, Harwood, Michelson, Song, & Holman, 2011).

Functional Impairment and Areas of Social Support and Achievement

Functional impairment (FI) was recently identified by the Division 29 Task Force as a viable participant-based factor in predicting treatment effectiveness (Beutler, Blatt, Alamohamed, Levy, & Angtuaco, 2006; Castonguay & Beutler, 2006a); however, FI has also been identified by the dysphoria work group of the Task Force on Empirically Based Principles of Therapeutic Change (Castonguay & Beutler, 2006b). Empirical support for the assessment of FI and application of this information in treatment planning are derived from more than 45 investigations (Beutler, Harwood, Alamohamed, & Malik, 2002). The combined sample size from these studies sums to more than 7,700 participants originating from both inpatient and outpatient settings. Almost all diagnostic categories were represented in these studies, and myriad types of psychosocial treatments, including a variety of treatment modes and formats, were administered (Harwood & Beutler, 2008; Harwood & Beutler, 2009a, 2009b).

Assessment of a patient's level of FI must include a determination of both deficits and areas of strength. The principal areas requiring assessment include levels of available support from others, access to social resources, the sense of attachment versus alienation from others, availability of social groups, proximity of family and friends, availability of role models, and the ability to maintain intimacy in relationship with others. These qualities are best assessed as objective qualities or external perspectives of behavior, rather than by an assessment of the person's subjective states (Strupp, Horowitz, & Lambert, 1997).

Assessing the positive presence of social attachments and personal resources, rather than just deficits and lack of attachments, is also important in this process. Thus level of social support is usually assessed in the course of assessing areas of impairment and strength. One's perceived level of a healthy social support system is a positive prognostic indicator in the STS model (Beutler et al., 2002; Beutler & Harwood, 2000). Treatment benefit corresponds with social support, suggesting that improvement in the quality of social support may also be a specific treatment factor (Castonguay & Beutler, 2006a). Indeed, social support has been identified as a patient prognostic indicator by the dysphoria work group of the Task Force on Empirically Based Principles of Therapeutic Change (Castonguay & Beutler, 2006c).

Lacking social support is very different from being actively abandoned or abused. Likewise, merely being in proximity to a friend or family member is significantly different from being actively engaged with other people, and having distal access to nonliquid financial resources is different from actually having money in hand.

The absence of social and financial resources, along with a sense of social isolation or alienation, are poor prognostic signs and suggest the need for enhancing the availability of these resources before other aspects of treatment can be expected to carry much weight. The presence of an intimate relationship, on the other hand, provides a degree of protection from the negative influences of the environment. All of these types of information are important in assessing a patient's level of social functioning.

Problem Complexity/Chronicity

The patient–treatment matching dimension of problem complexity and chronicity has received empirical support from at least 23 investigations (Beutler et al., 2002). The aggregated participant pool for these 23 investigations totaled nearly 2,000 and was comprised of populations from both inpatient and outpatient settings. Diagnoses varied across studies, as did treatment modalities and treatment formats. Relatedly, a variety of psychosocial and pharmacological treatments were examined in these investigations (Harwood & Beutler, 2008; Harwood & Beutler, 2009a, 2009b).

The recurrence and comorbidity of problems are critical in understanding and predicting the course of treatment and a given patient's prognosis (Beutler et al., 2000). The elements that comprise an assessment of these dimensions include (1) information about the recency and length of time during which the patient has had difficulties, (2) the degree of social disruption or stability in his or her life (e.g., maintenance of work and school roles) because of these problems, and (3) the variety of life roles affected by these difficulties.

Stated more simply, assessing the degree of problem complexity requires an understanding of the duration, relapse history, and comorbidity of the patient's presenting problem(s). These factors allow the clinician to estimate the degree to which the patient's problems impair multiple life roles (family, love, work, leisure). Notably, there are indices of positive patterns that tend to protect us from developing chronic and complex problems. An investment in leisure activities, the presence of at least one intimate tie, along with extended periods in which the patient has not manifested social and behavioral problems are positive factors that delimit the patient's problem complexity and improve his or her prognosis for recovery and maintenance.

In short, complexity/chronicity is typically assessed clinically and identified by patterns of recurrence, persistence of problems, comorbidity, and broad disturbances in interpersonal relationships (Beutler & Harwood, 2000). As a general rule, prognosis is attenuated by problem complexity/chronicity and by the absence of appropriate levels of patient distress (Beutler et al., 2000). Facilitating the development of social support increases the likelihood of positive change among patients with complex/chronic problems (Harwood & Williams, 2003).

Coping Style

Empirical support for the identification of a patient's dominant coping style and the application of this information in treatment selection is derived from at least 30 investigations involving participants from inpatient and outpatient settings (Beutler et al., 2000; Beutler, Harwood, Kimpara, et al., 2011). Aggregating the samples from the foregoing investigations brings the total number of participants to more than 5,600 individuals suffering from a wide variety of disorders. Depression was a primary diagnostic category represented; however, comorbidity was a prominent feature in many studies, with substance abuse, anxiety disorders, personality disorders, and severe psychopathology well represented among the samples.

Various treatment formats also were employed, including individual, group, and both married and cohabiting couples therapy (Harwood & Beutler, 2008; Harwood & Beutler, 2009a, 2009b). The formula we use to quantify one's dominant style of coping involves a ratio comprising several Minnesota Multiphasic Personality Inventory–2

(MMPI-2) clinical scales. More specifically, we use the following to construct an internalization ratio (IR):

$$IR = Hy + Pd + Pa + Ma/Hs + D + Pt + Si$$

An IR greater than 1.0 is considered high and characteristic of an externalizing coping style; the patient is likely to project his or her feelings and motives onto others and to be more undercontrolled and impulsive than the average patient. On the other hand, a ratio score that is less than 1.0 indicates that the patient is less impulsive, more self-critical, and more introspective than the average patient—his or her dominant style of coping is characterized as internalizing (Harwood & Beutler 2008; Harwood & Beutler, 2009a, 2009b).

When interacting with others, and especially when confronted with decisions, problems, and obstructions to their needs or goals, people engage in relatively consistent patterns of behavior that are designed to restore order and reduce discomfort. These characteristic styles of behaving are enduring, trait-like, social, and imply different objectives and motives. Thus they reflect aspects of personality. Efforts to cope with or resolve the problems that arise in daily life, including efforts to escape from stress, can be described as coping styles (Beutler, 1983). Simplistically stated, coping styles can be represented by two broad patterns: (1) attempting to resolve problems and escape pain by turning inwardly and thinking, or intentionally not thinking, about the problem, and (2) acting outwardly and directly, on or against the problem. Of course, most people do some of each. They think and they act. It is the relative balance of these two types of behaviors and the order of their use that define a person's coping style as either internalizing (a dominance of thinking and feeling) or externalizing (a dominance of acting).

People with an internalizing coping style may reason through a problem, question their own adequacy, insulate and remove themselves from the situation emotionally, divert their attention and thoughts to less threatening objects, or build defenses around their emotions to prevent their acknowledgment or expression. Any or all of these behaviors identify a process of looking inwardly for the solution to problems. Characteristically, internalizing includes, on the negative end of the continuum, self-blame, depression, emotional restriction, avoidance of others, social withdrawal, and exaggerated or restricted emotions; and on the more positive end, a search for internal resources. An externalizing coping style is associated with the relative reliance on behavior, rather than thoughts, reasoning, and feelings, to solve problems. Such people are typically described as "impulsive." An externalizing individual views the source of the problem as external; it is caused, controlled, and resolved by factors outside of the self. These individuals may place blame on others or the environment and feel victimized by fate or insidious intention. In response to these feelings, they may attack the source of frustration, become active in resolving the problem, get involved in an alternative activity, translate the problem explanation into some deficit of physical functioning, or impulsively flee the distressing environment (the "geographical cure").

These externalizing coping strategies are sometimes seen in people who have high needs for stimulation, those who do not carefully think through their decisions before acting, and those who have had problems with the law and authorities at school or work. These individuals are prone to (1) taking drugs to either escape or to keep themselves stimulated, (2) attacking and blaming other people for their problems, or (3) running

away. On the positive side, they also tend to take an active role in solving their problems and are not prone to fearful immobilization.

