

3

BASIC FEATURES OF CLINICAL ASSESSMENT

Chapter Preview

In this chapter we offer an overview of the clinical assessment process. We outline the range of assessment options available to clinicians, then describe the typical goals of assessment: diagnosis, description, treatment planning, and prediction. We also introduce factors affecting a clinician's choices about how to conduct an assessment, including the purpose of the assessment, the clinician's theoretical views, the psychometric properties of assessment instruments, and other contextual factors. We discuss the use of clinical judgment, focusing especially on errors that clinicians try to avoid. Assessments ultimately must be communicated to clients and third parties, so we conclude by discussing factors associated with reporting assessment results.

A Clinical Case

Dr. T was asked to take the case of Jessie, a 17-year-old male whose school counselor had become concerned that Jessie was showing signs of depression and rebellion. In addition to missing school because of alleged illnesses (his counselor does not believe him), Jessie has frequently been belligerent with some of his teachers. Jessie's parents drove him to the appointment, but they declined to meet with the therapist, saying only that they wanted to know what was wrong with their son and that they wanted him to get whatever treatment the insurance would cover. In the intake interview, Jessie expressed a strong dislike of school, saying it was a waste of time. He said that no one understands him there, and most of his classmates only ignore him or make fun of him. In school, he was mostly silent and sullen and did not speak with classmates. Outside of school, he spent most of his time with a 20-year-old friend who worked as a dishwasher at a local diner. Last week, the two young men had been caught in possession of alcohol. He admitted that he had been depressed and that for at least the last 3 weeks had "no interest in anything." He said that during this time he had slept for longer periods, sometimes as much as 14 to 16 hours per day. He also claimed to have had "some pretty bizarre thoughts about leaving this life and living in another dimension." Elaborating further, Jesse admitted to beliefs that he was not meant for this world, that he was being told (he could not or would not say by whom) that he might be better off dead. He also informed the therapist that his parents were in the process of getting a divorce.

How can the therapist best understand Jessie's problems? Which of the problems are the most pressing? What diagnosis most accurately describes those problems? Is Jessie suicidal or dangerous to others? What role do the parents play in his difficulties? How can an effective treatment be designed? In order to answer these questions, the psychologist must conduct an assessment.

Assessment is the collection and synthesis of information to reach a judgment. Almost everyone engages in some type of assessment at one time or another. For example, whether we realize it or not, we collect, process, and interpret information about the background, attitudes, behaviors, and characteristics of the people we meet. Then, in light of our experiences, expectations, and sociocultural frame of reference, we form impressions that guide decisions to seek out some people and avoid others.

Clinical psychologists collect and process assessment information that is more formal and systematic than that available to nonprofessionals. As discussed in Chapter 2, "Clinical

Psychology's Past and Present," before psychotherapy became a major role for clinicians, assessment was their most common applied activity. Although no longer the dominant activity of clinical psychologists, assessment remains a critical part of their training and practice. After all, assessment is required in order to delineate a client's problems, plan treatments, measure treatment effectiveness, conduct other kinds of research, and answer a multitude of questions asked of clinicians (Antony & Barlow, 2010). In this chapter, we consider what clinical psychologists have learned about the challenge of clinical assessment. In the following two chapters, we discuss specific types of assessments.

AN OUTLINE OF THE ASSESSMENT PROCESS

SECTION PREVIEW

Assessment involves a series of steps, beginning with a problem or referral and ending with the psychologist's communicating the results of the assessment to appropriate parties. In between, the clinician must make a number of judgments, such as which instruments to use and how best to convert clinical assessment data into a clinical judgment. At each step, the clinician must engage in activities to ensure that the most relevant data are gathered and analyzed.

Clinical assessment has been described in various ways (Tallent, 1992), but all of them portray it as a process of gathering information to solve a problem. All of them recognize that to be most effective, assessment activities should be organized in a sequence of systematic, logically related steps driven by a goal. Most assessments follow the general sequence outlined in Figure 3.1.

Receiving and Clarifying the Referral Question

Two related questions must be answered before clinical assessment can begin (McReynolds, 1975): What do we want to know, and how best can we find out about it? Answers to the first question—what do we want to know?—depend on who requested the assessment and for what purpose. The person or agency requesting the psychological assessment is called the *referral source*, and the question or issue to be addressed in the assessment is called the *referral question*. (Information that initiates an assessment is sometimes called the *presenting problem*, but we prefer to begin with a *referral question* because a specific question focuses on the assessment goal, while the presenting problem is often a longer description of the client's difficulties.)

The referral question is very important because it shapes the clinician's choice of assessment instruments and the interpretation and communication of results. Clinicians must therefore understand the context of the referral and often must help referral sources clarify the purpose of the assessment (Harwood, Beutler, & Groth-Marnat, 2011). In doing so, clinicians may need to educate others about what a psychological assessment can and cannot reveal. They may also need to educate them about the practical and ethical constraints involved in conducting assessments.

The referral question is the first step in shaping the ultimate goal of assessment, and, generally, the clearer the goal, the clearer the question. As you can see in Table 3.1, referral questions can be quite varied.

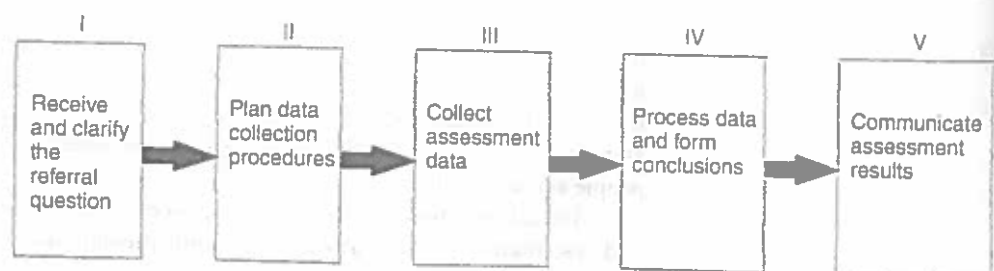


FIGURE 3.1 A Schematic View of the Clinical Assessment Process

TABLE 3.1 Examples of Referral Questions from Various Sources

Source	Referral Question
Therapist	Would this person be a good candidate for group psychotherapy? How effective have I been in treating clients with anxiety disorders?
Physician or treatment team	What is the correct psychological diagnosis for this client? What cognitive and emotional limitations does this person have following her accident, and what strengths does she have to draw on during rehabilitation?
Client	What, if anything, should we do to salvage this marriage? Would I be good at this kind of job?
School	How can I worry less and enjoy my life and my family more? What is an appropriate educational placement for this child? How should we intervene to help this student better manage violent tendencies and problems in relationships with peers and teachers?
Parent	Is my child suffering from depression and is she in need of counseling? How can we help my child cope with the loss of his mother? My child becomes very emotional and refuses to go to school: what can I do?
Court	Does this person pose an imminent threat of danger to himself or others? Which custody arrangement is in the best interest of this child? Is this client mentally able to understand the criminal charges against him and to assist an attorney in mounting a legal defense? Has this person sustained psychological injuries that would be relevant to litigation?
Employer, government agency, or other third party	Is this person suffering from a mental illness that would qualify him to receive disability payments? What form of treatment would best help this person to resume productive employment, and for which types of employment would the person be best suited? Will the survivors of this natural disaster benefit from rapid intervention, and if so, what kind of intervention?

