

PART I

GENERAL ISSUES

COPYRIGHTED MATERIAL

CHAPTER 1

Overview of Behavioral Assessment with Adults

MAUREEN C. KENNY, KRISANN ALVAREZ,
BRAD C. DONOHUE, AND CHARLES B. WINICK

HISTORICAL DEVELOPMENT

Behavioral assessment stems from early research in classical and operant conditioning that demonstrated how behavioral and emotional responses could be conditioned, or in effect learned (Watson & Rayner, 1920). Treatments based on behavioral principles were then initiated to ameliorate various problem behaviors (Jones, 1924). Despite the increase in the use of behavioral therapy in the late 1950s, behavioral assessment measures were not commonly utilized in a clinical context until the mid-1960s. Early behavioral assessment focused on the observation of behavior patterns. Specifically, assessment focused on quantifying the frequency, rate, and duration of behaviors that were very specifically described to enhance reliability (Ullmann & Krasner, 1965). The publication of *Complex Human Behavior* (Staats, 1963) and *Behavior Analysis* (Kanfer & Saslow, 1965) increased interest in the field.

In the 1970s, behavioral assessment expanded in application and focus. The concept that situational variables influenced behavior led researchers to investigate individuals' behaviors within larger social contexts such as family, school, and business. In addition, assessment expanded to include feelings, sensations, internal imagery, cognitions, interpersonal relations, and psychophysiological functioning (Lazarus, 1973). To adequately assess these content areas, indirect measures, such as self-reports and ratings by significant others were included in behavioral assessments (Cone, 1977, 1978).

Diversification of behavioral assessment continued throughout the 1980s and 1990s with incorporation of other disciplines and traditions. Indeed, in a review of articles published in behavioral assessment journals in the early to mid-1980s, only 15% indicated observation as a mode of assessment (Fernandez-Ballesteros, 1988). Such diversity garnered interest in the topic and resulted in an increase in journal articles related to behavioral assessment. As a result, in the span of 3 years, articles containing the term "behavioral assessment" increased from 50 articles in 1980 to 200 articles in 1983

4 GENERAL ISSUES

(Fernandez-Ballesteros, 1993). Inclusion of other disciplines was evident in a review of *Behavior Therapy* and *Behavior Research and Therapy* from 1988 through 1991 in which 53% of articles detailed the utilization of standardized trait questionnaires (Haynes & Uchigakiuchi, 1993). Moreover, 82% of the articles that included trait measures did so in conjunction with traditional behavioral assessment methodologies. Along a slightly different vein, objective measures of behavior (e.g., urine drug screens to detect drug use) have become customarily administered with subjective measures of behavior (e.g., ratings of satisfaction with drug use) in controlled treatment outcome studies (e.g., Azrin et al., 2001). Thus, behavioral assessment has developed over multiple decades and continues to expand through the incorporation of traditional forms of assessment and application to various fields.

CONCEPTUAL FOUNDATIONS

CONTRASTS WITH TRADITIONAL ASSESSMENT

Despite recent incorporation of traditional assessment strategies in behavioral assessment, differences between behavioral and traditional assessment models lie within contrasting conceptual foundations. Traditional assessment including psychodynamic and personality approaches assume that behavior is the result of stable, internal, psychological processes. The assumption of this methodology is that behavior is the expression of enduring, underlying personality traits. Therefore, emphasis of traditional assessment is on the measurement of internal experiences and underlying traits (O'Brien & Haynes, 2005). The stable nature of these traits does not lend itself to modification through therapy. In addition, traditional assessment requires the acceptance of abstract constructs used to conceptualize underlying traits that are not amenable to empirical investigation. These limitations, among others, led to development of behavioral assessment strategies.

Perhaps the most distinctive trait of behavioral assessment is the emphasis on situational determinants of behavior and exclusion of trait assumptions. Behavioral assessment highlights the importance of environmental contributions and assumes that behavior is situation specific. Assessment in this tradition emphasizes the measurement of environmental factors, situation specific behavior, and recently has incorporated the examination of cognitive processes (O'Brien & Haynes, 2005). Behavior is considered to be initiated and maintained through a dynamic process and thus is assessed across time and environments. This dynamic process is assessed through the investigation of the antecedents and consequences of behavior. These variables include stimuli that precede, co-occur, and follow behaviors of interest.

The emphasis on situational determinants in behavioral assessment lends itself to sensitivity to individual differences. Behavioral assessment accepts that individuals differ in past experiences and present environmental variables and allows for individualized application of assessment strategies (O'Brien & Haynes, 2005). Individualized application can be utilized to better understand the factors initiating and maintaining behavior as similar behavior may be a function of differing contextual variables across individuals. Unlike traditional assessment, such specificity provides for direct therapeutic application. Identification of contributing variables allows for behavior modification through the manipulation of the environment (Nelson & Hayes, 1979).

Behavioral assessment is also distinct from traditional assessment in its emphasis on empirical evaluation. The focus of behavioral assessment on specific actions

and contextual variables is better suited for empirical testing than conceptual trait characteristics utilized in traditional assessment (Goldfried & Kent, 1972). Application of empirical evaluation in behavioral assessment involves specific definitions of behavior, monitoring of contextual variables, systematic observation, and the use of validated instruments. Several studies evaluating the reliability and validity of diagnoses, behavioral observation, and self-report inventories highlight the importance of empirical validation in behavioral assessment (Follette & Hayes, 1992).

BASIC ISSUES IN BEHAVIORAL ASSESSMENT

Assessment is a necessary first step in behavioral therapy. In many ways, it is the complementary partner of behavioral treatment. Behavioral assessment ideally occurs prior to commencing treatment to assess requisite goals for treatment, during treatment to assist in measuring therapeutic progress and adequacy of selected treatments, and after treatment to determine final outcomes and appropriateness of treatment termination.

Behavioral assessment is typically organized in three distinct, and increasingly molecular stages. The first stage, or level, of behavioral assessment is focused on the identification of global behaviors (e.g., noncompliance, sleep disturbance) that are most contributory to the presenting problem. These behaviors are particularly distressing, disruptive to daily functioning, socially unacceptable, or dangerous or detrimental to health and well being. The focus of assessment then shifts to determining the syndrome or cluster of symptoms that influence the onset of problem behaviors. Assessment concludes with an assessment of the system. This view holds that the individual is a system, whereby one symptom disturbance affects functioning in another area.

Although behavioral assessment and the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, text revision (*DSM-IV-TR*; American Psychiatric Association, 2000) began on much different tracks (the former growing out of behavior therapy and the latter linked to the medical model), they have developed a working relationship and can be used in conjunction with one another. For instance, behavioral assessment is often relied on to assist in diagnosing *DSM-IV* diagnoses, as these syndromes are often based on quantifiable actions that are quite amenable to behavioral assessment practices. The nosologic scheme of the *DSM* sets forth a descriptive set of behaviors that are necessary for diagnosis in a particular area. This is analogous to the syndrome level of assessment discussed previously; where by a cluster of symptoms (which delineate a diagnosis) are identified as targets of treatment. Additionally, the *DSM* list of psychosocial stressors is consistent with behaviorism's emphasis on environmental factors (Tyron, 1998). The atheoretical nature of the *DSM* does not preclude explanations of disorders based on behavioral theory. However, the *DSM* is based on a nomothetic system (classifying groups of similar individuals), while behavioral assessment is idiographic (unique to the individual) (Nelson-Gray & Paulson, 2004). A synthesis of behavioral assessment and *DSM*, where behavioral assessment is used to create a treatment that may be best for a specific individual, is possible.