Interpersonal Compliance and Resistance

Reactance is frequently considered an extreme form of patient *resistance*—in this volume we use the terms interchangeably; however, the distinction is rather important in clinical practice. More specifically, it is often easier to elicit a negative/noncooperative response in a reactant patient than it is in a resistant one and, when elicited, the response tends to be stronger and more difficult to manage in those labeled high in reactance. Empirical support for the assessment of patient reactance level and the application of reactance-level information in treatment planning is derived from more than 30 investigations (Beutler et al., 2002). The aggregated participant pool from the aforementioned studies sums to more than 8,000 inpatients and outpatients.

The psychosocial treatment formats employed in the supporting investigations included individual, family, group, and parent training, with a variety of psychotherapies and pharmacological agents employed among myriad diagnostic categories (Harwood & Beutler, 2008; Harwood & Beutler, 2009a, 2009b).

One way to achieve what is needed to feel safe and to solve problems is to rely on those who seem to have more power and authority than we have. Compliance with a treatment recommendation or even entering treatment at all requires (1) some degree of interpersonal trust, (2) the ability to set aside ego issues, and (3) a willingness to comply with what the expert advises. At the other end of the spectrum of resistance, a person may (1) rebel against the influence of others, (2) act as if he or she can do everything unaided, and (3) even become oppositional when dealing with those who have either authority or control over him or her. Since such patterns tend to be repeated across situations, one's level of resistance or reactance reflects a quality of personality; however, most people are likely to resist when their behavioral options are eliminated. Thus situational demands also exert a significant influence in defining a person's level of resistance. Thus resistance is both a state and a personality trait.

Since most of the problems that cause people either to seek or be sent for help involve other people, it is very important to know the degree to which patients or clients are prone or likely to comply and cooperate with those who make suggestions or otherwise attempt to limit their choices. Individuals who are highly prone and vulnerable to being resistant tend to see others as unfriendly, controlling, and demanding. Compliant individuals, in contrast, may perceive the same people as being interested, friendly, and thoughtful. An assessment of these enduring perceptual dispositions can do much to help us predict how and if a person will be receptive to treatment or to the advice tendered by the psychological consultant.

Level of Subjective Distress

Empirical support for subjective distress is derived from at least 11 investigations that back the use of this patient dimension in both pretreatment planning and in the selection of specific interventions (Beutler et al., 2002). The combined sample size of these investigations, comprised of both inpatients and outpatients, totaled to more than 1,250. Various diagnoses were represented in these studies and several psychosocial and phar-

macological treatments were delivered as well (Harwood & Beutler, 2008; Harwood & Beutler, 2009a, 2009b).

Level of dysphoria or unhappiness is a rough indicator of the degree to which an individual perceives the presence of a problem. Discomfort often serves as a motivator for initiating treatment or some other action, and treatment itself may seek to (1) increase motivation by making a patient anxiously aware of his or her difficulty, or (2) reduce the distress to manageable levels. Thus subjective distress is an important aspect of patient functioning, and it is implicated in diagnostic, prognostic, functioning, and treatment decisions. Unlike level of functional impairment, however, this dimension must be assessed by accessing the patient's subjective state. Many patients may function quite well in their social world but still experience a good deal of anxiety about their performance, adequacy, and future.

Tapping both objective (functional impairment and resources) and subjective (dysphoria, euthymia, and felt support) dimensions of the individual's functional level is necessary for effective planning and decision making.

Applying Principles of Treatment Planning to Psychological Assessment

Obtaining multilevel sources of information has implications for how clinicians respond to the range of referral questions identified in this chapter. Beutler et al. (2000) have illustrated the nature of these dimensions by using them to formulate a set of cross-cutting principles that guide treatment decisions. These authors differentiate between basic and optimizing change principles—a distinction that reflects the specificity of the question being addressed in relation to the particular characteristics of one or more interpersonal relationships, especially the treatment relationship. Thus the basic principles can be used to guide assessment that addresses questions within the first two levels of information. Optimizing principles, on the other hand, guide recommendations about specific treatment relationships, the application of specific treatment models, and the fit of specific treatments to specific patients (differential treatment questions).

We identify specific principles that relate to treatment as we discuss various aspects of predicting patient responses to treatment, and we return to, and refine, these principles throughout this volume as we address various referral questions within the context of particular assessment procedures. Suffice it to point out that assessment methods vary in the type of information that they can usefully and validly assess. Thus the consulting psychologist often must select the methods used to emphasize these different levels of information and ensure that the information derived can be applied to the question(s) being asked. To know how to accomplish this task, the nature of psychological measurement must be clearly understood.

PSYCHOLOGICAL ASSESSMENT AND THE NATURE OF MEASUREMENT

Situational versus Trait-Like Responses

In clinical psychology and psychiatry, the basic task of measurement is translating the qualities that a patient brings to the clinician into a form that answers the questions posed by a referent. Psychological tests are measurement devices and methods that

are designed to accomplish this task. The use of modern psychological tests to provide answers to clinical questions and problems is a contemporary representation of a process that is as long as the history of humankind—the effort to identify the nature of individual differences and to account for both the similarities and uniqueness of each human's experiences. These efforts were, and are, embedded in the perennial attempts of peoples worldwide to predict and control their lives.

Throughout time, the speed at which people gained the abilities to predict and control events around them was governed by how well they overcame two major problems: (1) identifying those particular characteristics that are useful in describing the unique and similar qualities among people, and (2) distinguishing between the situational and personal contributors to (causes of) behavior. The first of these tasks underwrote the development of theories of personality and psychopathology, whereas the second underwrote the development of psychological tests.

The previous section of this chapter described the very basic list of concepts that our understanding of extant research suggests is minimally necessary in order to address the referral questions that are posed to psychological consultants. Now we face the task of identifying how we can move from making observations of a person's behavior to drawing a conclusion that this behavior represents certain distinguishing characteristics of the person rather than the situation in which he or she finds him- or herself.

In order to distinguish among situational and characterological (state vs. trait) aspects of behavior, it is helpful to maintain a constant test environment; doing so essentially eliminates, or at least holds constant, the influence of situational states. That is, across tasks, in a constant environment, clinicians may have some confidence that the resulting variations of behavior among participants or occasions are the result of participants' enduring, trait-like qualities. One of the major advantages of formal psychological measures over informal ones is the presence of standardized administration procedures. These procedures impose certain limitations and structures on a person's response, both by virtue of the explicit instructions used and by the implied expectations and consequences of the response. Such procedures are called "demand characteristics" of the testing environment. The responses to this structured and controlled assessment environment are compared to normative information based on a standardized sample. These norms tell us how people usually respond to these demand characteristics. Thus, by comparing a given person's response to these normative values, we can identify the degree to which the respondent's answers deviate from normal or average respondents. By inspecting the nature of the person's response, moreover, and comparing it to the nature of the demand characteristics of the instructions, we also can make some estimate of the degree to which this particular individual is sensitive to, and compliant with, the demands made by the environment. Departures from the expected response depict the ways in which characteristics of the respondent's beliefs and perceptions—qualities that are independent of the task itself—may alter the response to the demand characteristics of the task.

One way to view a response that departs from the established norms of a testing environment is to consider the variance as representing the insertion or modification of "attributed demand" to the demand characteristics. That is, the individual is responding not only to the environmentally imposed demands but also to some idiosyncratic, personal interpretation of these demands. Extreme departures from the conventional suggest the presence of attributed demands that are unconventional. We can see how

attributed demands may interfere with the environmental demand characteristics by viewing the two presentations that are illustrated in Figure 1.1. This figure shows two different ways that a respondent might react to the demand characteristics of the Bender Visual Motor Gestalt Test (BVMGT; Bender, 1938).

The instructions and physical properties of the BVMGT impose a demand characteristic that emphasizes accuracy and a distinct consideration of nine geometric figures. Each of the figures is presented on a small card, and the participant is instructed to "Draw each of the figures the best you can." There are no time constraints on the respondent's performance. The typical or normative response takes about 4 minutes and consists of systematically copying each of the nine figures, beginning in the upper left-hand side of the page and either proceeding down the left side of the page or across the top, separately constructing each figure. Thus the normative response is to reproduce the designs, one at a time, in an orderly sequence. Adherence to this normative response is seen in the first production (set A) shown in Figure 1.1.