Planning Data Collection Procedures

Answers to the second question—about how best can we find out what we need to know—come into play after the referral question and the clinician's role have been clarified. With a clear goal in mind, the clinician can now begin planning methods to collect data.

There is a vast range of data that might be collected about a client. To illustrate, consider the sheer number of things that can be asked about a person at several interrelated levels, from biological functioning to relationships with other people (see Table 3.2). The enormous diversity of possible assessment data means that we can never learn all there is to know about a client. Of course, clinicians would typically seek more information about the client in categories that are most relevant to the assessment goals.

Although the most important factor in the selection of assessment instruments is the referral question, other factors also affect the selection process. One of these is the quality of the assessment instrument or procedure. Obviously, clinicians would be better off selecting assessment methods whose psychometric properties (e.g., reliability, validity, usefulness) are the highest, but they might not always do so. For example, if one test had slightly lower (but still acceptable) reliability than another but provided more relevant information about a particular referral question, the clinician might select the one with slightly lesser reliability but greater relevance to assessment.

TABLE 3.2 Information Used in a Case Study Guide

1. Identifying data, including name, sex, occupation, income (of self or family), marital status, address, date and place of birth, religion, education, cultural identity.
2. Reason for coming to the agency, expectations for service.
3. Present and recent situation, including dwelling place, principal settings, daily round of activities, number and kind of life changes over several months, impending changes.
4. Family constellation (family of origin), including descriptions of parents, siblings, other significant family figures, and respondent's role growing up.
5. Early recollections, descriptions of earliest clear happenings and the situation surrounding them.
6. Birth and development, including age of walking and talking, problems compared with other children, view of effects of early experiences.
7. Health and physical condition, including childhood and later diseases and injuries; current prescribed medications; current use of unprescribed drugs, cigarettes, or alcohol; comparison of own body with others; habits of eating and exercising.
8. Education and training, including subjects of special interest and achievement, out-of-school learning, areas of difficulty and pride, any cultural problems.
9. Work record, including reasons for changing jobs, attitudes toward work.
10. Recreation, interests, and pleasures, including volunteer work, reading, respondent's view of adequacy of self-expression and pleasures.
11. Sexual development, covering first awareness, kinds of sexual activities, and a view of adequacy of current sexual expressions.
12. Marital and family data, covering major events and what led to them, and comparison of present family with family of origin, ethnic or cultural factors.
13. Social supports, communication network, and social interests, including people talked with most frequently, people available for various kinds of help, amount and quality of interactions, sense of contribution to others, and interest in community.
14. Self-description, including strengths, weaknesses, ability to use imagery, creativity, values, and ideas.
15. Choices and turning points in life, a review of the respondent's most important decisions and changes, including the single most important happening.
16. Personal goals and view of the future, including what the subject would like to see happen next year and in 5 or 10 years and what is necessary for these events to happen, realism in time orientation, ability to set priorities.
17. Any further material the respondent may see as omitted from the history.

Source: From Sundberg (1977). Sundberg, N.D., *Assessment of Persons*, 1st Ed., © 1997, pp. 97-98, 207. Reprinted and Electronically reproduced by permission of Pearson Education, Inc., Upper Saddle River, New Jersey.

Clinicians must consider, too, the characteristics of clients when deciding on assessments, selecting instruments that are appropriate for each client in terms of reading level, length, and the like. Similarly, clinicians must explain to clients the procedures and purposes of the assessment using language that clients can understand.

To plan assessments, then, the clinician must think broadly, weighing various features of assessment instruments against practical considerations of time, context, and usefulness to clients and other referral sources (Hunsley & Mash, 2010). The clinician must also plan to combine and integrate the results of assessments into a narrative that serves a specific clinical purpose.

Collecting Assessment Data

Once the referral question has been clarified, the appropriate assessment methods have been selected, and the client's cooperation has been secured, the data-collection stage of assessment can begin.

THE VALUE OF MULTIPLE ASSESSMENT SOURCES Clinical psychologists collect assessment data from four main sources: interviews, observations, tests, and historical records (case history data). However, they seldom rely on a single assessment source to create a working image of a client. Instead, they use multiple assessment channels to cross-validate information about a wide variety of topics. Thus, hospital records may reveal that a patient has been there for 30 days, thus correcting the patient's self-reported estimate of 2 days. Indeed, the whole story of a client's

problems is seldom clear until multiple assessment sources are tapped. For example, people have been observed to be assertive in social situations even though they described themselves in an interview or on a test as generally unassertive (Nietzel & Bernstein, 1976). It often takes multiple sources of assessment to separate those who do not engage in certain behaviors from those who cannot engage in them.

Another benefit of using multiple assessment sources appears when the clinician evaluates the effects of treatment. Suppose a couple enters therapy because they are considering divorce, and then, 3 months later, they do divorce. If the only outcome assessment employed in this case were "marital happiness," as expressed during interviews, the treatment might be seen as having worsened marital distress. However, observations, third-person reports, and life records might show that one or both partners find their newly divorced status liberating and that they are developing new interests and abilities.

Processing Data and Forming Conclusions

After assessment data have been collected, the clinician must determine what those data mean. If the information is to be useful in reaching the clinician's assessment goals, it must be transformed from raw form into interpretations and conclusions that address a referral question. The processing task is formidable because it requires a mental leap from known data to what is assumed to be true on the basis of those data. In general, as the leap from data to assumption gets longer, inference becomes more vulnerable to error.

Consider this: A young boy is sitting on a lawn cutting an earthworm in half. It would be easy to infer from this observational data that the child is cruel and aggressive and that he might become dangerous later in life. These inferences would be off the mark, however, for "what the observer could not see was what the boy, who happened to have few friends, thought as he cut the worm in half: 'There! Now you will have someone to play with!'" (Goldfried & Sprafkin, 1974, p. 305). In short, elaborate inference, especially when based on minimal data, can be dangerous.

Processing assessment data is difficult also because information from various sources must be integrated. Unfortunately, there are few empirical guidelines for how best to combine data from interviews, tests, observations, and other sources to reach integrated conclusions. So in forming their conclusions, clinicians often must rely heavily on clinical judgment, a topic we describe in some detail later in this chapter.