Behavioral assessment may include many different instruments and types of assessment strategies. The domains that often need to be attended to when working with a client may best be assessed in different ways including self-report, self-monitoring and recording, surveys, direct behavioral recordings, and physiological assessment (Barlow, 2005). Generally, the start of an assessment identifies what behaviors are

6 GENERAL ISSUES

of interest. This usually involves some behavioral excess (e.g., crying all the time) or deficit (e.g., unable to speak in social situations) (Tyron, 1998). Sometimes direct observation of the client is essential, as this method is generally the most validated method of specified behaviors, particularly when motivation is lacking. However, this method is relatively costly as compared with subjective self-report measures (i.e., questionnaires and other retrospective rating by others). Moreover, as compared with direct observation, subjective measures are easier to implement (i.e., direct observation often requires the assessor to travel to naturalistic environments). Assessment in the natural environment is preferred, but not always feasible and so artificial laboratory testing circumstances may be used (Tyron, 1998). Physiological responses (which can impact target behavior) can also be measured directly, requiring relevant equipment and expertise. The various methods of behavioral assessment will be reviewed later in the chapter.

FUNCTIONAL ASSESSMENT

Functional assessment is often confused with *functional analysis* (O'Neill, 2005). However, these terms are distinct from one another (Gresham, Quinn, & Restori, 1999; Herner, 1994). Functional assessment refers to the utilization of methods and techniques to identify target behaviors, including stimuli that precede (i.e., antecedents) and follow (i.e., consequences, target behaviors); conducting a functional analysis; and developing treatment interventions. Functional analysis is limited to the process by which environmental events are manipulated to estimate their impact on target behaviors, and is, therefore, a component of functional assessment.

IDENTIFICATION OF TARGET BEHAVIORS

The first step in the functional assessment process is the identification of target behaviors. Generally, the clinician and client work together to identify target problem behaviors. This process begins by broadly identifying behavioral patterns, and later specifying these behavioral patterns into well-defined behaviors (Donohue, Ammerman, & Zelis, 1998), including the manner by which these behaviors are performed, as well as their frequency, duration, and severity (O'Neill, 2005).

FUNCTIONAL ANALYSIS

Functional analysis involves formulating hypotheses regarding the function, or purpose, of target behaviors (Gresham et al., 1999). Relationships between the variables that proceed, co-occur, and follow target behaviors are examined through the ABC model of functional analysis (i.e., antecedents, behaviors, consequences). An extension of this model, the SORC model includes antecedent stimuli (S), factors inherent to organism (O), target response (R), and consequences (C) (Goldfried & Sprafkin, 1976). This model includes current environmental stimuli and consequences, as well as individual physiology and learning history. The way in which this information is obtained and utilized is dependent on the method of functional analysis.

Indirect functional analysis involves utilization of interviewing procedures that are specifically designed to identify the function of relationships existing between target behaviors and relevant antecedents and consequences (Iwata & Worsdell, 2005). In this process, clinicians formulate hypotheses about the maintenance of target behaviors

through inquiry about antecedents and consequences. Behavior analysts may assess interviewee beliefs about how the behavior is maintained, how the problem may be solved, and successful and unsuccessful strategies that may have been used to cope with target behaviors. This method is easily implemented. However, verbal reports have been reported to be unreliable (Sturme, 1994).

Descriptive functional analysis involves systematic observation of target behavior to aid in hypothesis formulation (Sulzer-Azaroff & Mayer, 1977). Data obtained in this method is utilized in the formulation of hypotheses regarding the function of target behaviors. These hypotheses are confirmed or disconfirmed by examining the correlations between occurrence of target behaviors and antecedents and consequences. Descriptive functional analysis is a more reliable method than indirect functional analysis, as it does not rely upon self-report. However, this method relies on correlations and fails to provide a means of experimentally testing hypotheses (Iwata & Worsdell, 2005).

Experimental functional analysis is the method of experimentally manipulating the environment to develop and test hypotheses that are pertinent to the maintenance of target behaviors. Environmental events may be presented or withdrawn for short time periods to assess the affect on target behaviors (O'Neill, 2005). Behavior occurrence during these trials is compared to identify the function of the target behaviors and the maintaining variables (Iwata & Worsdell, 2005). Although this method is relatively time consuming, experimental manipulation offers great promise when accurate assessment is wanting.

TREATMENT INTERVENTIONS

Once target behaviors and their functional relationships have been identified, treatment intervention may be developed and implemented. When multiple target behaviors are identified, the clinician may choose to initially focus on a single behavior or address multiple behaviors concurrently. It is customary to initially target the behavior causing greatest distress to the client, thus assisting the client in being able to attend to other behaviors. Another approach is to first target the behavior most amenable to change so that the client is encouraged by the effect of treatment (O'Leary & Wilson, 1975). When the client lacks motivation, the clinician may choose to first address the behavior that is most disruptive to the client's significant others (Tharp & Wetzel, 1969). In addressing multiple behaviors concurrently, clinicians must examine whether behaviors are influenced by one another. Related behaviors are referred to as *response classes*. Response classes include behaviors that have seemingly different characteristics, but have similar effects on the environment or contextual variables (O'Neill, 2005).

SELF-REPORT ASSESSMENT

Information from self-report assessment strategies (e.g., questionnaires, self-monitoring) is obtained directly from the client. Self-report methods are cost-effective, easily administered, and amenable to the assessment of motoric, physiological, and cognitive processes (Cone, 1978). Self-report assessment generally requires less time and resources than other forms of assessment, such as behavioral observation. The primary criticism of self-reports is that client's may provide inaccurate information due to various factors, including intentional false reporting and memory deficits. However,

8 GENERAL ISSUES

Barlow, Hayes, and Nelson (1984) argued the client's perception of their problem is an important component of the therapeutic process, and Evans (1986) noted that the theoretical assumptions of behavioral assessment do not include unconscious processes that would lead to distortion of self-reports of behavior.

SELF-REPORT MEASURES/INSTRUMENTS

Self-report measures present clients with statements regarding the presence, frequency, or severity of their thoughts and behaviors. Measures differ in their number of items, but frequently utilize a Likert-type response scale (e.g., 7 = Extremely happy, 1 = Extremely unhappy). Standardized rating scales and problem behavior checklists are generally paper-and-pencil measures. However, some measures may be administered via computer. Computer-administered assessments often have the advantage of automated scoring, which can reduce the time of assessment interpretation. Although most measures are completed by the client, some measures may be completed by significant others or family members to gain a greater understanding of the problem.

Self-report measures vary in their degree of specificity. Some instruments assess multiple domains (e.g., depression, anxiety, conduct problems) broadly, while others assess specific behavioral categories more comprehensively. Determination of the type of measure to employ is dependent upon the objective of assessment. General measures are beneficial in the initial stages of assessment to identify the areas in which the client is experiencing greatest relative difficulty. Specific measures are useful when the clinician is interested in assessing detailed information about a client's identified problem and usually are administered once the initial target behaviors have been identified during broad-based assessment. Determination of which measure to use should include an examination of the reliability and validity of potential instruments because self-report inventories vary in degree of psychometric validation.

The Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1994) is a broad-based self-report checklist that assesses nine dimensions (i.e., Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism) and three global indices of distress (i.e., Global Severity Index, Positive Symptom Distress Index, and Positive Symptom Total). Clients respond to 90 items measuring symptom presence and severity on a 5-point rating scale ranging from "not at all" to "extremely." The SCL-90-R may be utilized to initially identify problems and measure treatment progress and outcomes. The SCL-90-R includes computer-based administration requiring approximately 12 to 15 minutes. The SCL-90-R has been reported to have above satisfactory ratings of internal consistency (Derogatis, Rickels, & Rock, 1976; Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988), and high levels of convergent-discriminant validity (see Derogatis, 1983). Most SCL-90-R item stems are behaviorally specified, (e.g., shouting or throwing things) although some are not amenable to functional analysis (e.g., never feeling close to another person).

Beck Depression Inventory II The Beck Depression Inventory Second Edition (BDI-II; Beck, Brown, & Steer, 1996a) is a specific self-report inventory used to assess the presence and severity of various symptoms of depression. The measure consists of 21 items; each item has four alternative responses from which the individual can choose, which are then scored, 0 to 3. The BDI-II is available in a paper-and-pencil

version. Total administration requires approximately 5 minutes and may be conducted by the clinician or self-administered by the client. The BDI-II has been reported to have good reliability, and high content and convergent validity (Beck, Brown, & Steer, 1996b). Although frequently utilized to measure depression, many of the items are designed to measure cognitive processes (e.g., I feel that I am a total failure as a person) limiting its application to traditional functional assessment.