The person who drew the second set of designs (set B) was responding to a different set of demand characteristics, demonstrated by this production's departure from the normative expectation. If the BVMGT was administered using the usual instructions—that is, if it imposed the usual demand for accuracy and separation—then we must assume that the respondent inserted some attributed demands into the process. The job of the clinician is to learn something about these attributed demands, and thereby about the respondent, by noting the degree to which the response departs from the norm as well as the nature of that departure. A clinician viewing this second production might conclude

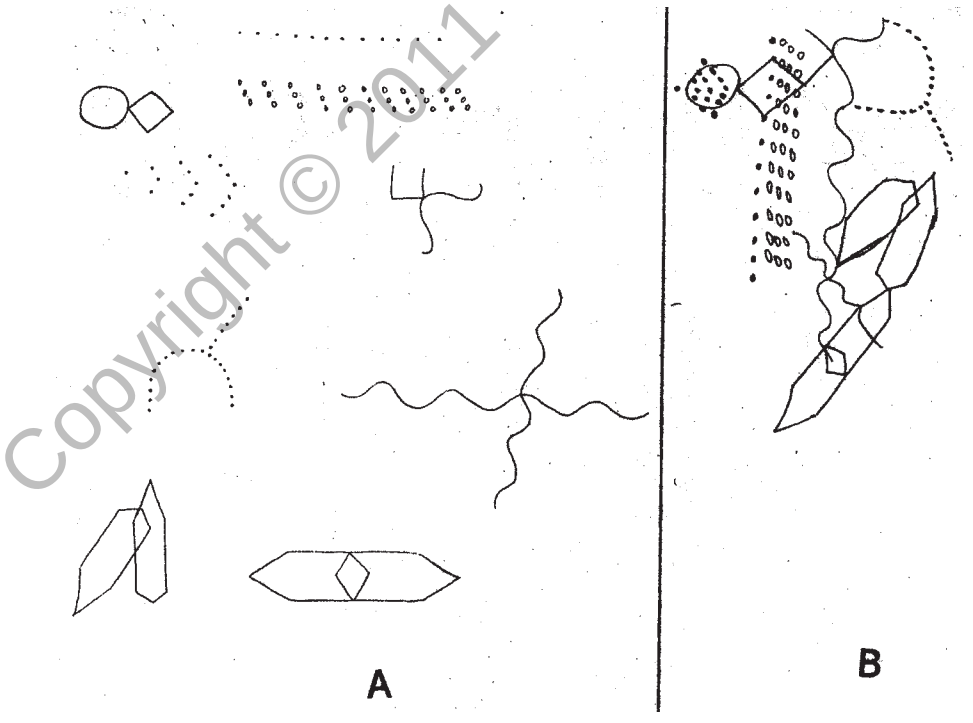

FIGURE 1.1. The Bender Visual Motor Gestalt Test under two demand sets.

that the patient's response is unconventional and poorly organized. It does not reveal a realistic appraisal of, or responsiveness to, the explicit demand characteristics contained in the instructions. In particular, the response shows little sensitivity to the demand of maintaining the boundaries among the separate stimuli. A clinician might conclude that the person who completed set B approached the task with an attitudinal response set that modified the explicit demands by adding attributed demands or expectations to the instructions.

These attributed demands were constructed from the respondent's beliefs and attitudes. In other words, this person did not realistically appraise the demands established by the examiner's request because of interference from internal expectations or trait-like qualities. It is these trait-like qualities that are revealed in the production, assuming that the explicit instructions remain constant. A person who changes the demand characteristics of the environment by some internal process of imposing new and idiosyncratic rules and assumptions may be resistant, schizophrenic, neurobiologically impaired—or exceptionally creative.

It is noteworthy that the responses in set B are not disorganized; they are simply organized according to a different set of principles. Indeed, the respondent organized the figures into a design. Perhaps the patient attributed a relational demand to the task that competed with, and overshadowed, the stimulus demand of separateness. Clearly, unless we know the explicit and stimulus-determined demand characteristics of the assessment environment, we cannot interpret the responses that come from psychological measures.

Classifying Behavior through Measurement

The results of psychological measurement, based on formal tests or informal clinical observations, are expressed as either categorical or dimensional classifications (see Kamphuis & Noordhof, 2009). An example of a categorical classification is the application of a psychiatric diagnosis. A person is assigned to a diagnostic class or type on the basis of a simple dichotomy of fit or no-fit with certain criteria. Dimensional assessment, on the other hand, assumes that certain qualities are best described as existing in some varying amount in most or all people. It is assumed, therefore, that qualities of this type are only roughly and inaccurately described by a categorical classification. Dimensional assessment is exemplified in quantitative estimates of the magnitude of such attributes as anger, depression, maladjustment, anxiety, neuroticism, extraversion, fear, and so forth.

Interpreting these scores forces us to grapple with the problem of translating them into concepts that have clinical meaning to others. Tests are administered and scored according to established rules and conventions; however, attributes such as problem-solving ability, anxiety, conflict, and personality are inferred, not observed directly. That is, they are hypothetical constructs whose existence can only be estimated from observed behaviors or reported experience. Their value is measured not only by how well they describe or predict behavior, but by how much consensual agreement exists in their meaning among those with whom clinicians communicate. This is why it is valuable to use terms that can be translated easily across various theories. The concepts we presented earlier are a beginning point for the transmission of cross-cutting knowledge. It is valuable to use these concepts and terms to communicate results in common language in order to maximize the usefulness of the communication process itself.

The tasks of integrating, ordering, and transmitting information from both informal and formal measurement procedures are considerably less systematic and technical than those of administering and scoring psychological tests. In this volume, we prefer to use the term psychological *assessment* rather than psychological *testing* to capture this broader clinical function. Psychological assessment includes the use of clinical skills beyond the mechanical administration of tests and computation of scores. The use of this terminology represents the recognition that the measurement instrument of greatest value, in the final analysis, is the clinician, not the test. The clinician who conducts a psychological assessment is a consultant, not a technician who merely administers a test; a consultant delivers opinions, not a summary of procedures.

The Process of Psychological Assessment

The requirements of psychological assessment can be illustrated by addressing the four steps of the assessment process identified at the beginning of this chapter. These steps require (1) an understanding of the social systems through which patients enter or seek assessment; (2) an understanding of the nature of measurement and familiarity with the measurement devices available; (3) knowledge of methods of interpreting these observations; and (4) familiarity with the process of communicating the findings and opinions to others. The four steps, again, are as follows:

1. Identifying the problem to be addressed.
2. Selecting and implementing methods for extracting the information needed.
3. Integrating sources of information around the problem.
4. Deriving and reporting conclusions, opinions, and recommendations.

The first of these steps is basic to the psychologist's role in many contexts; the second is the technical function of test selection, administration, and scoring. We assume that the advanced graduate students and professionals for whom this text is intended are familiar with basic psychometric principles and with the technical skills of responding to a referral request, selecting instruments, administering these tests, and scoring them. In this chapter we review only the basic principles associated with the first two steps. Chapter 2 considers the selection of specific tests. The rest of the book is devoted to the third and fourth steps, the advanced principles of formulating, integrating, and communicating opinions—the activities of the consulting psychologist.

Identifying the Problem

We have said that psychological testing is the application of measurement to the description of individual differences; however, in clinical practice, the nature of psychological assessment is more complex than this simple statement suggests. Good clinical assessment begins with translating requests for consultation into questions that can be meaningfully answered by clinical methods. Patients are referred to psychologists for many different reasons, not all of which are stated by the persons making the referrals. Moreover, not all of the ways in which people differ from one another are either likely to be of interest to a clinician or amenable to clinical assessment methods.

For example, let us return to the questions asked at the beginning of this chapter. To a political analyst, the most important distinctions among George W. Bush, Bill Clinton,

and Barrack Obama may be their political affiliations. To a working person, it may be their social background, to a member of a racial minority group, it may be their ethnicity.

Classifying these men by such attributes as political affiliation, social background, or ethnicity represents a method of categorical measurement and scoring. However, the resultant classification provides little help in deriving an answer to clinical questions. A mental health clinician may find that questions such as "Does any of these men have a clinical disorder?," "Are they depressed?," and "Does any of them pose a danger to himself or others?" are more relevant. That each of these men has been considered by some to be incompetent, bigoted, or dishonest all make these questions salient even for former and present presidents of the United States.