Communicating Assessment Results

The final stage in the assessment process is the creation of an organized presentation of results called an *assessment report*. To be of greatest value, assessment reports must be clearly written and clearly related to the goal that prompted the assessment in the first place. If that goal was to classify the client's behavior into a diagnostic category, information relevant to diagnostic classification should be highlighted in the report. If assessment was aimed at determining a client's likely cooperativeness in and responsiveness to psychotherapy, the report should focus on those topics. These simple, self-evident prescriptions are sometimes ignored in assessment reports, especially when assessment goals were never explicitly stated or when clinicians do not fully understand what their assessment instruments can, and cannot, reveal about clients.

SECTION SUMMARY

Assessment progresses in a series of steps that begin with a referral question. The referral question may come from a client, a third party, or the therapist, but the clinician must ensure that the question is clear and can be addressed reasonably by the available assessment instruments. Once the referral question is clarified, planning for assessment begins. Planning involves selection of appropriate instruments—not a simple task given the number of factors affecting clinicians' choices—and preparation of the client. Data collection involves accumulating evidence from historical records, interviews, observations, and tests. Each of these sources has special strengths and limitations. In the data-processing stage, the accumulated data are analyzed and integrated into a coherent assessment. In the final stage, the clinician communicates conclusions and provides recommendations in ways that are clear and useful to the referral source.

THE GOALS OF CLINICAL ASSESSMENT

SECTION PREVIEW

Most referral questions relate to diagnosis, description, treatment planning, or prediction. We first consider diagnostic classification, the labeling of psychological problems. Psychological diagnosis relies primarily on criteria described in the Diagnostic and Statistical Manual of Mental Disorders (DSM) of the American Psychiatric Association. A diagnosis or label alone seldom tells clinicians all they want to know about a client, so clinicians often assess clients' strengths, weaknesses, and social connections. Diagnoses, descriptions, and other information help clinicians design treatments for each client, which is another goal of assessment. Measuring the outcomes of treatment is yet another assessment goal, an increasingly important one in this era of accountability and evidence-based practice. Finally, clinicians sometimes conduct assessments to make prognoses or predictions, difficult judgments that can be improved if clinicians attend to the lessons provided by years of empirical research on the topic.

Diagnostic Classification

As discussed in Chapter 2, once clinical psychologists began working with adult clients during and after World War I, they were often asked to perform clinical assessments for the purpose of diagnosing mental disorders in psychiatric patients. This process is variously called diagnostic classification, psychodiagnosis, differential diagnosis, or diagnostic labeling. Today, diagnostic classification remains a significant part of clinical research and practice.

Accurate psychodiagnosis is important for several reasons. First, proper treatment decisions often depend on knowing what, exactly, is wrong with a client (Harwood, Beutler, & Groth-Marnat, 2011; Hays, 2013). Second, research into the causes of psychological disorders requires reliable and valid identification of disorders and accurate differentiation of one disorder from another. Finally, classification allows clinicians to efficiently communicate with one another about disorders in a professional "shorthand" (Sartorius et al., 1996).

The use of a common nomenclature (naming system) is extremely important in any field. Imagine, for example, how confusing it would be if medical patients who developed pharyngitis (a sore throat) because of A-beta-hemolytic *Streptococcus* (a bacterium) received a different diagnosis depending on which state or country they lived in. Fortunately, if you have this condition and are accurately assessed, you will be given the same diagnosis (strep throat) whether you are in New York, Peoria, Hong Kong, or Helsinki. That is because the World Health Organization's (WHO) *International Classification of Diseases* (ICD) provides common codes that identify each disease. In clinical psychology and psychiatry, the comparable system is the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, or *DSM*.

A BRIEF HISTORY OF THE DSM Various systems for classifying mental disorders had been used since the early 1900s, but classification of mental disorders became more formalized in 1952 when the American Psychiatric Association published the first *Diagnostic and Statistical Manual of Mental Disorders*. Known as *DSM-I*, it remained in use until 1968, when—to make the *DSM* more similar to the World Health Organization's *ICD*—it was replaced by *DSM-II*.

DSM-I and *DSM-II* provided a uniform terminology for describing and diagnosing abnormal behavior, but they offered no clear rules to guide mental health professionals' diagnostic decisions. Accordingly, when *DSM-III* appeared in 1980, it included a set of criteria for assigning each diagnostic label. These criteria—which referred mainly to specific symptoms and symptom durations rather than inferred causes—were increased in number and specificity in *DSM-III-R*, which was published in 1987. Clients were to be diagnosed with a particular disorder only if they met a pre established number of criteria from the full list of criteria associated with that disorder.

DSM-III and *DSM-III-R* also introduced *multi-axial* diagnoses, allowing clinicians to describe clients along different dimensions, or *axes*, providing a more complete picture of clients' problems and the factors affecting them. Axis I was for disorders that had a noticeable onset and course, such as when a person who is functioning well develops debilitating panic attacks within a relatively short period of time. Axis II disorders included those that are long-standing and resistant to change, such as various pervasive developmental problems (e.g., autistic spectrum disorders) and personality disorders. Axes III through V allowed diagnosticians to identify physical conditions that might

affect a person's mental state and functioning (e.g., cancer), the severity of recent psychosocial stressors (e.g., divorce or job loss), and the person's psychological, social, and occupational functioning during the past year. *DSM-III* and *DSM-III-R* improved the reliability of most diagnoses, but they also had shortcomings. So in 1988, only a year after *DSM-III-R* appeared, the American Psychiatric Association established a task force to begin work on *DSM-IV*.

New editions of the *DSM* are always major enterprises; they take years and require coordinating the efforts of thousands of people. To develop *DSM-IV*, planners organized numerous experts into 13 work groups, each of which followed a three-step procedure for studying a different set of disorders and how best to diagnose them (Widiger, Frances, Pincus, Davis, & First, 1991). Each group first reviewed all the clinical and empirical literature relevant to a given disorder and used their findings to guide initial suggestions for changes in diagnostic criteria for that disorder. When literature reviews failed to resolve issues, the work groups sought to conduct analyses on existing patient data sets. Another step involved asking clinicians to use diagnostic criteria in clinical settings with real clients. Conducted at more than 70 sites worldwide, these focused *field trials* examined issues such as how alternate wordings or alternate thresholds (cutoffs) affected reliability, prevalence rates, or concordance with the parallel diagnoses from *ICD-10* (Nathan & Lergenbucher, 1999). The *DSM-IV*, with its revised diagnostic criteria, was published in 1994. It was followed in 2000 by a text revision (*DSM-IV-TR*) that did not change its diagnostic criteria, but did update information about the prevalence, onset, course, and other aspects of various disorders (APA, 2000).