SELF-MONITORING

Self-monitoring is the process of “systematically observing and recording aspects of one’s own behavior” (Bornstein, Hamilton, & Bornstein, 1986, p. 176). Functionally related internal and external environmental events may be recorded to gain better understanding of the client’s behavior (Cone, 1999). This form of self-report assessment is frequently used in clinical psychology. Indeed, in a survey of behavioral practitioners, 83% reported using self-monitoring with their clients (Elliott, Miltenberger, Kaster-Bundgaard, & Lumley, 1996).

Clients may use self-monitoring to track thoughts, feelings, and behaviors. Client recordings may include the frequency and duration of target behavior, as well as relevant antecedents and consequences to assist in understanding the function of behavior. Because clients are responsible for monitoring their behavior, the client and clinician should agree on operational definitions of the target behavior and relevant contextual variables to be monitored (Korotitsch & Nelson-Gray, 1999). Some authors reported finding increased accuracy when clients were trained in how to monitor behavior (Hamilton & Bornstein, 1977; Nelson, Lapinski, & Boykin, 1978). Thus, clients should have a thorough understanding of the recording process.

The way in which self-monitoring is structured may vary. Clients may record behavior at every occurrence, during a specific time period, or during specific situations. Barton, Blanchard, and Veazy (1999) recommend that the time period monitored by clients be specific to the monitored behavior and purpose of the data. Cone (1999) suggested that clinicians consider: (a) when and how frequent behavior should be monitored, (b) the time period for monitoring, (c) what is to be recorded, (d) the format of the record, (e) whether and how to use cues, (f) the number of different variables to be recorded, and (g) any techniques for assuring compliance.

Clients most often record their behavior on paper-and-pencil forms, although counters, tape recorders, and journals are also useful in the monitoring process. Paper-and-pencil forms and counters can quickly capture the frequency of behavior. Unlike counters, paper-and-pencil forms allow detailed information including severity ratings and contextual variables to be recorded. Tape recorders and journals offer the additional advantage of comprehensive verbatim responses, although much of the information recorded may be irrelevant to the treatment plan.

Self-monitoring may assist in the initial assessment process through identification of target behaviors and relevant contextual variables. Monitoring can also be used to establish baseline rates of behavior (Ciminero, Nelson, & Lipinski, 1977; Korotitsch & Nelson-Gray, 1999). This gathering of information further contributes to the development of a functional analysis through the formulation of hypotheses regarding target behavior and maintaining variables. When diagnostic classification is unclear, self-monitoring can aid in appropriate diagnosis. Treatment plans can then be developed accordingly (Korotitsch & Nelson-Gray, 1999).

10 GENERAL ISSUES

Utilization of self-monitoring during treatment assists clients in gaining increased awareness of behavior occurrence and behavior management (Korotitsch & Nelson-Gray, 1999; Wilson & Vitousek, 1999). Indeed, merely attending to the occurrence of behavior may alter behavior in the direction of the therapeutic goal (Baird & Nelson-Gray, 1999). Monitoring can also be used in treatment to monitor treatment gains and outcomes. Recordings throughout treatment can be compared to base rate information to assess behavior changes. If behavior change is not reflected in recordings, intervention strategies may be altered.

Self-monitoring is an inexpensive, convenient technique that requires less of the clinician's time and resources than other methods (e.g., behavioral observation). In addition to convenience, self-monitoring is less intrusive. Since clients record behaviors themselves, there is no need for the presence of observers, and recordings are less likely to be biased by observer perception. Clients are also benefited by being provided with a method of attending to their behavior and responses. Bornstein and colleagues (1986) noted the immediate self-derived feedback outside of therapy through self-reporting methods empowers clients. Self-monitoring also has a wide range of application.

Despite benefits of self-monitoring, the technique is not without criticism. Reactivity to monitoring may occur. Response reactivity occurs when clients purposely alter their report of behavior as a result of attention (Ciminero et al., 1977). Frequency or severity of problem behavior may be inflated to signal need for treatment (Craske & Tsao, 1999), or underreported if the client fears negative consequences for failing to reflect treatment gains. Contrarily, clients may underreport problem behavior to impress the therapist or "get out" of therapy commitments. Accuracy may also be affected by the limitations of memory. If the client records retrospectively, aspects of the monitored variables may not be remembered. Further, easily remembered events might be recorded as occurring more frequently (Craske & Tsao, 1999). To increase accuracy of data, observer load is an important consideration (Hayes & Cavior, 1977, 1980). Baird and Nelson-Gray (1999) suggest it is best to initiate the monitoring process with a few behaviors, and gradually increase the number of monitored behaviors with the passage of time.

BEHAVIORAL INTERVIEWS

Behavioral interviewing uses the framework of learning principles (i.e., operant and classical conditioning) to elicit information from the client, and initiate functional analysis of the problem (Sarwer & Sayers, 1998). The interview is often the first step in identifying the problem behavior(s), generating hypotheses, and gathering information (Glass & Arnkoff, 1989). Most clients are cooperative and provide detailed information regarding their problems and the antecedents and consequences. However, there are times when clients are not able to fully describe what is wrong. In instances such as these, it may be helpful for the clinician to ascertain motivations for the referral.

Interviews used in behavioral assessment vary in structure. The content and organization of unstructured interviews is dependent upon the clinician and the goal of the interview. Semi-structured and structured interviews are less flexible. Semi-structured interviews include standardized questions and formats, but do permit the clinician leeway to ask follow-up questions, select questions to ask from an available list, and permit idiosyncratic queries. Structured interviews are fully structured

interviews that clearly delineate the format of the interview and the questions to be read verbatim by the clinician. They do not allow rephrasing or additional questions to be posed to the client.

Unstructured Interviews Clinicians frequently use unstructured interviews at the outset of the assessment and therapeutic process to identify client's problem. The clinician works with the client to identify, and operationally define the problem. Lazarus (1973) proposed the BASIC ID approach to assessing content areas in interviews [i.e., behavior (B), affect (A), sensation (S), imagery (I), cognitions (C), interpersonal relationships (I), and possible drug utilization (D)]. Assessment of medical history is customary, as medical conditions may contribute to problem behavior. In keeping with behavioral assumptions, interviews are chiefly focused on current problem behavior, contextual variables, and treatment utility. The history of the problem may be assessed, but the emphasis is on recent events.

Kratochwill (1985) provides a four-stage format relevant to discussing topics during an unstructured interview:

1. Problem identification is the specification of problem to establish base rates and treatment goals.
2. Problem analysis is the assessment of contextual variables and client resources.
3. Plan implementation is the proposed data collection for evaluation of treatment progress.
4. Treatment evaluation concerns how pre- and postlevels of behavior will be compared.

Witt and Elliott (1983) suggested a more detailed organization for unstructured interviews:

1. The clinician begins by providing an overview of the interview agenda and a rationale for problem specification.
2. Clinician and client work together to identify and clearly define problem behaviors.
3. The problem behavior is further defined by assessing frequency, duration, and intensity.
4. Relevant antecedents and consequences are identified.
5. Realistic treatment goals and deadlines are discussed and agreed upon by the clinician and client.
6. The client's strengths are identified.
7. Any behavior to be recorded and the method of the recording are identified.
8. The way in which treatment efficacy will be assessed is discussed.
9. The clinician summarizes the information discussed and obtains agreement from the client.

These suggestions provide the clinician with a format for organizing unstructured interviews. However, clinicians have the freedom to vary the content of the interview, the specific questions posed, and the amount of time spent on each area. Unstructured interviews should be tailored to the client's needs and responses during the interview process. Through individualization of the interview, the clinician may obtain information from the client that would not have otherwise been disclosed.