There are many responses to the claims of being a bigot, dishonest, or incompetent, but some of the related questions raised may have little relevance to whether or not they could discharge the functions of the office of president of the United States. A clinician who is asked to assess an individual's personality must begin by distinguishing between questions whose answers will shed light on the person's abilities to carry out responsibilities, tasks, and duties, and those whose answers will serve merely as distractions or matters of interest to others. To assess relevance, clinicians must first determine what question or questions are being asked by whoever is referring the "patient" to the psychological expert. In addition, clinicians also must gauge the consequences of various answers in relation to the person's level of functioning in various life and social roles. Only when clinicians know the likely implications of various answers can they judge (1) whether the questions themselves are relevant, and (2) if so, whether psychological assessment would be a useful method for deriving answers to these questions. Central concerns include (1) estimating whether the questions asked can be answered within the time allotted, (2) using methods that are available, and (3) ascertaining if the findings will be used in a way that is advantageous to the patient.

Clinicians must always be aware of who is requesting the information and the purpose to which the information will be put. Clinically relevant referral questions seek six general types of information:

1. Descriptions or formulations of the pattern of current behaviors
2. Causes of the behaviors observed
3. Changes that can be anticipated in these behaviors over time
4. Ways in which these patterns may be modified
5. Patterns and areas of deficit
6. Resources and strengths of the person

In other words, the questions address the objectives of determining diagnosis, etiology, prognosis, differential treatment, degree of functional impairment, and assets, respectively, as outlined at the beginning of this chapter. Diagnostic questions may be phrased as requests to rule in or rule out certain diagnoses, or they may ask how certain symptoms and behaviors are related to one another.

Questions of etiology may take the form of inquiring about whether or not traumatic brain damage is present or whether a patient's disturbed interpersonal relationships could be attributed to a recent loss or trauma. Both diagnostic and etiological questions seek to clarify the nature (e.g., interrelationship, severity) of problematic behaviors.

Questions about whether a given condition is likely to dissipate with time, or whether a given person is at risk for a future problem, are seeking prognostic data to facilitate the prediction of the normal course of change and development in various behaviors and symptoms. Questions of the fourth type, which solicit information to facilitate differential treatment planning, are related to prognostic issues and ask clinicians to anticipate what will happen with the patient's symptoms under certain imposed conditions (e.g., "Is this patient a good candidate for psychotherapy?"; "Should antidepressant or antipsychotic medication be used?"). In addition, some treatment questions are designed to prevent future problems (e.g., "Will education prevent this at-risk person from developing alcoholism?").

Questions about functional impairment may include inquiries regarding the patient's premorbid level of performance (e.g., "How much of this person's impairment predated the trauma?"), and those attempting to estimate some future level of performance (e.g., "What is this patient's employment potential?"; "What level of achievement can we expect of this individual?"). Questions of these types seek to determine (1) expectations others may reasonably hold for patients after their acute symptoms have dissipated, or (2) the cost factors associated with a disability.

Finally, some questions combine information about the patient, his or her environment, and the likely course of the patient's problems. Answering many of these questions may require clinicians to integrate information about the patient's assets and strengths. Thus the professional making the referral may ask the clinician for information on probable risk and protective factors, assets, and strengths on which to build an effective treatment program. No diagnostic, prognostic, or etiological picture of a patient is complete without an indication of strengths and assets.

When health care professionals seek consultation from other professionals, they frequently use shorthand communication methods, often without realizing that those with whom they are consulting must be familiar with these abbreviated communications in order to respond to them adequately. Hence a responding clinician must learn to distinguish the stated reasons for referring a patient from the unstated ones. Stated requests are often too general or too specific to allow the responding clinician to address adequately the covert or unstated needs.

For example, the most frequent requests from referring psychiatrists are couched in very broad terms such as "diagnostic testing" or "personality assessment," which are too general to be easily addressed. Such requests do not allow a psychologist to select an efficient way of responding. The request for nothing more specific than "diagnostic testing" could result in an 8-hour neuropsychological evaluation, the administration of 30 different projective tests, 15 hours of interviews and paper-and-pencil tests, and a two-night sleep and penile plethysmographic study—if all diagnoses described by the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; American Psychiatric Association, 1994) were to be considered systematically. These procedures not only are very expensive but constitute an inefficient use of time, because the referent usually has a narrower view of the most likely diagnostic options that he or she wishes to have considered.

On the other hand, some explicit requests are so specific that they do not allow the psychologist enough latitude to develop a reasoned response. A request for the MMPI or projective testing usually indicates a request for assistance in making a differential diagnosis; however, such specificity prevents the responding psychologist from selecting

the most useful measures for addressing this issue and precludes consideration of the concomitant influence of characteristics that may be obtained more reliably and validly from other methods. For example, restricting an assessment either to the MMPI-2 or to a projective method such as the Rorschach would be inadequate if the underlying goal is to find out how the patient functions in his or her family. These tests do not directly consider the family context. Moreover, even under the best of circumstances, a request for a specific test or type of test will be insufficient if test results are not considered in light of the patient's living circumstances and intellectual abilities, both of which must be ascertained through the use of other assessment methods. A given profile on the MMPI-2 will warrant very different interpretations if the patient has borderline intellectual abilities and is living in a group home than if the patient has superior intellectual resources and is living independently.

Reframing or translating the explicit request into a question that reflects the actual problem facing the referent simplifies the tasks of the consulting psychologist. Reframing a request either for "diagnostic testing" or for the "MMPI," for example, will probably result in an answerable question such as the following: "Is this person's depression of the unipolar or bipolar type?" Similarly, restating the request for "personality evaluation" or "projective testing" as an answerable question will probably result in something like this: "Is this patient able to cope with the stress of job loss without becoming psychotic?"

The first task of the clinician upon receiving a request for evaluation, therefore, is to contact the referent and discuss the request in sufficient detail that an answerable question emerges or can be developed. An answerable question is one that elicits concepts and issues that are within the domain of psychological experience and that can be measured. Good referral questions often pose hypothetical consequences for various potential answers. The answers obtained must possess the qualities of sensitivity (i.e., rightly identifying when the respondent is different from the norm) and specificity (i.e., rightly specifying when the respondent is similar to the norm). In order to translate explicit requests into questions that possess these qualities, the clinician usually needs to obtain information concerning the patient's background, current and anticipated treatments, and the time frame in which the answers are needed. For example, the clinician might ask the referent to (1) elaborate on the patient's current problem, (2) explain why he or she (the referent) thinks that psychological assessment will help, (3) specify what the clinician needs to find out, (4) describe how the information obtained will be used, and (5) indicate what decisions are pending, given the results. Background knowledge in normal development, abnormal psychology, developmental psychopathology, comparative treatments, and differential treatment efficacy will help the clinician frame questions to define the nature of the referral.

SELECTING AND IMPLEMENTING THE ASSESSMENT TOOLS

Fundamentally, psychological assessment boils down to the task of measuring and classifying observations. The processes of selecting and administering psychological tests are formalized extensions of what we all do in daily life. We meet a person at a cocktail party (identify the situational demands on behavior); observe the person as he or she interacts with us (observe samples of behavior); compare the person's response to that of others or to our prior experiences in similar situations (measure and compare); and conclude that

the person is likely to be friendly or unfriendly, likable or unlikable (generalize to unobserved or future situations). We have observed, measured, and classified (diagnosed) the person; queried about his or her history (explored etiology); developed expectations of future responses (determined prognosis); predicted how the person would respond to certain information about us (assessed a differential response to treatment); and drawn conclusions about the person's strengths and weaknesses (identified resources as well as functional impairments). In daily life, our safety and existence often depend on our ability to observe, measure, and classify accurately. If we perform these tasks poorly, we may be socially insensitive (inaccurate measurement), mistakenly assume that others will not hurt us (inaccurate prediction), or find ourselves becoming anxious when others' behavior changes abruptly (inaccurate generalization).

The distinctions between these day-to-day assessments and professional psychological assessment lie primarily in the degree of measurement precision used by clinicians and the theoretical origin of the constructs used to lend understanding, prediction, and control to the realm of behavioral events. A psychologist uses concepts founded on formal psychological theories rather than those contained only within privately derived theories. But, like the cocktail party observer, the clinician looks beyond each subject's responses to the nature of the situation in which the response occurs. Behaviors are judged within their context. All psychological assessment assumes that both the test environment and the associated behaviors constitute representative samples of external environments and of concomitant behaviors in these environments. Hence, it is assumed that the relationship between relevant test demands and resultant scores will be recapitulated on a magnified scale in an external environment. In other words, clinicians assume that the important elements of the testing environment correspond with similar elements of the real world, and that the symbolized meaning of test scores will be associated with a predictable set of behaviors within these real-world environments.