The developers of *DSM-IV* and *DSM-IV-TR* sought to extend the improvements in reliability begun with *DSM-III-R*'s use of observable criteria. They hoped to develop categories that relate well to specific outcomes (predictive validity) or specific underlying neurological or genetic constellations (construct validity). But despite its improved empirical foundations, the *DSM-IV* diagnostic system continued to draw criticism. For one thing, although many Axis I disorders showed good to excellent reliability, disorders within the schizophrenic spectrum and some childhood and adolescent disorders did not. And Axis II disorders, particularly the personality disorders, continued to show unacceptably low inter-rater and test-retest reliability (Marini et al., 2000).

Critics especially questioned whether the either-or categorizations imposed by *DSM-IV* criteria were the best way to understand psychopathology. If individuals met a certain number of criteria, they were said to "have" a disorder; if not, they do not "have" it. In fact, though, psychological disorders can be present in varying degrees. For example, there is no clear borderline between being depressed and not depressed. Critics contended that diagnoses should be considered as extremes along one or more underlying dimensions (Widiger & Trull, 2007).

Clinicians from various theoretical orientations also criticized the *DSM-IV* for not providing enough useful information about disorders. Behaviorally oriented clinicians were concerned that *DSM-IV* classification ignored the context in which symptoms occur, thus providing no basis for understanding the meaning or function that a particular pattern of disordered behavior might have in different social circumstances (Follette, 1996; Wulfert, Greenway, & Dougher, 1996). Psychodynamically oriented clinicians criticized *DSM-IV*'s strong emphasis on observable symptoms. They were concerned that clinicians think they understand clients once a diagnostic label is applied, when in reality they may not understand the client's *subjective* experience of having the disorder (Packard, 2007). Neurologically oriented clinicians also had criticisms, arguing that emphasizing observable symptoms fails to take into account discoveries about the brain circuitry associated with specific disorders (Bracha, 2006).

Critics also noted that the *DSM* excludes certain conditions. Every version of the *DSM* has been based on, and applied to, North American populations. None included symptom clusters seen in other cultures (Hall, 2005). Also missing from the *DSM* were the many types of *relational disorders*, such as couples conflict, parental discipline problems, child neglect, sibling conflict, domestic abuse, incest, and the like. These conditions produce considerable psychological suffering (First, 2006).

Even the very number of diagnoses in *DSM* has been criticized. The first edition, *DSM-I*, contained 106 diagnoses; in *DSM-IV*, there are 365 (see Figure 3.2). This 300% increase in diagnostic categories over four decades has led some to wonder if that many new disorders have been discovered and scientifically verified, or if the increase is more representative of the expansion of the mental health field and its practitioners' need to identify and treat a wider variety of conditions (see Houts, 2004, for an interesting discussion).

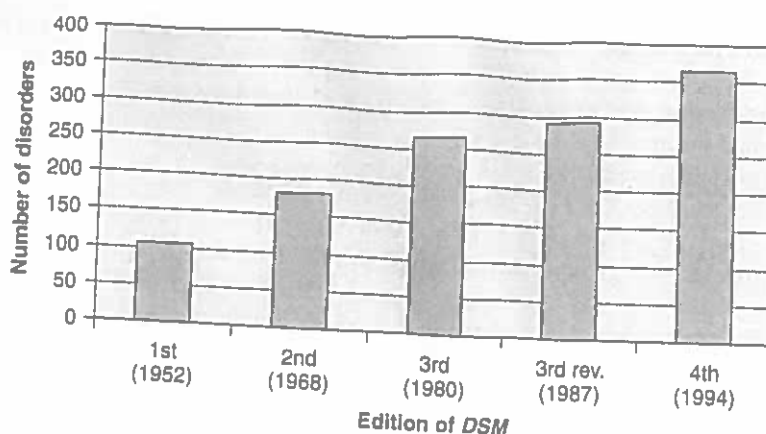


FIGURE 3.2 Number of Diagnoses in the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* Across Editions (DSM-5 did not significantly change the number of possible diagnoses; rather it modified the diagnostic criteria for certain disorders and reorganized disorders into 17 broad categories)
Source: Houts (2004).

THE DSM-5 Like its predecessors, the newest edition of the *DSM*, *DSM-5*, was designed to be an improvement over previous editions. It too depended on input from numerous work groups, research studies, and field trials. Publication of the *DSM-5* in May of 2013 was preceded by public comment periods, where interested persons expressed their views on proposed changes (see www.dsm5.org). Like its predecessors, this new edition serves as the primary diagnostic system for mental health professionals in North America.

Among the most important changes introduced in *DSM-5* is the new organization of its chapters. Some *DSM-IV* categories have changed, as have some of the diagnoses within those categories. For instance, in *DSM-IV*, bipolar disorders and depressive disorders both appeared in the mood disorders chapter. In *DSM-5*, these disorders appear in separate chapters—one for bipolar and related disorders and one for depressive disorders. Obsessive-compulsive disorder was in the anxiety disorders chapter of *DSM-IV*, but now has its own chapter that also includes body dysmorphic disorder, hoarding disorder, hair-pulling (“trichotillomania”) disorder, and skin-picking (“excoriation”) disorder (American Psychiatric Association, 2013). The table of contents has also changed (see Table 3.3). The first chapter of *DSM-5* is now the one on neurodevelopmental disorders many of which are diagnosed in infancy or childhood. These include intellectual disability (intellectual developmental disorders; formerly called mental retardation), autism spectrum disorders, attention deficit hyperactivity disorder, specific learning disorders, motor disorders, and various communication disorders. Disorders first identified in childhood are also placed first within each of the other chapters. Finally, the multiaxial system has been dropped because the developers concluded that there was not sufficient scientific justification for it.