12 GENERAL ISSUES

A major advantage of unstructured interviews is the flexibility to assess multiple areas and spontaneously alter the interview format based on information obtained. In addition to identifying problem behavior and contextual variables, unstructured interviews provide an opportunity for clinicians to formulate and inquire about hypotheses regarding the function of behavior. Depending on the information obtained, the clinician may determine whether further assessment measures are necessitated. Flexibility of the unstructured interview allows the clinician to explore client initiated content, and provide empathy, which may assist in establishing rapport. The interaction also provides an opportunity for the clinician to gain an impression of the client, and may serve as an observation measure if the client displays problem behavior during the interview.

Semi-Structured and Structured Interviews Semi-structured and structured interviews provide an organized approach to interviewing. Both forms of interview provide clinicians with a standardized set of questions, sequence for asking questions, and behavior ratings. The difference lies in the degree of structure. Semi-structured interviews allow clinicians to rephrase questions and ask additional questions for clarification. Structured interviews require clinicians to ask questions verbatim without deviation. This may be beneficial for interviewers with less experience and clinical skill, provided they receive training on the specific interview.

As compared with unstructured interviews, clinicians may achieve more accurate diagnoses through the use of semi-structured and structured interviews as they assure coverage of diagnostic criteria. Diagnostic accuracy is particularly important in research, and clinical settings for which accurate classification is crucial (i.e., forensic settings). Structured interviews are also used to verify diagnostic impressions gained during unstructured interviews. If clinicians are unsure of a client's diagnosis, they may use a semi-structured or structured interview to assist in differential diagnosis. The emphasis on diagnostic criteria, however, does not provide a thorough examination of contextual variables.

Although structured interviews have been criticized for failing to provide an opportunity to build rapport (Segal & Coolidge, 2003), they have demonstrated greater reliability and validity than unstructured interviews (Rogers, 2001; Segal & Coolidge, 2003). Reliability across interviewers is an important aspect of assessment when working with managed care systems that require diagnostic certainty and base reimbursement on diagnosis. Reliable measures also provide a more accurate measure of treatment outcome. However, it is important to note that psychometric properties vary across specific interviews, as indicated below.

STRUCTURED CLINICAL INTERVIEW FOR DSM-IV AXIS I DISORDERS The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) is a semi-structured interview for assessing *Diagnostic and Statistical Manual* disorders (DSM-IV) (American Psychiatric Association, 1994) Axis I disorders. Multiple versions of the interview are available depending on the type of client to be interviewed and the setting in which the interview is to be conducted (e.g., patient/nonpatient client, research/clinical setting). The clinical version (SCID-CV; First, Spitzer, Gibbon, & Williams, 1997) provides open-ended prompts with close-ended follow-up questions to assess both current (i.e., within the past month) and lifetime symptomology. The interview is comprised of 6 major modules: Mood Episodes, Psychotic Symptoms, Psychotic Disorders, Mood Disorders, Substance Use Disorders, and Anxiety and other Disorders, and requires 30 to 90 minutes for administration. However, the clinician

has the discretion to utilize all modules, or chose the modules that are most relevant to the client's presenting problem. The SCID-I has been used as the "gold standard" for validity (e.g., Shear, Greeno, & Kang, 2000; Steiner, Tebes, & Sledge, 1995). Reliability is generally good, and varies across modules (Zanarini, et al., 2000).

DIAGNOSTIC INTERVIEW SCHEDULE FOR DSM-IV The Diagnostic Interview Schedule (DIS-IV) for *DSM-IV* (Robins et al., 2000) is a fully structured interview for assessing *DSM-IV* (American Psychiatric Association, 1994) disorders. All questions and probes are provided, and presented verbatim. Questions are close-ended as a means of limiting variability. The DIS-IV requires 90 to 120 minutes to administer. Clinicians may choose which modules to administer, or utilize the optional termination points within each disorder section if the client does not endorse sufficient symptoms. Given the complexity of the interview, training is recommended.

DIRECT BEHAVIORAL OBSERVATION

Behavioral observation of specified behaviors is the hallmark of behavioral assessment (Algozzine, Konrad, & Test, 2005) and is the preferred method of obtaining information about problem behaviors (Groth-Marnat, 2003). Jordan and Franklin (2003) report behavioral observation is one of the most effective measures of frequency and duration of behavior. Although other methods are more commonly utilized, these methods are more likely to be influenced by reporting biases. The more removed from the time and place of the occurrence of the behavior, the more indirect the assessment and the greater chance of error. Behavioral observation should ideally occur in naturalistic environments (e.g., home, school, work), and should involve a range of appropriate and inappropriate social behavior (Algozzine et al., 2005). In conducting in vivo assessment the observer has little control over the behaviors that are observed. In vivo observations are most effective when assessing high frequency, global behaviors and are valuable for measuring change following clinical intervention (Groth-Marnat, 2003).

Defining the target behavior, in measurable, observable terms is a necessary first step to behavioral observation (Groth-Marnat, 2003). Any definition of a behavior should include specific examples of behavioral performance to assist in achieving sufficient reliability between raters. For some behaviors, such as measuring how many alcoholic beverages the client consumes (e.g., 12 ounces of beer, 5 ounces of wine), this is an easy process. Other behaviors, such as assertiveness, are more difficult and must be clearly defined (e.g., refusing an offer to ingest an alcoholic beverage). The target behaviors for observation will be derived from information obtained from the interview, anecdotal observations (e.g., observation of argument with spouse in waiting room), self-report inventories, and other assessment strategies (Groth-Marnat, 2003).

An observational assessment generally focuses on the here and now. Thus, the clinician's attention is focused on the immediate behavior, its antecedents (what happened just before the behavior) and its consequences (what happens right afterward). In conducting informal observations, an observer may observe behavior for a specific period of time and provide a narrative summary of relevant actions, often called a *narrative recording*. For example, a clinician working with an individual who self-reports social skill deficits may observe this person at a party. The clinician would then return to the office and make notes of observed actions. This type of observation is very helpful in defining more specific areas for future assessment to be measured

14 GENERAL ISSUES

quantitatively (Groth-Marnat, 2003). Formal observations identify previously defined behaviors. In a formal observation, the observer decides ahead of time who will be observed, what they will be doing, when and where the observation will take place, and how the behavior will be counted. The behavior must be operationalized prior to the observation so that the observer is able to measure it and keep a record. An example of a well-defined behavior would be, "asking a coworker how she is feeling during a work break," whereas a poorly defined behavior would be, "being friendly." Once the target behavior occurs, the clinician records its occurrence, along with relevant antecedents (e.g., presence of one person who smiles, solving a work-related problem successfully) and consequences (e.g., person returns salutation).

It is necessary to fully understand the various dimensions of behavior prior to conducting formal observations (Algozzine et al., 2005). For instance, *frequency* is the number of times an individual engages in the target behavior within a given time frame; the clinician must be clear about when the behavior begins and ends (e.g., Sam swore at his supervisee three times during the 8-hour work day). Rates per minute, hour or day can be derived from frequency data, thereby allowing comparisons across observations. Duration is how long the individual engages in the target behavior during the specified time period (e.g., Sam engaged in a conversation with his coworker for 2 minutes during the 1-hour lunch). Latency refers to the length of time that elapses between the antecedent stimulus and the onset of the behavior (e.g., Thirty minutes after Sam arrived home, he began crying). The *magnitude* of the behavior is the force, intensity, or severity of the behavior, but relies on a qualitative judgment (1 = Extremely angry, 5 = Extremely happy). What the behavior looks like motorically is the *topography* of behavior, while *locus* is where the behavior occurs.

PARTICIPANT AND NONPARTICIPANT OBSERVERS

O'Brien, Kaplar, and Haynes (2005) describe both nonparticipant and participant observers. Nonparticipant observers are generally individuals who are trained to record the occurrence of behaviors and contributing variables. These individuals are not, however, part of the client's environment. They receive formal training and are hired to conduct the observations and collect data on the problem behavior and causal variables. One downfall of this type of assessment is the cost associated with the training and hiring of such individuals. Participant observers are involved in the client's natural environment. They may include family members, coworkers, teachers, hospital staff, or other caregivers. Due to their regular involvement in the client's life, they are able to conduct observations in many different settings across substantial periods of time. There are drawbacks to utilizing these individuals. Given their multiple roles and responsibilities, they are generally able to record a limited number of events. In addition, their reports may be biased due to their relationship with the client, thus affecting objectivity of their recordings (O'Brien et al., 2005).