Because of the various ways they are structured, some tests are better suited than others to assess the domain of cognition; others are best suited to assess the domain of overt behavior; and still others are best suited to tap the domain of emotion. For example, the Wechsler Adult Intelligence Scale–IV (WAIS-IV) is a good instrument for assessing cognitive activities, but it is not well suited to determining the nature of the examinee's emotional experience; the MMPI-2 is well adapted to deriving information about behavioral traits but is less suited to understanding transitory, situational responses; mood scales are well constructed to evaluate various emotions but poorly suited to assess intellectual and cognitive abilities. Although domains of cognitive, affective, behavioral, and social experiences are not independent of one another, they have some unique qualities and vary in their salience from one environment to another. A psychologist must know both the domain of behavior that is best assessed by a given test and the nature of the environment to which that response domain is best generalized.

As noted earlier in this chapter, test environments are designed to place certain explicit and implicit demands and constraints on the patient's responses. These demand characteristics vary along at least three dimensions that parallel aspects of various external or real-world environments. Observing how patients respond to tests that embody the characteristics of different points along each of these dimensions has value for predicting the nature of behavior. Specifically, these test environments are designed to vary in the degree to which they (1) are structured or ambiguous, (2) attend to internal or external experience, and (3) place stress on the respondent.

Depending on the nature of the referral question, various aspects of these dimensions should be emphasized in the selection of tests. Ambiguity in an environment provides information about the respondent's ability to organize and interpret experience. Tests that vary in ambiguity may suggest something about the patient's ability to use cognitive resources, such as abstract and logical thought, to integrate experience. Likewise, observing how the patient responds to methods that focus variously on internal and external experience may provide information about his or her coping styles, level of impulsivity, vulnerability to threat, and accessibility to experience. This information may be important in addressing questions about intellectual abilities, personality disorders, the diagnosis of mood disorders, and suitability for insight-oriented versus behavior-oriented psychotherapies.

Finally, observing whether the patient responds with compliance, defiance, resistance, or decompensation to various levels of stress imposed in the testing environment may provide information about his or her stress tolerance, adequacy of protective defenses, resistance potential, and impulse control. Of course, in most instances, the questions asked are complex, and the patient's response requires the clinician to generalize to environments that vary in several or all of these qualities. Hence instruments are usually selected to permit the systematic observation of response variations at several points along each of these dimensions.

Methods of psychological assessment also differ in the sensitivity and accuracy with which they measure and predict behavior. The clinician's task is thus to select systematic methods for sampling the aspects of situations to which he or she wants to generalize, and to ensure that the behaviors observed in these situations are measured reliably and validly. In order to accomplish this latter task, the clinician must be familiar with good measurement procedures, including (1) the scaling methods used, (2) measurement sensitivity, (3) measurement specificity, (4) availability of normative data, (5) reliability of observations, and (6) validity of the observations. Each of these six areas deserves brief review.

Scaling Methods

Whether the goal is to assess the nature of the clinical question being asked, define the type of situation to which generalizations must occur, or measure an attribute such as intelligence or anxiety, the first and most fundamental quality of measurement is identity. Simply put, the measurement instrument—be it a clinician's judgment or test score—must translate samples of observed behaviors into a form that fairly represents distinctive qualities of individuals. The failure to identify or classify observations prevents accurate inferences from being made about past, present, or future behavior.

Measurement applies numbers to individuals or attributes as a means of establishing identity among observations. In increasing order of sophistication, there are four methods of preserving identity: nominal, ordinal, interval, and ratio. These four methods are often described as scaling methods because they order and classify observations. Nominal scaling assigns individuals or behaviors to categories. The remaining methods of assigning identity reflect variations of dimensional measurement.

The best example of nominal scaling, as applied to people, is our system of diagnostic categories. A DSM diagnosis of major depression identifies a cluster of related symptoms, differentiates those who have the condition from those who do not, and sug-

gests a particular course of development and treatment. Diagnostic labels define discrete categories or types of people, and have general application to a wide range of individuals who seek assistance from mental health practitioners. Diagnostic labels are limited in their value because they fail to make some important discriminations among those who meet the criteria for the diagnosis (i.e., those with major depression differ from one another in important, treatment-relevant ways), and they do not provide any information about the large number of individuals who fail to meet the criteria for a diagnosis but who still seek, and can benefit from, mental health services.

That is, nominal scaling methods, such as diagnosis, identify who has a condition but do not allow us to compare individuals either within or across groups. Using a nominal, diagnostic scale, for example, we could not say that "depression is more than schizophrenia" any more than we could say that "apples are more than oranges"—they are entirely different classes.

Ordinal scaling, on the other hand, is a measurement method that identifies the relative ranking of observations. We can say that depression is "more prevalent" than schizophrenia, that one person has "more depressive symptoms" than another, or that there are "more apples than oranges in Washington State." This ordinal or ranking method preserves the hierarchy that exists among the observations as well as the identity of each. However, although it allows the definition of a dimension, it does not tell us how much more frequently depression is observed than schizophrenia, how much more depressed one person is than another, or how many more apples than oranges are grown in Washington State. Doing any of the latter tasks requires either interval or ratio measurement. These latter scaling methods allow identity, ranking, and comparison.

In clinical assessment it is often important to determine either the category or diagnosis (nominal scale) and how many of the component symptoms are present or how severe they are (one of the three dimensional scales—ordinal, interval, or ratio) in absolute rather than relative terms. To do so, we must construct instruments that apply continuous ratings, in the form of numbers, to our observations. If we can assume that the differences between numbers are the same all along the continuum (the principle of equal intervals), then we can compare one score against another and conclude something about both the presence and the magnitude of observed differences.

Again, both interval and ratio scaling methods allow this latter type of magnitude comparison. Ratio measures can be applied only to characteristics that exist in a continuous quantity and do not exist at all (i.e., the scale has an absolute zero). Most psychological qualities do not possess both of these aspects. It is difficult to envision zero levels of anxiety or depression, for example. Unlike physical distance measures, where "0" means that two measurements are identical, it is not possible to measure most psychological properties by a ratio scale. Psychological characteristics are more similar to temperature than to physical distance; in the measurement of psychological qualities, as in that of temperature, a score of "0" is only one point along a scale in which lower scores are always possible. It is conceivable for someone to become even less depressed than a person who scores "0" on a test, much as temperatures can be measured below 0.

Sensitivity, Specificity, and Normative Value

Although necessary, identity as a property of a measurement scale is not sufficient for adequate assessment. If we reflect back on the question of how to describe the three

presidents, we can see that the categorical (nominal) identification of political alliance is of little help in assessing clinical referral questions, because it is not sufficiently sensitive to individual variations and is a poor predictor of the degree to which emotions and behaviors may depart from normal expectations. Both Democrats and Republicans can be emotionally healthy, disordered, or dangerous; not all Democrats are like Barrack Obama and not all Republicans are like George W. Bush.

Using the example of the U.S. presidents, we can illustrate three other important concepts in measurement: sensitivity, specificity, and normative value. In a way, we have already discussed the concept of sensitivity. A measurement possesses sensitivity if it can classify a person's uniqueness. That is, a sensitive measure is one that correctly identifies an individual as having a given characteristic or as being a member of a given group. Sensitivity is best understood when applied to categorical measurement and, in this case, is the percentage of true positives—that is, the percentage of time the measure identifies a quality as being present when, in fact, it is. Thus, the sensitivity of a measure is an estimate of the degree to which it is able to identify a condition as present, when it is, in fact, present.

To illustrate the concept of sensitivity, let us first imagine that we have constructed a self-report test consisting of a single question: "Have you ever been President of the United States?" If this test is then administered to George W. Bush, Bill Clinton, and Barack Obama, we may expect that all three will answer affirmatively. By checking their responses against public records for a normative base, we can then determine that we have successfully identified all three of these individuals accurately and that this quality is quite unique. They are true positives in that they not only have responded positively to the question, but they actually belong to the class of people defined by our historical criterion. Hence we can conclude that our test has high (even perfect) sensitivity.

The more politically minded may correctly point out that although all three men were indeed elected to the office of president, each can claim unique accomplishments. With a series of subsidiary questions, cross-referenced against historical records, we could develop three subscales for our test. Bush, but neither Clinton nor Obama, can be identified with a subscale that asks whether he "directed the invasion of Iraq"; Clinton, but neither Bush nor Obama, can be identified with a subscale that asks whether he secured a budgetary surplus; and Obama, but neither Bush nor Clinton, can be identified as having issued a proclamation to close Guantanamo Bay Prison. Thus a measurement system made up of these categories will still possess 100% sensitivity, in that all three former presidents can be accurately classified and distinguished from one another.