Consider again the case of Jessie presented at the beginning of this chapter. A tentative *DSM-5* diagnosis for him is presented below. It is based on the information available in the vignette, but the clinician would also investigate further the extent of Jesse’s substance use and the “pretty bizarre thoughts” that he reported. The latter would be particularly important because part of the *DSM-5* diagnosis involves determining whether there are psychotic features and whether those features are mood congruent or mood incongruent. In this case, the clinician judged that such features are present, so Jessie’s *DSM-5* diagnosis might look like this:

D 01 Major Depressive Disorder, single episode, with mood-congruent psychotic features
Severity level: Moderate (based interview and a Patient Health Care Questionnaire)
Additional concerns: Relational problem, parental divorce

Because of the significant changes in *DSM-5*’s chapter organization, disorder names, and criteria, clinicians who have been using *DSM-IV* will need to retrain themselves to learn the new system, particularly if insurance companies, governmental agencies, and other groups rapidly adopt the changes. Existing psychological tests, structured interviews, and other diagnostic instruments will have to be revised, and new ones will have to be developed so as to conform to

TABLE 3.3 Significant Changes in the *Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5)***Basic Organization**

- Disorders presumed to be more influenced by neurological and neurodevelopmental factors now appear first, a change designed to increase similarity with the *International Classification of Diseases (ICD-11)*
- Disorders with similar features are listed adjacent to each other
- Multi axial categorization of disorders using Axes I-V is discontinued

Disorder Categories

- New disorders such as disruptive mood dysregulation disorder and binge eating disorder are now listed
- Disorders such as Asperger's disorder, childhood disintegrative disorder, and pervasive developmental disorder were dropped as distinct diagnoses; their symptoms are now included in a single diagnosis of autism spectrum disorder
- Depressive disorders and bipolar disorders are placed in separate categories
- Obsessive—compulsive disorder and posttraumatic stress disorder (PTSD) are moved to categories separate from anxiety disorders
- There is a new category of trauma and stressor-related disorders, which includes diagnoses such as PTSD, adjustment disorders, and reactive attachment disorder

Diagnostic Criteria

- The diagnosis of major depressive disorder no longer includes an exemption for bereavement
- Substance abuse and substance dependence categories will be merged

changes in DSM-5 diagnostic criteria. There will no doubt be numerous books, practice manuals, workshops, conferences, seminars, and the like to help mental health professionals make the transition to DSM-5.

CRITICISMS AND ALTERNATIVE DIAGNOSTIC PROPOSALS Like its predecessors, DSM-5 has generated intense interest and considerable controversy. As you might guess, not everyone is pleased with the development of the new DSM. Among the strongest criticisms are those expressed by members of the American Psychological Association's Division 32 (Society for Humanistic Psychology). In an open letter to the developers of DSM, they complained about what they perceived as lowered diagnostic standards, the introduction of new disorders, and the lack of empirical grounding for some disorders. Each of these, they contend, increases the "medicalization" of psychological disorders and de-emphasizes understanding of the sociocultural factors associated with disorders (Clay, 2012). Others have cautioned that psychiatrists may be inclined to expand certain diagnostic categories, that they may not be entirely objective and evidence based in their decisions (Frances & Widiger, 2011; Miller, 2012). Defenders of DSM-5 counter that this is an unfair characterization of the motives of honorable professionals who worked hard to create a more effective diagnostic system (Black, 2013).

Clinicians and researchers concerned about the shortcomings in DSM-IV and DSM-5 have proposed alternative diagnostic systems. Several have proposed that rather than assigning diagnoses based on whether or not clients show a specified—some say arbitrary—number of symptoms, clinicians should base diagnoses on measurement of client characteristics on a selected set of dimensions that are relevant for each disorder. Critics contend that such *dimensional approaches* would avoid the DSM's either-or dichotomization.

Arguments for a dimensional approach have been especially persuasive in the diagnosis of personality disorders, where diagnostic reliability has tended to lag. Though DSM-5 does incorporate some dimensional assessment (American Psychiatric Association, 2013), critics contend that the dimensional approach has not been applied broadly enough. Others (e.g., Westen, 2012) suggest that clinicians should diagnose on the basis of the degree to which a given client's characteristics resemble prototypical examples of particular disorders. So rather than noting the presence or absence of symptoms from a disorder checklist, clinicians could rate, from 1 to 5, the degree to which a client's behavior matched a prototypical description of each disorder.

Other diagnostic proposals call for describing psychopathology in terms of specific theories or contextual principles. For instance, Bornstein (2006) explored how concepts such as ego strength, defense style, and mental representation of the self and others have considerable diagnostic utility. This kind of thinking prompted the publication of an alternative diagnostic system, the *Psychodynamic Diagnostic Manual (PDM)* in 2006. This manual was developed by a task force whose members come from five major psychoanalytic groups. Advocates of this alternative system argue that it promotes more depth-oriented assessment. They claim that it also encourages clinicians to go beyond overt symptoms during the diagnostic process and to generate more complex pictures of clients, which in turn allows for more sophisticated case formulation and better treatment planning (Brabender & Whitehead, 2011; Huprich & Meyer, 2011).

The field of positive psychology has provided another alternative diagnostic approach. Advocates of this approach agree with some other critics that diagnosis based on problematic symptoms alone is insufficient. What is missing, they say, is the systematic assessment of clients' strengths (Ehde, 2010; Joseph & Wood, 2010). In fact, they argue that the "real" psychopathologies, the ones that "cut nature at the joints," might be better understood as the absence of the character strengths, not as the presence of symptoms described by *DSM* (Duckworth, Steen, & Seligman, 2005). One result of this approach is *The Handbook of Positive Psychology Assessment* (Lopez & Snyder, 2003) which describes approaches to assessment based on positive psychology. Another is Peterson's (2006) "*Values in Action (VIA) Classification of Strengths: The un-DSM and the Real DSM.*"

It is no wonder that the diagnostic classification enterprise continues to be a hotbed of controversy. After all, it attempts to differentiate a wide variety of complex, socioculturally defined disorders caused by multiple (often unknown) factors using a relatively small set of shorthand labels organized in a way that will help clinicians make the best possible decisions about their clients. No diagnostic system is likely to ever accomplish all these goals. So although various alternative approaches may each have their advantages, it is unlikely that any of them will replace the current categorical classification system. The categorical system is consistent with the medical model that dominates the health care industry, and it offers an efficient shorthand through which clinicians can communicate with each other. The challenge for competing diagnostic schemes, and for the newest versions of the *DSM*, is to improve upon diagnostic reliability, validity, and utility while still maintaining a system that is efficient.

Description

As just noted, classification should be reliable and valid and serve as an efficient shorthand for communication among clinicians and researchers. But for many clinicians, diagnostic labels are not enough. They want to know more, and consequently, alternative diagnostic proposals often shade over into *descriptive assessment*, which these clinicians see as more important than diagnostic classification.

Descriptive assessment by a cognitive-behavioral therapist, for instance, might focus on outlining factors such as antecedent conditions, environmental incentives and disincentives, alternative sources of reward, cognitive complexity, and attributional style. A psychodynamic therapist might focus descriptive assessment on ego strengths and weaknesses, cognitive functioning, defense mechanisms, quality of family and other relationships, and characteristics of the self (Gabbard, Litowitz, & Williams, 2012). Of course, diagnostic classification and descriptive assessment can go hand in hand. Bertelsen (1999), for instance, suggests that clinical assessment should work at two separate levels: one concerned with diagnostic classification, the other concerned with evaluating multiple factors that influence the course of treatment.