Direct behavioral observation can also be conducted with more than one observer at a time, making it a very effective measurement strategy (Jordan & Franklin, 2003). Two observers might be used in a residential or hospital setting. For instance, both a nurse and a mental health technician might be trained to observe a client's behavior, and would record their observations. Then, interobserver agreement could be obtained as a measure of reliability. The use of two observers is time consuming, expensive, and not always practical, usually demanding an assessment of interrater/interobserver reliability.

By-Products By-products occur when an observer codes stimuli that are indicators of the behavior of interest (e.g., urinalysis to assess cocaine use, bruises to assess perpetration of child maltreatment, movie ticket stubs to assess attendance at a movie; Jordan & Franklin, 2003). Advantages of this method include precise quantification, relative noninvasiveness, and ease of implementation. However, Bloom, Fischer and Orme (2003) note some problems with behavior byproducts including that the information is generally limited to quantity, many behaviors do not have by-products, and the clinician must be certain that the by-product is reflective of the target behavior. The validity of information gained from the collection of by-products is also questionable (e.g., tampered urinalysis, bruises reflect accidents rather than perpetration of abusive behavior, movie tickets are collected prior to movie) because it is sometimes difficult to verify.

Methods of Data Recording There are distinct methods of recording behavior (Algozzine et al., 2005). *Event recording*, the simplest and most efficient system, measures the occurrence of behavior. The observer must wait for the behavior to occur, and then record relevant information. This method is useful for aggressive actions, greetings, or verbal expressions (Groth-Marnat, 2003). It is especially effective when recording behaviors that have low frequencies, different types of behaviors, and when measuring change over long periods of time. However, it may be difficult to use this method for situations in which target behaviors have poorly defined beginnings and endings, and attention is difficult to sustain. *Interval recording* measures whether the behavior occurs within a predetermined time interval. This can be done in whole or part, meaning the interval can be separated into a specific number of time intervals (blocks of time) or the entire time period is used. For example, a period of 30 minutes can be divided into three equal, 10-minute intervals. Then the observer monitors the client to see if the target behavior occurs during that time interval (Jordan & Franklin, 2003).

Audio or video recording is also used as a method of observation that is less intrusive than an observer presenting in the room, and reduces the chances of reactivity from the client. Given the complexity of behavior, and the rapidness with which it may occur, video taping of behavior can allow for subsequent analysis without fear the observer "missing" important data. Sometimes, observers are not able to record behavior in the units of measurement that are needed (e.g., seconds) as accurately as an instrument can (Tyron, 1998), and instruments may be less invasive than observers. Technological advances have assisted observers by providing computers for recording purposes or digital-data storage devices which can provide counts of behavior and numerous calculations about rates, intervals, and reliability (Baer, Harrison, Fradenburg, Peterson, & Milla, 2005). However, the user of such equipment must be prepared for the rather lengthy dubbing process after the recording takes place and the possibility of equipment failure.

Analog Observation When direct observation is not possible, analog observations may occur. In these situations, a controlled clinical situation analogous to the natural environment is created so the target behavior may be observed (e.g., simulated bar to observe alcohol use behaviors). Indeed, there are many instances in which behaviors of interest may not occur in the presence of an observer (e.g., binge eating, fire setting, antisocial acts). Thus, analog assessment is one acceptable alternative. Importantly,

16 GENERAL ISSUES

the results of analog assessments may not be generalizable to the natural environment (Gresham et al., 1999).

Limitations There are some significant limitations associated with behavioral observation. The accuracy of observational data can be compromised if the target behavior is not clearly defined, the observer is not trained or objective, and the time intervals are not equal or numbers of response opportunities are not equal (Algozzine et al., 2005). Direct behavioral observation is expensive (O'Brien et al., 2005) and labor intensive (observers must be recruited, trained, supported, and the data entered and analyzed). Additionally, observer reactivity occurs when the presence of the observer can influence the behavior being observed (e.g., a conspicuous observer) or the client increases socially desirable behaviors while decreasing undesirable ones (Tyron, 1998). The observer may also have a lapse in attention and discuss the data with others (Groth-Marnat, 2003). Nevertheless, one of the chief advantages of behavioral observation is that the observer is permitted to examine behaviors that clients or significant others may be unable to report with sufficient detail or accuracy.

PSYCHOPHYSIOLOGICAL ASSESSMENT

Given the tradition of the "mind-body" connection, the use of instruments to assess physiological processes underlying psychological disorders is important. Physiological responses are activities of the body's muscles that can be measured with precision through technical and mechanical indicators of client behavior (Jordan & Franklin, 2003). For example, assessment of cardiovascular measures such as heart rate and blood pressure are important in the assessment and treatment of anxiety disorders (Sturgis & Gramling, 1998). Utilization of psychophysiological recordings permits direct, observable, and measurable data. Many psychological disorders are accompanied by physiological changes (sleep disturbances, gastrointestinal changes, heart rate, body temperature) that can be measured through instrumentation (Iverson, Stampfer, & Gaetz, 2002). The validity and reliability of such measurements is generally high. Much like behavioral assessment in general, psychophysiological measures can be used to initially assess the target problem, and later, to monitor outcome (Turpin, 1991).

The choice of which psychophysiological responses to measure depends on whether the target behavior is considered a primary defining characteristic or only part of the symptoms associated with the behavior (Sturgis & Gramling, 1998). For example, heart rate increase in anxious clients is only part of the symptom picture, while increased blood pressure in hypertensive clients is a defining feature of the disorder. Selection of the responses for measurement should also be guided by whether they involve the autonomic nervous system or the hypothalamus-pituitary-adrenal cortex system. The former relates to stimuli eliciting a startle response or requiring a physical or cognitive effort, while the latter is stimulated from more chronic stressors which are uncontrollable. The clinician must also consider the invasiveness, convenience, and acceptability of the measuring device (Sturgis & Gramling, 1998). Some commonly measured responses are heart rate and blood pressure, galvanic skin response (skin conductance), brain activity (electroencephalography; EEG); muscle response (electromyography; EMG), respiration, eye movement, and hand temperature. These measurements allow psychological problems to be translated into specific physiological indices (Groth-Marnat, 2003).

Anxiety disorders are one domain in which assessment of psychophysiological indices seems extremely important. Alpers, Wilhelm and Roth (2005) report that a comprehensive assessment of a client with anxiety requires both self-report and physiological assessment. Turpin (1991) states that adding psychophysiological techniques to "clinical assessment, particularly within the context of anxiety disorders, may yield unique information that might help to determine both the direction and effectiveness of therapeutic interventions" (p. 366). One such example is circadian heart patterns that have been found to be associated with a number of anxiety disorders (Iverson et al., 2002). These authors report that this method is a "noninvasive, inexpensive, minute-by-minute index of autonomic nervous system dysfunction" (p. 202). Mussgay and Ruddel (2004) suggest a thorough cardiovascular examination of all clients with anxiety disorders that may assist in devising a more complex treatment regime, including exercise. They performed a continuous assessment of EEG, blood pressure, breathing patterns and other cardiovascular indices that allowed the calculation of heart rate variances. Further, Alper et al. (2005) found that respiratory and autonomic system measures are valid diagnostic and treatment outcome criteria for phobic disorders.

Another area in which psychophysiological measurements are made is sexual disorders. For men, a penile strain gauge is used to obtain a direct measure of penile circumference. As the penis expands in the aroused state, the strain gauge records the changes on a polygraph (Barlow & Durand, 2005). This could be used to diagnose primary or secondary erectile dysfunction. Additionally, biofeedback measures are used with clients experiencing headaches.