Now in terms of mental disorders or personality traits, the case is more difficult than it is with our former and current presidents. Whereas it is possible to check the presidents' answers against public records, we can check a patient's answers only against defined criteria, such as DSM-IV criteria. But the criteria themselves are only definitions and therefore dependent on cultural and historical norms. Examples of the relativity of definitions are labels such as "homosexuality," which was included as a descriptor of people in DSM-II, or "posttraumatic stress disorder," which was introduced in DSM-III after observation and acknowledgment of the psychological effects of participating in the Vietnam War. Although this caveat has only limited influence for the actual assessment procedure, it serves to underscore the fact that the certainty of constructs varies, and that caution regarding diagnostic labels is warranted, especially if the empirical support for the diagnostic construct is weak.

Although our presidential designation system possesses impeccable sensitivity, in that it accurately assigns each of the presidents to a categorical class of which he is the only member, psychological assessment requires that the measurement used also be capable of identifying those who do not belong to the targeted group. The ability to identify accurately those who do not have a certain quality or group membership is referred to as the scale's specificity. Making a determination of our test's specificity is impossible at this point because we have not yet tried it out on people who have not been President of the United States, have not secured a national budgetary surplus, have not waged war, and have not negotiated the termination of a prison for accused terrorists.

If we asked a large group of randomly selected people the four questions posed to the three presidents, we would find that all (or most) would say "no" to all of the questions. In checking the public records (our criteria), we would probably find that, in fact, none has actually been President of the United States, none has secured a national budgetary surplus, none has directed troops to invade Iraq, and none negotiated the closure of a prison for accused terrorists. Hence we could conclude that our test possesses the quality of specificity—it has successfully identified those in our sample who have not been President of the United States—as well as sensitivity.

Once we administer our test to such a large group, it gains some normative value. If we assume that the million people we have asked are representative of those in the United States, we can infer that most people would answer "no" to the questions, and that those who say "yes" will be unusual. But because there is so little variability in responses to our scale (1,000,000 people say "no" and only three people say "yes" to our questions), our scale does not allow us to say much of anything about the large number of people who have not been president. As this example illustrates, there must be both response variability and a normative value in order for the meaning of responses to be assessed.

To illustrate the importance of these concepts in a different way, consider the following: Say we hold up a pen and ask a classroom full of graduate students to identify what it is (a categorical rating). The characteristics of a pen are so constant and well known that there will be little variation among students' answers. Because all or most of the responses will be the same, we would be able to conclude little about these students beyond the probability that they are sensitive to their environments and familiar with pens; however, suppose one student says, "That is a prince who has been enchanted by a wicked witch." It is the departure of this student's response from the usual or normative response that would allow us to draw conclusions about how realistic his or her perceptions are. If the student comes from an unusual cultural background, in which witches and demons are believed to inhabit all objects, then his or her response may be seen as normal or usual within that particular culture, and our ability to interpret its unique meaning is lost. This illustration underlines the need to consider the meanings of responses in terms of the respondent's social norms and history.

Perhaps it is an unfortunate characteristic of psychological assessment that deviation from the norm is more informative than compliance with the usual. If we could rule out the possibility that this student's response is usual or normal within the cultural or religious environment in which he or she lives, by assessing a large number of people who are from the same culture, we could then conclude that the student's response represents some unusual characteristic of him or her. The more unusual the response, compared to the norm that represents the culture with whom the student identifies and in

which he or she lives, the more clearly we can conclude that a variant response indicates some form of clinical abnormality.

For example, suppose that our student looks frightened, jumps up, and runs out of the room when we hold up the pen. We may infer, with some degree of confidence, that the student is fearful of, and has negative attitudes toward, wicked witches, above and beyond his or her beliefs about pens. If he or she shares with the majority culture a primitive, animistic religion and background, then the unusual nature of this response may be assumed to reflect a deficit in the ability to objectively analyze, interpret, and respond to routine events; however, we can see that it is the deviation or variation of the response that gives us this ability, since we still can say little about the large number of students who have given the expected response, "pen." Even with perfect sensitivity and specificity, in other words, our "Pen Test" may have very limited value, because it only tells us something about those who deviate from the norm.

Since no one can be expected to be "average" in everything, we usually construct tests on which there are many ways to deviate from the average or norm. For example, in our illustration of the "President's Test," the large group of randomly selected individuals represents a normative sample because the characteristics of the sampled individuals are likely to be similar to those of the larger population. As in the case of the students in the "Pen Test" example, however, their responses do not distinguish them from one another. In response to the question about having been elected President of the United States, almost all of them have said "no." To be able to draw conclusions about individuals within this group, we must find ways in which their individuality is manifest. If we add an item to our test that asks, "How many people have financially benefited from your decisions during the past year?", we would obtain a number (i.e., a score) from each of our respondents that would manifest response variability. The arithmetic mean of these responses would provide a normative value against which we could compare our three former presidents and all others in the group, even without knowing the accuracy of their estimates. Moreover, the scores (i.e., the number of people who have benefited) among our sample of, say, 1,000,003 people would probably fall within a bell-shaped or normal distribution curve. Some people, like our presidents, will identify a large number of people as having benefited from their actions, whereas others will indicate that few or none has so benefited. Because our sample is both large and representative of the larger population, the distribution of scores is likely to be quite well distributed and characteristic of people in the United States. That is, the mean and distribution of the sample are likely to closely approximate what we would find if we were to ask this question of everyone in the United States.

After first computing the standard deviation of our sample, which is an estimate of how the responses are distributed (assuming that the scale measures a normally distributed characteristic), we can describe each individual within our sample by computing an effect-size score. This score simply describes, in decimal form, the number of standard deviations separating the individual from the mean of the sample. Because of their visibility and positions of power, it is likely that all three of the presidents in our example will be at wide variance with most of the rest of our sample. They will have highly positive effect-size scores (i.e., they will be several standard deviations above the mean) in the number of people who have financially benefited from their decisions. By inspecting these scores, we can compare the self-rated influence of any of the three presidents and that of any other person in our group.

Comparing individuals to normative standards based upon large numbers of randomly selected (i.e., representative) individuals does not help us understand either what caused any particular observed deviation or the accuracy of the scores given. The questions we still face include a determination of whether or not scores obtained in this way are likely to be accurate or clinically meaningful. Alternatively, do these scores vary as a function of some still unknown quality of the environment? To what degree is their accuracy influenced by momentary distraction? Do these scores indicate a stable aspect of respondents' personalities or intelligence? Are they likely to be influenced by current distress levels or impediments such as a recent bad night's sleep? In reference to our presidents, for example, are their estimates of the numbers of people affected by their decisions a product of their need to feel important, or do these estimates give an accurate indication of their influence? In other words, the measurement must be both reliable and valid.

Reliability and Validity

A measurement provides identity and sensitivity if it reflects the unique features of the patient's experience; it possesses specificity and normative value if it identifies the degree of similarity between an individual and others. The central purpose of psychological assessment is to gather information that allows us to generalize to situations that we cannot observe directly or that have not yet occurred.

We know that behavior does not occur in a vacuum; it arises in response to an environmental quality as well as to the individual qualities and characteristics of the person. Behaviors result from the combination of myriad unique elements, including the interplay among environment, personality, intelligence, presence and degree of pathology, and the *perceived* social pressures of the moment. Hence, if we are able to provide a constant (i.e., identical) environment for every person who completes our test, the resulting differences in behavior are likely to reflect individual qualities of personality, intellect, and expectations. Psychological tests attempt to provide such a constant environment for individuals, in order to permit testers to (1) observe the variations among their responses, and (2) infer the nature of each of their unique characteristics. We formalize the procedures of observation; study and standardize the environments in which a patient's behavior is sampled; and work to ensure that our instruments for observing and measuring are sensitive, specific, and have normative value. Thereby, our observations can be judged to represent samples of how individuals differ in their responses to environments.