Description-oriented assessment makes it easier for clinicians to pay attention to clients' assets and adaptive functions, not just to their weaknesses and problems. Accordingly, descriptive assessment data are used to provide pretreatment measures of clients' behavior, to guide treatment planning, and to evaluate changes in behavior after treatment. Descriptive assessment can also improve measurement in clinical research. For example, in an investigation of the relative value of two treatments for depression, assessments that describe clients' posttreatment behaviors (e.g., absenteeism, self-reported sadness, and depression test scores) are of greater value than diagnostic labels (e.g., depressive versus nondepressive).

Unfortunately, the movement toward broad description of persons may never dominate clinical assessment, especially in inpatient psychiatric settings and other managed-care facilities.

The reason is time. As skyrocketing health care costs have increased economic pressures to limit hospital stays and to concentrate on focused assessments and short-term treatments, time-consuming and comprehensive patient evaluations are just too expensive. The degree to which descriptive approaches to assessment gain widespread usage will ultimately depend on their empirically demonstrated usefulness—the degree to which clinicians can perform assessments efficiently and use them to design better research and treatments.

Treatment Planning and Treatment Assessment

Diagnostic and descriptive assessment can be used to plan treatments. In the simplest model, a diagnosis (e.g., depression) leads to a preferred treatment (e.g., cognitive-behavioral psychotherapy), just as a given medical illness (e.g., strep throat) might lead to a preferred treatment (e.g., antibiotic). Identifying ideal connections between diagnoses and psychotherapy methods has been a main goal of the empirically supported treatments movement (see our discussions in Chapter 1, “What Is Clinical Psychology,” and Chapter 10, “Research on Clinical Intervention”).

While efforts to match specific treatments with specific diagnoses have certainly improved the empirical base of psychotherapy, the enterprise has not worked out as cleanly in practice as many had hoped. As the previous discussion on diagnosis implies, diagnostic classification is imprecise, and it is difficult to argue that specific treatments work for specific diagnoses when the diagnoses themselves keep changing. Further, extensive research on psychotherapy effectiveness (discussed in Chapter 10) shows that treatments often have general and overlapping effects, some of which have more to do with characteristics of the persons involved than with the disorders to which they are applied (Beutler & Malik, 2002; Norcross & Lambert, 2011; Singer, 2013). Because of these problems, clinicians have sought to incorporate other factors into treatment-related assessment. The key is to identify which factors, apart from the diagnosis and the “brand” of psychotherapy, best predict how an intervention will work.

In short, treatment planning assessment goes beyond the basic medical-model question: *Which treatments work best for which disorders?* Instead, it addresses the more detailed question, famously stated by Gordon Paul (1967): “*What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?*” (p. 44). In Chapter 6, “Basic Features of Clinical Interventions,” and in Chapter 10, “Research on Clinical Intervention,” we discuss treatment planning further and we address answers to Paul’s important question in much more detail.

Treatment-related assessment can also focus on the question of how well treatment has worked. Indeed, clinicians working in today’s accountability-driven climate are increasingly asked to provide evidence of their effectiveness. Generating that evidence requires assessment. As an example, consider a case described by Alan Kazdin (2006).

Gloria, a 39-year-old woman, was self-referred for depression. An initial assessment was conducted, and a treatment plan involving combined cognitive-behavioral and interpersonal therapy, administered on weekly visits, was devised. The therapist also had Gloria come about 20 minutes early each week and complete brief measures designed to track her progress. One measure was a rating on a series of statements developed collaboratively between Gloria and the therapist, which they called the G-Scale. Another was an abbreviated version of the Beck Depression Inventory (BDI; Beck, Steer, & Garbin, 1988), and the last was an abbreviated version of the Quality of Life Inventory (QOLI; Frish, 1998).

Figure 3.3 represents the session-by-session changes in Gloria’s scores over several weeks of treatment.

There are many ways to assess the results of treatment. Questionnaires, client self-reports, tests, and a variety of other measures can be used to evaluate treatment progress. For most clinicians, the goal is to record outcomes quickly, efficiently and accurately, and to periodically review the data to assess their treatment’s strengths and weaknesses. Having these outcome data is essential even without the pressure applied by third party payers because conscientious self-assessment is part of the professionalism normally expected of clinical psychologists, and indeed of all health care providers.

Assessments can also be used as an explicit component of treatment. There is substantial evidence that discussing assessment methods and results with clients can have great therapeutic value (Finn & Hale, 2013). Imagine, for instance, Gloria and her therapist discussing the assessment results

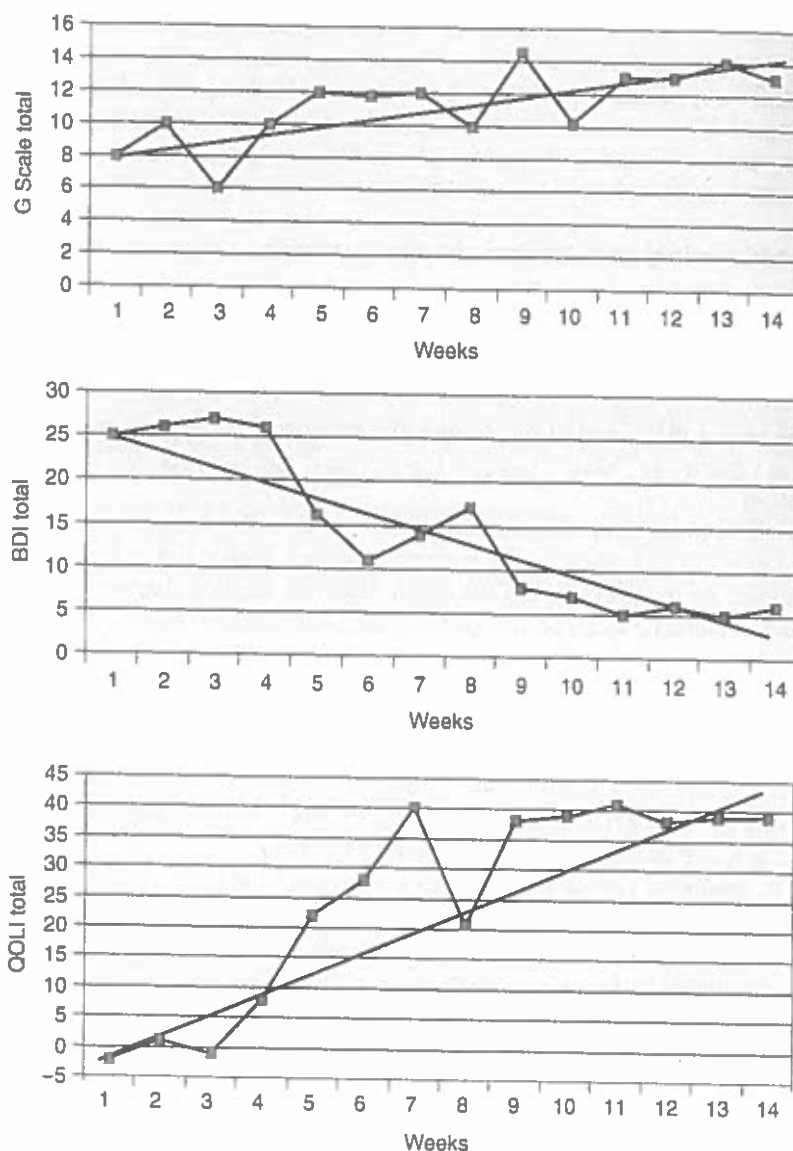


FIGURE 3.3 Assessment of Gloria's Psychotherapy Progress on Three Measures. (The first two weeks can be considered baseline measures. A linear regression line has been fitted to the data points on each graph.)