Given that there are an increasing number of disorders that may require psychophysiological assessment, the laboratory is no longer the main setting for such research (Turpin, 1991). There are now ambulatory methods for psychophysiological recordings that allow clients to move about freely and can be used in the client's home or work setting, thus moving assessment to the real world and increasing its generalizability. Strides have been made to produce user-friendly software and hardware as well as compact computers that are available to most any clinician at a modest cost. As advances are made in ambulatory measuring devices, these would be of greater interest to clinicians who work outside of laboratory settings.

Despite its advantages, there are some limitations to physiological assessment. Skill, expertise, and training are essential for those who are conducting these assessments. At times, the measures produce inconsistent results due to procedural or technical difficulties or the nature of the behavior (Barlow & Durand, 2005). Given this, clinicians must be highly trained in the area in which they are obtaining these measurements. There is also often an adaptation that takes place when the client arrives at the clinic for psychophysiological assessment. Most measurements require that the client remain still (e.g., blood pressure measurement), which may be difficult, especially if the situation itself is anxiety provoking. The use of the instrumentation may also impact the physiological responses of the client. If the assessment requires multiple transducers, video cameras, and other devices, the greater the physiological response of the client (Sturgis & Gramling, 1998). Some clients may also demonstrate habituation responses (reduction in behavior following presentation of stimuli) over time. Finally, the instrumentation is generally recording physiological responses that are very minor and the equipment must be sensitive to recording only those responses that are meaningful and not extraneous.

18 GENERAL ISSUES

Reliability and validity are a concern with psychophysiological measurements also. To ensure not only the safety of the client, but to obtain reliable and valid measurements, proper preparation of the skin, electrode placement and proper care of the instruments is necessary. In a naturalistic setting where physiological assessment is taking place, it may be extremely difficult to filter out “noise” or sources of artifact. For example, temperature, and food and beverage intake may all affect psychophysiological measures and are not easily controlled. However, heart rate is one psychophysiological measurement that is highly reliable and easy to measure, even when a client is ambulatory (Sturgis & Gramling, 1998).

RELIABILITY AND VALIDITY

As mentioned previously in this chapter, behavioral assessment differs from traditional assessment; although it shares some of the same issues with respect to reliability and validity, it also possesses unique concerns in these areas. Reliability and validity are both fundamental to ensure accuracy of assessments. Potential sources of error, which affect reliability and validity, in turn could lead to biased case conceptualizations and interventions (Haynes, 2006).

The relationship between reliability and validity is complex. Reliability is necessary, but not a sufficient condition for validity. Best practice within behavioral assessment should aim at a balance between validity and reliability, with a priority given to validity while cognizant of the importance of reliability (Baer et al., 2005). If a measurement is reliable, but is not valid (not what the assessor intended to measure) then it has little value. Demonstrating reliability and validity of assessment data is important to ensure confidence that the results of the assessment have some degree of consistency (ability to be reproduced) while accurately represent the construct measured (behavior; Shriver, Anderson, & Proctor, 2001).

RELIABILITY

Reliability refers to the consistency or reproducibility of the results obtained from an assessment method. The ways in which reliability are established relate to determining how much measurement error is present in different conditions. Some of the standard methods of determining reliability used in traditional assessments are internal consistency, equivalent forms, test-retest, interrater, and standard error of measurement (Shriver et al., 2001). Reliability is related to the degree of confidence that similar results would be found over different occasions, raters or samples of the same behavior (Linn & Gronlund, 2000).

Behavioral assessment strives to take into consideration psychometric considerations, however, many concepts require different interpretations. For instance, normative comparisons are not as important in behavioral assessment as they are in traditional settings due to the emphasis of the former on a client’s own level of functioning (Groth-Marnat, 2003). Additionally, inconsistency and instability in scores, which traditionally would be viewed as affecting reliability, are interpreted differently by behavioral assessors (Silva, 1993). While many may report this as error in the observation system, it may be interpreted as instability of the behavior or true changes in the behavior (Haynes, 2006). However, having two raters (whose interrater reliability would be examined) would determine potential errors in the observation system. With regard to test-retest reliability, low test-retest reliability is likely to be interpreted

as due to environmental conditions rather than error during data collection (Groth-Marnat, 2003).

VALIDITY

Validity is defined as an evaluation of the adequacy and “the appropriateness of the interpretations of the results of an assessment procedure” (Linn & Gronlund, 2000, p. 75). Essentially this means the information gained from the test or assessment and inferences derived from there are appropriate, useful, and meaningful (Osterlind, 2006); the test or instrument measures what it is intended to measure. The validity of an assessment is not simply described as existing or not, rather it is based on a particular use and or purpose. Validity is best thought of in terms of degree, rather than all or none. The most common descriptions of validity use the terms, high, moderate, and low validity.

In traditional assessment, there has been a tripartite focus of validity: content, construct, and criterion related. However, these constructs may not be as applicable in behavioral assessment. Content validity refers to how well this assessment procedure or task represents the domain of behavior to be measured. Initially in behavioral assessment, content validity was approached informally by including in questionnaires and observational strategies what was rationally considered to be studied. For instance, a measure of depression would include statements that were believed to represent what symptoms most depressed people are experiencing (Groth-Marnat, 2003). Construct validity refers to “empirical evidence that supports the posited existence of a hypothetical construct and indicates that an assessment device, does in fact, measure that construct” (Popham, 2002, p. 363). This can be done in several ways including intervention studies, differential-population studies and related-measures studies (see Popham, 2002, for a complete review). Finally, criterion related validity refers to the relationship between the current measure and some future measure. Difficulties in this area have been similar to those raised in traditional assessment, namely, generalizability to different populations, settings and methods of administration (Groth-Marnat, 2003). Generalizability becomes an issue when behaviors that are observed in an office setting are representative of behaviors that would be displayed at home (Tyron, 1998). Behaviorists should obtain enough information across different settings in order to be considered valid and generalizable.

High face validity, sometimes thought to affect responding, can be an advantage in behavioral assessments. Face validity refers to the fact that “the appearance of a test seems to coincide with the use to which the test is being put” (Popham, 2002, p. 65). Knight and Godfrey (1995) report that behaviorally based memory tests that are high in face validity increase the likelihood that the client will be motivated to perform well and receive appropriate encouragement from others (caregivers, staff, clinicians) to perform their best. Face-valid measures are likely to be perceived as relevant to the client’s problem, as they use behaviorally referenced items related to the clients’ concerns.

Self-Report Historically, self-report measures, initially used to gain a direct report from the client, relied on content and face validity (Groth-Marnat, 2003). This method received much criticism for response bias, questionable reliability and validity, lack of norms, and client honesty in responding. Clients may under or over report behavior, be inaccurate in their own self recording of behavior, and fall prey to distortion when

20 GENERAL ISSUES

engaging in retrospective reporting (O'Donohue, Beitz, & Byrd, 2006). However, it may be better to think of self-report questionnaires as assessment tools, rather than psychometrically sound tests. To assist in conducting a thorough assessment, self-report measures should be administered with other measures from other sources, and processed through a lens of reliability and validity by the assessor. In fact, when multiple measures are applied reliability is increased regardless of the measurement device.

Behavioral Observation In direct observation, reliability is often reported as the percentage agreement between observers. Observer agreement tells us the degree to which two similarly trained observers record behavior. This can be achieved by having different observers independently observe and record the behavior and then compute the interobserver reliability (e.g., percentage agreement or nonagreement). Sometimes two observers may not agree on the presence or absence of the target behavior (Tyron, 1998). Baer et al. (2005) discuss one way to ensure that high reliabilities are not achieved by chance (high interrater reliability). They recommend checking to see how often the observers agree the target behavior occurred and how often it did not occur, respectively, occurrence reliability and nonoccurrence reliability.