The next task we encounter as clinicians requires us to distinguish between transient and enduring characteristics of people's behavior—states and traits. Mercurial behavioral characteristics, those that change from moment-to-moment or occasion-to-occasion, reflect state-like qualities and are judged to be attributes that are largely influenced by one's unique perception of the environment. Durable characteristics that remain constant over time and across situations indicate the presence of traits and are judged to be a personality quality that is minimally reactive to the environment. Situational anxiety is a state; eye color or avoidant personalities are traits. Qualities that have both state and trait properties lie somewhere in between, and each of these qualities may change at various rates and degrees in response to an individual's unique perception of environmental demands and social pressures.

Without knowing whether the behaviors we observe are likely to be enduring or situational, we do not know the degree to which our observations or the meanings of test scores can be generalized. The methods of classifying and measuring, in other words, must also possess the ability to be replicated; this is the quality of reliability. Reliability is an index of consistency or purity of measurement; it is usually expressed in the form of a correlation; however, because personal qualities vary in how much they are influenced and changed by the nature of the situation, different types of consistency or reliability are relevant for different measures. Test–retest reliability is indicated by high correspondence or similarity of responses on two different occasions. If our students say "pen" every time they are asked, we can infer that their familiarity with the object derives from enduring knowledge—a base of knowledge that supersedes changes occurring in the environment. Conversely, if their responses are strikingly different in two different situations or at two different times, we can conclude that whatever their responses are measuring is a temporary or passing state in their experience.

A measurement device with response variability allows us to estimate the likelihood that a given response will recur if the test is administered repeatedly on several different occasions. This estimate is based on the standard error of measurement, which is derived from knowledge of the reliability of the test responses. The higher the correlation between scores for a test on two separate occasions, the higher the reliability, and the smaller the error of measurement—that is, there are fewer unintended influences affecting the scores. The standard error of measurement, expressed as a standard deviation, is an estimate of variability in test scores likely to reflect the scores of a single individual if he or she took the test on many different occasions. It is used to estimate the possibility that the variation we observe in each of this individual's responses is an accidental occurrence. We can see in this example how test–retest reliability, as applied to dimensional measurement, is similar to the concept of sensitivity, as applied to categorical measurement. It is an estimate of how sensitive the test is to variations in the condition being assessed.

Another form of reliability is applied to a test when we want to assure ourselves that an entire test or subtest is measuring the same factor. For this purpose, we compute the test's internal consistency, which is an estimate of the degree to which each item or subpart of the test measures the same factor as the rest of the test items or subparts. Internal consistency is usually expressed as a correlation between the items and the total score and reflects how well items "hang together." With tests that are designed to measure several different attributes by way of subtests, internal consistency is estimated by the relationship of the items to the subtest scores rather than to a total score. It is expected that these part–whole correlations will be higher than the correlations between items and total scores on subtests, which are designed to measure an attribute differing from one to be measured by an individual item. Internal consistency is sample dependent, and high internal consistency may simply reflect redundancy in item content.

Equivalent-forms reliability is a method of assessing consistency that combines some of the principles of test–retest and internal consistency reliabilities. In this method, we may construct two forms of the test and compute the degree to which they measure the same factor, either when administered at the same time or when administered on two separate occasions. This form of reliability is used when there is some reason to believe that the act of responding to the test on one occasion will determine how a person responds on the second occasion. This concern arises when the response is affected

either by memory or by corrective knowledge that is gained while the person is taking the test. For example, a subtest from the Wechsler Memory Scale–IV (Pearson, 2009) presents paired words and then asks the participants to recall the second word of each pair when the first is repeated to him or her. This paired learning task is repeated three times in each administration, and it is likely that the learning that takes place will be carried over to another occasion. Therefore, a different but comparable list of words (an alternate form of the test) is used when the test is repeated, in order to avoid this problem.

Sometimes the test scores are subjective, as in the case when a clinician must rate clients' drawings or the meaning of behaviors. We want to assure ourselves that different raters make similar ratings (i.e., see similar meanings). Thus we may ask the raters to judge the amount of similarity and then compare the judges' ratings to derive an index of agreement between those ratings. We compute this index of agreement to support the notion that the same factor is being measured by each judge's rating. In such instances, the type of reliability that is desired is called *interrater* reliability.

Although high reliability estimates tell us that a similar quality is being measured, and although comparisons of different types of reliability indicate whether the quality measured is a stable quality of the person, the test, the situation, or the rater, they do not tell us what it is that we are measuring. The accuracy with which a test measures the factor or quality we want it to measure is called *validity*. Validity is the most basic and yet the most difficult criterion to meet in test construction. Because the concept or attribute we are addressing is usually abstract and subjective, there are generally no direct measures of its essence. Therefore, it is almost impossible to completely establish a test's validity. But by first identifying the particular type of validity that is of principal concern, and then applying some established procedures to the task of measuring this type of validity, we can obtain an estimate of a test's validity that is sufficient for our purposes.

The nature of validity, like that of reliability, varies because of the different purposes to which we want our test to apply. The main types are called *content, construct, criterion*, and *incremental* validity. In clinical appraisal, our desire to identify and distinguish between those behaviors that are situational versus those that endure across situations is often made more difficult by the various meanings of the words and labels we use to define characteristics to different people. In order to be useful, the terms we employ must have the same meaning across situations and cultures. The behaviors and acts that are called *aggression* or *sexual* in one culture must also be identifiable as such in another, even though both the normative levels and the acceptability of these behaviors may differ with the cultural values and norms. The political designations that define the presidents in our earlier example do not have these qualities; they are culture-specific and historically determined, and whatever attributes can be legitimately associated with them do not translate across cultures. The political platforms of the Christian Democrats in Italy, for example, may carry little similarity to the collective beliefs of a U.S. Democrat who is also a Christian. Nor is someone like Barrack Obama, who has been called a socialist by his critics, likely to be very similar to either a Greek Socialist Party member or a past head of the U.S.S.R. or China.

The task of establishing the meaningfulness of the content is central to the derivation of an assessment device's content validity. This form of validity deals with the subject of the test and is an effort to define the relevant aspects of the characteristic or construct that is being measured. To the degree that the items appear to the common observer to

be related to the quality that is targeted for measurement, we may say that the test has *face validity*; however, not all content validity relies on its apparent similarity to the targeted construct or quality. Sometimes the quality that we are measuring cannot be measured by obvious items alone. In order to keep the definition of various words constant and to ensure that the content is accurate in relation to our needs, we often define the meanings that we want for different terms by referring to a formal psychological theory.

Once the terms used in our test items are defined through either their face validity or theoretical content, their meaning must be operationalized. That is, their meaning must be identified in terms of some behavior that is observable to others. Since these terms are often derived from formal theories in the field, their translation into observable behaviors typically is based on the ratings of experts who are familiar with the theory from which the definitions and terms are being extracted. The importance of this point can be illustrated by our "President's Test." Let us assume that one of our items asks the former and current presidents to rate their success. The meaning of this term to the three men may be judged very differently if the term is extracted from economic theory and judged by economists; if it is drawn from a particular political party's platform and judged by the presidents themselves; or if it is borrowed from communication theory and judged by news correspondents. Therefore, the theoretical as well as the practical meanings of terms must be considered in assessing content validity.

Another form of validity that bears even more directly on the theoretical meaning of the qualities being assessed is *construct validity*, which refers to the degree to which the measurement device accurately identifies the presence of a quality or construct. Since constructs are theoretical rather than observable entities, construct validity is usually established by demonstrating that the measured trait or state bears the expected relationships to members of a network of other constructs within our chosen theory. The nature of these relationships is defined by the theory from which the construct has been extrapolated. For example, if our theory defines a president's success by how closely he or she follows a conservative agenda, then one measure of the construct validity of our test is how well it correlates with a measure of political conservatism. As is often true in the establishment of construct validity, scores on an established test are often used as a criterion in order to determine whether the same abstract constructs are also present in the new test. Sometimes this attribute is referred to as *convergent validity*, in that it is a demonstration that two tests converge or measure similar properties.

More than convergent validity is required in order to establish that a test has construct validity, however. In addition to demonstrating that the new test correlates with other tests that measure the same theoretically derived construct, a demonstration of construct validity also requires evidence that the new test does not correlate highly with tests that are designed to measure different constructs. This demonstration is called *discriminant validity*. For example, a person may score high on our test because he or she desires to appear conservative in an environment that is politically conservative. In that case, a portion of the test score may reflect the desire to fit in with others (i.e., social desirability responding) rather than either success or conservatism. If we can demonstrate that our test does not correlate with a measure of social desirability, we have adequately demonstrated its discriminant validity and provided more support for its construct validity.