Note: G Scale is a scale developed by the client, Gloria, and the therapist; BDI is the Beck Depression Inventory (selected items), and QOLI is the Quality of Life Inventory. *Source:* Kazdin, Alan E., *Research Design In Clinical Psychology*, 4th Ed., © 2003, p. 322. Reprinted and Electronically reproduced by permission of Pearson Education, Inc., Upper Saddle River, NJ 07458.

in Figure 3.3. Greater involvement in assessment might encourage her to be more objective and accurate in her self-monitoring; it might increase her trust in the therapist and in the treatment; it can provide her with another avenue of therapeutic interaction. Assessment expertise is a traditional strength of clinical psychologists, so it makes sense to use this strength to therapeutic advantage (Youngstrom, 2013).

Prediction

A final goal of clinical assessment is to make predictions about human behavior. Such predictions might include forecasts about how the symptoms of a client's disorder might change with or without treatment (prognosis), about future performance (descriptions of how someone will perform in a given job or situation), or about dangerousness (descriptions of the likelihood of

someone behaving violently toward the self or others). In any of these predictions, clinicians must have valid information about how the characteristics revealed by assessment relate to the behavior being predicted. Without that link, prediction would be little more than guesswork.

PROGNOSIS Most often, prognosis refers to a prediction about the outcome of treatment, but it can also refer more generally to predictions about changes in symptoms without treatment or under certain circumstances (e.g., a head injury might change the prognosis for treatment of an anxiety disorder).

DSM diagnoses already contain considerable information about prognosis, including results of long-term studies of the course of a disorder as well as other information related to its typical onset, chronicity, those most at risk, and the like. And although the *DSM* does not provide specific prognostic statements, it is not difficult for clinicians (or students) to recognize that some disorders tend to be more debilitating than others, more chronic or prone to relapses, or more responsive to treatment or to certain positive or negative life circumstances.

So clinicians can build upon the basic information in the *DSM* to improve their prognoses. For example, information about the client's level of social support and subjective distress can alter a prognosis. Relatively high levels of support and moderate, but not debilitating, distress improve prognoses whereas relatively low levels of support and intense distress worsen them. A prognosis can also be influenced by such client factors as impulsivity and coping style, resistance to therapist suggestions, treatment setting (e.g., inpatient versus outpatient), the "fit" between client and therapist (e.g., shared values or cultural background), and the client's beliefs about treatment (Hundt et al., 2013). The degree to which a prognosis should be adjusted by these or other factors should be guided by empirical research evidence about the degree to which each factor has been shown to influence treatment outcomes in similar cases.

PREDICTING FUTURE PERFORMANCE Clinicians are sometimes asked by businesses, government agencies, police and fire departments, and the military to help them select people who are most likely to perform well in certain jobs. In such cases, the clinician must first collect and/or examine descriptive assessment results to provide data on which to base predictions and selections. This step is critical and often underappreciated by those who believe psychologists should be able to make predictions in any domain simply on the basis of general psychological training. In order to know how someone will perform in a given job or situation, psychologists must have empirical evidence about which characteristics reliably predict which performances. That means that for each job or domain, psychologists have to do their homework; they cannot rely on their assumptions or general clinical judgment. A classic example of how descriptive and predictive assessment can overlap was provided by Henry Murray's use of specialized tests, interviews, and observations to select soldiers who would be the most successful spies, saboteurs, and other behind-enemy-lines operatives during World War II (Office of Strategic Services, 1948). Murray's assessment program was so comprehensive that it took several days to complete and measured everything from intelligence to skill at planning murder.

Similar, though less extensive, prediction-oriented clinical assessment programs also appeared in large-scale postwar screening programs designed to select civilian and military employees (Institute of Personality Assessment and Research, 1970), graduate students in clinical psychology and psychiatry (Holt & Luborsky, 1958; Kelly & Fiske, 1951), and Peace Corps volunteers (Colmen, Kaplan, & Boulger, 1964). Because such assessment programs influence decisions affecting large numbers of people, they must be evaluated not only for their predictive validity but also for their impact on the people being assessed and on the organizations that utilize them.

PREDICTING DANGEROUSNESS Predictions of dangerousness, long a part of clinical practice and research, are often called forensic evaluations. For instance, a clinician might be asked to evaluate whether an eighth grader who brought a handgun to school represents a continuing homicide risk and should be educated in a secure facility instead of in a regular school (Vincent, 2006). Clinical researchers might develop instruments designed to predict the likelihood that an adult sex offender will commit another offense if released (Langton et al., 2007).

Anna is a 20-year-old university student. Her parents became concerned about her mental state after two phone conversations during which Anna stated she was not leaving her dorm, was not communicating with others, and not attending class. They traveled 200 miles to see her and following a long and sometimes acrimonious discussion, convinced Anna to seek professional help. Anna first went to the university counseling center, where the results of an initial assessment suggested mixed anxiety and depression. Further information revealed, though, that Anna had recently shown signs of mania, so a diagnosis of bipolar I was suggested. During the interview, she was basically cooperative, but poorly groomed and at times agitated or despondent. She said that she probably needed help. Because the university did not have facilities to handle a case as potentially severe as hers, Anna agreed to be transferred to the city hospital's inpatient unit. During an intake interview there, a social worker noticed several scratch marks on her wrists that were of apparently recent origins. She admitted inflicting them, but said they "weren't that big a deal." In addition to finalizing a diagnosis and attempting to engage Anna in planning treatment, the next steps for the clinical staff was to determine Anna's risk of suicide and to take appropriate precautions if it is high.

As in the case of Anna, presented below, predictions become especially harrowing because they involve life-or-death situations, either for the potentially suicidal client or for those whom the client might harm upon release from a mental hospital. How accurate are such predictions?

In fact, clinical psychologists often find it difficult to predict dangerousness accurately (Coid et al., 2011; Fazel, Singh, Doll, & Grann, 2012; Hilton, Harris, & Rice, 2006). One reason is that the *base rate*, or frequency with which dangerous acts are committed in any group of people, is usually very low. The following example shows why this is important.