Assuming reliability, validity across observers is a concern because not all observers possess the same threshold for observing the target behavior (Baer, et al., 2005). Ways to handle this are to more specifically define the topography of the behavior or create subcategories of behavior that are all recorded. Validity may also be comprised by observer distraction, so efforts should be made to create an environment that is minimally distracting for observers, while maintaining a typical environment for the behavior to occur. Additionally, a type of validity that may be unique to behavioral assessment is observer validity that is how accurately the observer records the behavior in a direct behavioral observation (Tyron, 1998). Finally, reactivity of the client can become a significant source of concern (O'Donohue et al., 2006). By virtue of being observed, clients may behave differently than if they were not. Indeed, it may result in genuine changes in behavior. Here the distinction between reliability and validity becomes clearer. The reliability of the information obtained by the observation is high, but the validity is questionable. The observer would be wise to not record the subject's responses and behaviors that appeared to be due to reactivity.

Reliability across different behavioral observational strategies varies based on a number of factors including observer expectancies, interference from nontarget persons, as well as a number of observer errors (i.e., halo effects, primacy effects, leniency in scoring). When there are highly structured procedures, reliability is increased. Understandably, observation conducted in a laboratory or controlled setting will have higher reliability than one conducted in a natural environment. There are ways for bias to be reduced and reliability to be increased, such as through highly structured procedures as systematic sampling (a procedure that allows each person in the population to have a known and equal probability of being selected).

To measure behaviors reliably, the definition of a behavior to record must be made clear to observers. For observable behaviors, staff who are conducting observations should be trained adequately to minimize the effects of reactivity of subjects, accurately record behavior and avoid interpretations, and leave the interpretations to the clinician. Measurement becomes much less reliable when dealing with internal cognitions where direct observation is not possible and the clinician relies on self-report by the client.

SUMMARY

Behavioral assessment is a psychological assessment paradigm that is based on theory, research, and practice. Founded in behavioral theory, over the past 20 years, there has been considerable research to support behavioral assessment methods. Behavioral assessment is concerned with the preceding events and resulting consequences of behavior, as well as ways to change these behaviors. Given the wide variety of treatment options available to clinicians, a thorough, broad-based assessment is a critical first step to ensure treatment success. There are several options available to the clinician desiring to conduct a behavioral assessment including functional assessment of behavior, self-report measures and instruments, interview techniques, direct behavioral observation, and psychophysiological recordings. Functional assessment includes identifying target behaviors, conducting a functional analysis, and developing treatment strategies. Self-report assessment obtains information directly from the client and includes interviews, self-report measures, self-monitoring and recording. Although there are multiple benefits to these methods, clients may not always provide accurate information on their behavior. Behavioral observation can be one of the most useful methods of assessment, due to its high validity and reliability. As with all forms of assessment, it has a number of limitations, including cost, observer bias, and necessary training. Some behavioral problems can be assessed through psychophysiological assessment that focuses on direct, observable data of physiological changes. Advanced methods of assessment including ambulatory recording devices make this type of assessment accessible to most clinicians. Psychometric concepts of traditional assessments, such as validity and reliability, have a distinctive role in and their relevance to behavioral assessment were discussed, with an emphasis on a balance between the two and a consideration of the unique issues inherent in behavioral assessment. Behavioral assessment is important before, during, and after intervention.

REFERENCES

- Algozzine, R., Konrad, M., & Test, W. (2005). Direct observation. In M. Hersen, G. Sugai, & R. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy: Vol. 3. Educational applications* (pp. 1266–1269). Thousand Oaks, CA: Sage.
- Alpers, G. W., Wilhelm, F. H., & Roth, W. T. (2005). Psychophysiological assessment during exposure in driving phobic patients. *Journal of Abnormal Psychology, 114*, 126–139.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Azrin, N. H., Donohue, B., Teichner, G., Crum, T., Howell, J., & DeCato, L. (2001). A controlled evaluation and description of individual-cognitive problem solving and family-behavioral therapies in conduct-disordered and substance dependent youth. *Journal of Child and Adolescent Substance Abuse, 11*, 1–43.
- Baer, D., Harrison, R., Fradenburg, L., Peterson, D., & Milla, S. (2005). Some pragmatics in the valid and reliable recording of directly observed behavior. *Research on Social Work, 15*(6), 440–451.
- Baird, S., & Nelson-Gray, R. O. (1999). Direct observation and selfmonitoring. In S. C. Hayes, D. H. Barlow, & R. O. Nelson-Gray (Eds.), *The scientist practitioner: Research and accountability in the age of managed care* (2nd ed., pp. 353–386). Boston: Allyn & Bacon.

22 GENERAL ISSUES

- Barlow, D. H. (2005). What's new about evidence-based assessment? *Psychological Assessment, 17*, 308–311.
- Barlow, D. H., & Durand, M. (2005). *Abnormal psychology: An integrative approach*. Belmont, CA: Wadsworth/Thompson Learning.
- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings*. New York: Pergamon Press.
- Barton, K. A., Blanchard, E. B., & Veazy, C. (1999). Self-monitoring as an assessment strategy in behavioral medicine. *Psychological Assessment, 11*, 490–497.
- Beck, A. T., Brown, G., & Steer, R. A. (1996a). *Beck Depression Inventory II Manual*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Brown, G., & Steer, R. A. (1996b). *Beck Depression Inventory II (BDI-II)*. San Antonio, TX: Psychological Corporation.
- Bloom, M., Fischer, J., & Orme, J. (2003). *Evaluating practice: Guidelines for the accountable professional*. Boston: Allyn & Bacon.
- Bornstein, P. H., Hamilton, S. B., & Bornstein, M. T. (1986). Self-monitoring procedures. In A. R. Ciminero, C. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 176–222). New York: Wiley.
- Ciminero, A. R., Nelson, R. O., & Lipinski, D. P. (1977). Self-monitoring procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 195–232). New York: Wiley.
- Cone, J. D. (1977). The relevance of reliability and validity for behavioral assessment. *Behavior Therapy, 8*, 411–426.
- Cone, J. D. (1978). The behavioral assessment grid (BAG): A conceptual framework and taxonomy. *Behavior Therapy, 9*, 882–888.
- Cone, J. D. (1999). Introduction to the special section on self-monitoring: A major assessment method in clinical psychology. *Psychological Assessment, 11*, 411–414.
- Craske, M. G., & Tsao, J. C. I. (1999). Self-monitoring with panic and anxiety disorders. *Psychological Assessment, 11*, 466–479.
- Derogatis, L. R. (1983). *The SCL-90-R Manual II: Administration, scoring, and procedures*. Towson, MD: Clinical Psychometric Research.
- Derogatis, L. R. (1994). *SCL-90-R: Symptom Checklist-90-R*. Minneapolis, MN: National Computer Systems.
- Derogatis, L. R., Rickels, K., & Rock, A. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry, 128*, 280–289.
- Donohue, B., Ammerman, R., & Zelis, K. (1998). Child abuse and neglect. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 183–202). New York: Plenum Press.
- Elliott, A. J., Miltenberger, R. G., Kaster-Bundgaard, J., & Lumley, V. (1996). A national survey of assessment and therapy used by behavior therapists. *Cognitive and Behavioral Practice, 3*, 107–125.
- Evans, I. M. (1986). Response structure and the triple-response-mode concept. In R. O. Nelson & S. C. Hayes (Eds.), *Conceptual foundations of behavioral assessment* (pp. 131–155). New York: Guilford Press.
- Fernandez-Ballesteros, R. (1988). *Que escriben los evaluadores conductuales?* [english translation] Unpublished manuscript, Universidad Autonoma de Madrid.
- Fernandez-Ballesteros, R. (1993). Behavioral assessment: Dying, vanishing or still running. *European Journal of Psychological Assessment, 9*, 159–174.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders-Clinician Version (SCID-CV)*. Washington, DC: American Psychiatric Press.