To illustrate with a clinical example, scores on a particular test of depression should correlate highly with scores on other tests of depression, but they should not correlate

with a test of a supposedly different quality, such as anxiety. Unfortunately, this is a poor but important example of discriminant validity, because although depression and anxiety are theoretically distinctive concepts, few types of psychological measurement (including the ratings of clinical judges) can distinguish them. This fact illustrates an important "Catch-22" problem in demonstrating the validity of measurement: All of these validation estimates assume that it is already possible to measure the construct or concept under investigation. If we can already measure it, why develop the test in the first place? If we cannot measure it now, the new test cannot be shown to be valid.

This problem has led some theorists to suggest that only conceptual or face validity is necessary under most conditions. That is, the test is valid if it appears to be valid and if it is reliable. Alternatively, this problem points to the need for still another type of validity based upon some external criterion. *Criterion validity* is usually subdivided into two types, concurrent validity and predictive validity, depending on whether the test is expected to relate to external criteria that are present at the time the test is administered, or to those that are expected to occur at some time in the future. If our test of success, founded upon a conservative political theory, correlates with party affiliation, it may be said to have concurrent validity. If, on the other hand, it correlates with who wins the next presidential election, it may be said to have predictive validity.

The concepts of specificity and sensitivity are related to criterion validity. If a test of diagnosis is sensitive, it accurately identifies those who have the qualities that define the diagnosis—an external criterion. If it possesses specificity, it accurately identifies those who do not have the qualities that define the diagnosis—also an external criterion. Both of these examples of criterion validity are also examples of concurrent validity. A test that is able to predict the likelihood that a person will develop a set of symptoms meeting diagnostic criteria at some time in the future has predictive validity. More specifically, if a test of recurrent depression successfully predicts future depression, it may be said to have predictive validity. In clinical practice, the assessment of the course or differential treatment response of a patient relies on predictive validity. Indeed predictive validity may be the most important, but perhaps the most difficult, type of validity to demonstrate.

Finally, *incremental validity* is the demonstration that the test provides more substantial knowledge of, greater ability to predict the behavior of, or more accurate identification of individuals than is possible by using more easily obtained information. Thus, if a test predicts suicide risk with greater accuracy than clinical judgment, the test can be considered to have incremental validity. Whereas most forms of validity are expressed as correlations or accuracy ratings, incremental validity is usually expressed as a partial correlation—a correlation expressing the relationship between the test and a criterion, while the influence of other variables, or of prior knowledge, is held statistically constant.

SUMMARY

In this chapter we have briefly reviewed the nature of psychological tests, presented basic descriptive concepts from the STS approach (Beutler, Clarkin, & Bongar, 2000); introduced the patient dimensions that are related to treatment outcome in the STS procedure, and addressed two of the four tasks required of those who conduct psychological assessment: defining the referral question and selecting the measurement device. This

latter issue has been addressed only in terms of the qualities that are needed in order to measure accurately what we seek to measure. We have pointed out the difficulty of translating the shorthand requests for consultation that are frequently received into meaningful questions. We have also pointed out that psychological tests and measures must possess the qualities of scaling identity, sensitivity, specificity, normative value, reliability, and validity. Each of these concepts has been described and illustrated.

Although we have pointed out some of the considerations that are necessary in the selection of test procedures for answering referral questions, we have said little about the usefulness of specific tests, because their usefulness depends on the questions being asked. Specifically, we have pointed to the need to select tests that represent, in some significant and meaningful way, the environment to which generalizations are to be made and that evoke behaviors that are also representative of the behaviors likely to occur in these environments. The remainder of this book addresses the use and usefulness of several specific measures, the interpretive integration of these measures, and the communication of findings. Early chapters provide a conceptual description of the dimensions of the environment that must be considered in the selection of specific tests, and a system for organizing observations in preparation for rendering an integrated summary of findings from different sources and methods. Subsequent chapters describe several specific measures and their integrative interpretation. We wish to alert the reader to the fact that interpretations must be selected to address the referral questions asked. Because referral questions are so varied, the thoughtful clinician must use discretion and judgment when extracting and using the information gathered from specific tests for the purposes of responding to any given consultation request.

REFERENCES

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

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

Beutler, L. E. (1983). *Eclectic psychotherapy: A systematic approach*. New York: Pergamon Press.

Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology, 69*, 197–204.

Beutler, L. E., Blatt, S. J., Alamohamed, S., Levy, K. N., & Angtuaco, L. A. (2006). Participant factors in treating dysphoric disorders. In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 13–63). New York: Oxford University Press.

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

Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patients*. New York: Oxford University Press.

Beutler, L. E., Goodrich, G., Fisher, D., & Williams, O. B. (1999). Use of psychological tests/instruments for treatment planning. In M. E. Maruish (Ed.), *The use of psychological tests for treatment planning and outcome assessment* (2nd ed., pp. 81–113). Hillsdale, NJ: Erlbaum.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy: A practical guide to systematic treatment selection*. New York: Oxford University Press.

Beutler, L. E., Harwood, T. M., Alamohamed, S., & Malik, M. (2002). Functional impairment and coping style. In J. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient needs* (pp. 145–170). New York: Oxford University Press.

Beutler, L. E., Harwood, T. M., Kimpara, S., Verdirame, D., & Blau, K. (2011). Coping style. In J. C. Norcross (Ed.), *Relationships that work: Therapist contributions and responsiveness to patient needs* (2nd ed., pp. 336–353). New York: Oxford University Press.

Beutler, L. E., Harwood, T. M., Michelson, A., Song, X., & Holman, J. (2011). Resistance and reactance level. In J. C. Norcross (Ed.), *Relationships that work: Therapist contributions and responsiveness to patient needs* (2nd ed., pp. 261–278). New York: Oxford University Press.

Beutler, L. E., & Williams, O. B. (1999). *Systematic Treatment Selection (STS): A software package for treatment planning*. Ventura, CA: Center for Behavioral Health Technology.

Beutler, L. E., Williams, O. B., & Norcross, J. C. (2008). *Innerlife.com: A software package for treatment planning*. Available at *www.innerlife.com* and *www.webpsychcorp.com*.

Castonguay, L. G., & Beutler, L. E. (2006a). Principles of therapeutic change: A task force on participants, relationships, and technique factors. *Journal of Clinical Psychology, 62*, 631–638.

Castonguay, L. G., & Beutler, L. E. (2006b). Common and unique principles of therapeutic change: What do we know and what do we need to know? In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 353–369). New York: Oxford University Press.

Castonguay, L. G., & Beutler, L. E. (2006c). Therapeutic factors in dysphoric disorders. *Journal of Clinical Psychology, 62*, 639–647.

Fisher, D., Beutler, L. E., & William, O. B. (1999). Making assessment relevant to treatment planning: The STS Clinician Rating Form. *Journal of Clinical Psychology, 55*, 825–842.

Harwood, T. M., & Beutler, L. E. (2008). EVTs, EBPs, ESRs, and RIPs: Inspecting the varieties of research-based practices. In L. L'Abate (Ed.), *Toward a science of clinical psychology: Laboratory evaluations and interventions* (pp. 161–176). New York: Nova Science.

Harwood, T. M., & Beutler, L. E. (2009a). How to assess clients in pretreatment planning. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 643–666). New York: Oxford University Press.

Harwood, T. M., & Beutler, L. E. (2009b). Cognitive-behavioral therapy and psychotherapy integration. In K. S. Dobson (Ed.), *Handbook of cognitive-behavioral therapies* (3rd ed., pp. 94–130). New York: Guilford Press.

Harwood, T. M., & Williams, O. B. (2003). Identifying treatment relevant assessment: Systematic Treatment Selection. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (2nd ed., pp. 75–81). New York: Guilford Press.

Kamphuis, J. H., & Noordhof, A. (2009). On categorical dimensions in DSM-V: Cutting dimensions at useful points? *Psychological Assessment, 21*, 294–301.

Kopta, S. M., Newman, F. L., McGovern, M. P., & Sandrock, D. (1986). Psychotherapeutic orientations: A comparison of conceptualizations, interventions, and treatment plan costs. *Journal of Consulting and Clinical Psychology, 54*, 369–374.

Pearson, D. T. (2009). Wechsler Memory Scale—Fourth Edition administration and scoring manual. San Antonio, TX: Author.

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery*. Washington, DC: American Psychological Association.