- Follette, W. C., & Hayes, S. C. (1992). Behavioral assessment in the DSM era. *Behavioral Assessment, 14*(3/4), 293–295.
- Glass, C. R., & Arnkoff, B. D. (1989). Behavioral assessment of social anxiety and social phobia. *Clinical Psychology Review, 9*, 75–90.
- Goldfried, M. R., & Kent, R. N. (1972). Traditional versus behavioral assessment: A comparison of methodological and theoretical assumptions. *Psychological Bulletin, 77*, 409–420.
- Goldfried, M. R., & Sprafkin, J. N. (1976). Behavioral personality assessment. In J. T. Spence, R. C. Carson, & J. Thibaut (Eds.), *Behavioral approaches to therapy* (pp. 295–321). Morristown, NJ: General Learning Press.
- Gresham, F. M., Quinn, M. M., & Restori, A. (1999). Methodological issues in functional analysis: Generalizability to other disability groups. *Behavioral Disorders, 24*, 180–182.
- Groth-Marnat, G. (2003). Behavioral assessment. In G. Groth-Marnat (Ed.), *Handbook of psychological assessment* (pp. 103–128). New York: Wiley.
- Hamilton, S. B., & Bornstein, P. H. (1977). Increasing the accuracy of self-recording in speech anxious undergraduates through the use of self-monitoring training and reliability enhancement procedures. *Journal of Consulting and Clinical Psychology, 45*, 1076–1085.
- Hayes, S. C., & Cavior, N. (1977). Multiple tracking and the reactivity of self-monitoring: Pt. I. Negative behaviors. *Behavior Therapy, 8*, 819–831.
- Hayes, S. C., & Cavior, N. (1980). Multiple tracking and the reactivity of self-monitoring: Pt. II. Positive behaviors. *Behavioral Assessment, 2*, 283–296.
- Haynes, S. N. (2006). Psychometric considerations. In M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 17–41). Burlington, MA: Elsevier Academic Press.
- Haynes, S. N., & Uchigakiuchi, P. (1993). Incorporating personality measures in behavioral assessment: Nuts in a fruitcake or raisins in a Mai Tai? *Behavior Modification, 17*, 72–91.
- Herner, R. B. (1994). Functional assessment: Contributions and future directions. *Journal of Applied Behavior Analysis, 27*, 401–404.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 885–892.
- Iverson, G. I., Stampfer, H., & Gaetz, M. (2002, Fall). Reliability of circadian heart pattern analysis in psychiatry. *Psychiatric Quarterly, 73*, 195–203.
- Iwata, B. A., & Worsdell, A. S. (2005). Implications of functional analysis methodology for the design of intervention programs. *Exceptionality, 13*, 25–34.
- Jones, M. C. (1924). The elimination of children's fears. *Journal of Experimental Psychology, 7*, 382–390.
- Jordan, C., & Franklin, C. (2003). *Clinical assessment for social workers: Quantitative and qualitative methods* (2nd ed.). Chicago: Lyceum Books.
- Kanfer, F. H., & Saslow, G. (1965). Behavioral analysis: An alternative to diagnostic classification. *Archives of General Psychiatry, 12*, 529–538.
- Knight, R., & Godfrey, H. (1995). Behavioral and self report methods. In A. Baddeley, B. Wilson, & F. Watts (Eds.), *Handbook of memory disorders* (pp. 393–410). Oxford, England: Wiley.
- Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment, 11*, 415–425.
- Kratochwill, T. R. (1985). Selection of target behaviors in behavioral consultation. *Behavior Assessment, 7*, 49–61.
- Lazarus, A. A. (1973). Multimodel behavior therapy: Treating the "BASIC ID." *Journal of Nervous and Mental Diseases, 156*, 404–411.

24 GENERAL ISSUES

- Linn, R., & Gronlund, N. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Mussgay, L., & Ruddel, H. (2004). Autonomic dysfunctions in patients with anxiety throughout therapy. *Journal of Psychophysiology*, *18*, 27–37.
- Nelson, R. O., & Hayes, S. C. (1979). Some current dimensions of behavioral assessment. *Behavioral Assessment*, *1*, 1–16.
- Nelson, R. O., Lipinski, D. P., & Boykin, R. A. (1978). The effects of self-recorders' training and the obtrusiveness of the self-recording device on the accuracy and reactivity of self-monitoring. *Behavior Therapy*, *9*, 200–208.
- Nelson-Gray, R., & Paulson, J. (2004). Behavioral assessment and the DSM system. In S. Haynes, E. Heiby, & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment: Vol. 3. Behavioral assessment* (pp. 470–488). Hoboken, NJ: Wiley.
- O'Brien, W. H., & Haynes, S. N. (2005). Behavioral assessment. In M. Hersen & J. Rosqvist (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy* (Vol. 2, pp. 704–711). Thousand Oaks, CA: Sage.
- O'Brien, W. H., Kaplar, M., & Haynes, S. (2005). Behavioral assessment. In M. Hersen & J. Rosqvist (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy: Vol. 1. Adult clinical applications* (pp. 82–90). Thousand Oaks, CA: Sage.
- O'Donohue, W., Beitz, K., & Byrd, M. (2006). Overview of behavioral assessment with adults. In M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 3–16). Burlington, MA: Elsevier Academic Press.
- O'Neill, R. (2005). Functional behavioral assessment of problem behavior. In M. Hersen & J. Rosqvist (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy* (Vol. 3, pp. 1322–1328). Thousand Oaks, CA: Sage.
- Osterlind, S. (2006). *Modern measurement: Theory, principles, and applications of mental appraisal*. Upper Saddle River, NJ: Pearson Education.
- Popham, W. J. (2002). *Classroom assessment: What teachers need to know* (3rd ed.). Boston: Allyn & Bacon.
- Robins, L. N., Cottler, L. B., Bucholz, K. K., Compton, W. N., North, C. S., & Rourke, K. (2000). *The Diagnostic Interview Schedule for DSM-IV (DIS-IV)*. St. Louis, MO: Washington University School of Medicine.
- Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York: Guilford Press.
- Sarwer, D., & Sayers, S. (1998). Behavioral interviewing. In A. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 763–778). Boston: Allyn & Bacon.
- Segal, D. L., & Coolidge, F. L. (2003). Structured interviewing and DSM classification. In M. Hersen & S. Turner (Eds.), *Adult psychopathology and diagnosis* (4th ed., pp. 72–103). Hoboken, NJ: Wiley.
- Shear, M. K., Greeno, C., & Kang, J. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry*, *157*, 581–587.
- Shriver, M., Anderson, C., & Proctor, B. (2001). Evaluating the validity of functional behavior assessment. *School Psychology Review*, *30*(2), 180–192.
- Silva, F. (1993). *Psychometric foundations and behavioral assessment*. Newbury Park, CA: Sage.
- Staats, A. W. (1963). *Complex human behavior*. New York: Holt, Rinehart and Winston.
- Steiner, J. L., Tebes, J. K., & Sledge, W. H. (1995). A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disorders*, *183*, 365–369.
- Sturgis, E., & Gramling, S. (1998). Psychophysiological assessment. In A. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 126–157). Boston: Allyn & Bacon.
- Sturmey, P. (1994). Assessing the functions of aberrant behaviors: A review of psychometric instruments. *Journal of Autism and Developmental Disorders*, *24*, 293–304.

- Sulzer-Azaroff, B., & Mayer, G. R. (1977). *Applying behavior-analysis procedures with children and youth*. New York: Holt, Rinehart and Winston.
- Tharp, R. G., & Wetzel, R. J. (1969). *Behavior modification in the natural environment*. New York: Academic Press.
- Turpin, G. (1991). The psychophysiological assessment of anxiety disorders three-systems measurement and beyond. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(3), 366–375.
- Tyron, W. (1998). Behavioral observation. In A. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 79–103). Boston: Allyn & Bacon.
- Ullmann, L. P., & Krasner, L. A. (1965). *Case studies in behavior modification: A systematic extension of learning principles*. New York: Holt, Rinehart and Winston.
- Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, 3, 1–14.
- Wilson, G. T., & Vitousek, K. M. (1999). Self-monitoring in the assessment of eating disorders. *Psychological Assessment*, 11, 480–489.
- Witt, J. N., & Elliott, S. N. (1983). Assessment in behavioral consultation: The initial interview. *School Psychology Review*, 12, 42–49.
- Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., & Schaefer, E., et al. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of Axis I and II diagnoses. *Journal of Personality Disorders*, 14, 291–299.