Assume that a clinician is 80% accurate in predicting homicidal behavior. Assume also that the base rate for homicide in the population the clinician examines is 10 murders per 10,000 people. The accuracy of predictions can be evaluated in terms of the pattern of four possible outcomes. If the clinician predicts dangerousness and the person indeed behaves dangerously, the outcome is called a *true positive*. If the clinician predicts that there is no danger, and the person does not behave dangerously, we have a *true negative* outcome. If the clinician predicts no danger, but the person does act dangerously, a *false negative* outcome occurs. Finally, when the clinician predicts dangerous behavior but dangerous acts do not occur, we call it a *false positive*. As indicated in Table 3.4, the clinician would correctly predict who 8 of the 10 murderers would be. However, if the clinician is 80% accurate, he would incorrectly categorize 20%, or 1,998, of the other 9,990 persons as potential murderers even though none of them would commit violence. So the 8 true positives must be viewed in light of the 1,998 false positives in which the clinician predicted homicides that did not occur.

In this example, then, the clinician's *positive predictive power* to identify murderers would be woefully low (less than 1% of predicted murderers would actually kill someone), while the accuracy of predicting nondangerousness, called *negative predictive power*, would be greater than 99.9%. To maximize true positives, then, the clinician could predict that no one will commit a murder, a prediction that would be correct 99.9% of the time. However, all the errors would be false negatives, and society usually believes that such mistakes (i.e., predicting that dangerous people are safe) are more serious than false positives (i.e., predicting that safe people are dangerous).

To err on the side of caution, then, clinicians tend to overpredict dangerousness. Of course, if there are no serious consequences to false-positive errors, overprediction is not a problem. But typically, errors in either direction can have significant consequences. Imagine, for instance, trying to predict which young adults are likely to engage in school violence, a task that takes on special importance in the wake of the murderous rampages that have taken place at Columbine High School, Sandy Hook Elementary School, and other places. School shootings are extremely low-base-rate events, so trying to identify at-risk individuals will inevitably result in many false-positive

TABLE 3.4 Measuring the Accuracy of Clinical Predictions of Dangerousness

Clinician's Prediction	Ultimate Outcome	
	Homicide	No Homicide
Homicide	8 (true positives)	1,998 (false positives)
No homicide	2 (false negatives)	7,992 (true negatives)

errors. For every correctly identified killer, dozens or hundreds of other youths are at risk of being incorrectly stigmatized with the same label (Cunningham, Sorensen, & Reidy, 2009; Mulvey & Cauffman, 2001).

It is tempting to dismiss the effects of low base rates on prediction—after all, they are “only statistics”—and to cling to the belief that when push comes to shove, the “experts” will be able to accurately predict school violence, suicide, or other violent acts. But experts do not cling to such beliefs. They know that the base-rate statistics reflect how difficult it is for clinicians (and others) to predict dangerousness accurately (Hart & Cooke, 2013). The American Psychiatric Association has taken the position that psychiatrists have no special knowledge or ability that allows them to accurately predict dangerous behavior (1983).

Those who attempt to predict dangerousness do try to minimize errors by following certain guidelines. First, clinicians usually prefer to make predictions about level of risk (e.g., high, moderate, low) rather than about whether a particular individual will or won't commit a particular act. Second, they use validated assessment instruments designed to predict the behaviors in question. There are several instruments for predicting dangerousness, for example, and though none of them have better than a modest predictive accuracy (Yang, Wong, & Coid, 2010), their results appear to be better guides than making guesses on the basis of intuition and limited experience. Researchers seeking to improve prediction of dangerousness increasingly focus on combining assessment evidence from four domains: (a) the defendant's dispositional tendencies, such as anger or impulsiveness; (b) clinical factors, such as evidence of mental or personality disorders; (c) historical factors, especially a record of violence; and (d) contextual factors such as the strength of social support from family and friends (Cassel & Bernstein, 2007).

So in cases like Anna's, a clinician will ask clients about their suicidal thoughts, their desire for death (Baca-Garcia et al., 2011), whether they have already made efforts at suicide, and if so, what kind and how often (Ougrin, et al., 2011). They will also ask clients for their own estimates of their potential for suicide (Peterson, Skeem, & Manchak, 2011). The responses to each of these questions can then be used to adjust upward or downward the estimated risk of suicide derived from a validated suicide screening questionnaire. In addition, clinicians refer to published literature on evidence-based risk assessment and prevention in an effort to improve the accuracy of their predictions of dangerous behavior (e.g., Pompili & Tatarelli, 2011).

SECTION SUMMARY

Clinical assessment goals include diagnostic classification, description, treatment planning and evaluation, and prediction. Diagnosis is typically based on the latest version of the Diagnostic and Statistical Manual of Mental Disorders, the DSM-5, though alternative diagnostic systems are being considered to remedy perceived shortcomings of the current system. Description involves a fuller characterization of a client than is typically available from a diagnostic classification. Both diagnosis and description can be used to help plan treatments. Assessments can also be used to measure treatment effectiveness and to demonstrate it to others. Finally, clinicians are sometimes asked to make predictions about future behavior, including clients' dangerousness, often a difficult task.

CLINICAL JUDGMENT AND DECISION MAKING

SECTION PREVIEW

Clinicians make judgments by combining information from different sources. In doing so, they can rely on empirically based methods of decision making, on clinical experience and intuition, or on a combination of the two. While clinical intuition is essential in many situations (e.g., spontaneous events that occur during the course of therapy), research shows that clinical psychologists have no special capacity for intuition: they are prone to the same kinds of error as are other human beings when relying on impressionistic thinking. So wise clinicians incorporate empirically based, actuarial models into their judgments in situations when such models are available (e.g., prediction of dangerousness).

The mass media have long cast clinical psychologists as experts who can astutely translate obscure signs into accurate statements about a person's past, present, or future. As in the case of violence, some events are particularly difficult to predict. But whether a clinician is trying to judge the likelihood of a client attempting suicide, a stalker harming a victim, a child being abused by a parent, a brain-injured client succeeding in a new job, or a particular form of psychotherapy alleviating a sufferer's anxiety, all predictions involve clinical judgment.

Clinical Intuition

As our discussion of prediction of dangerousness suggests, empirical research does not support the idea that clinicians have special inferential capabilities. Donald Peterson made this point forcefully 45 years ago: "The idea that clinicians have or can develop some special kinds of antennae with which they can detect otherwise subliminal interpersonal stimuli and read from these the intrapsychic condition of another person is a myth which ought to be demolished" (Peterson, 1968, p. 105). There is now a large body of research which has addressed the accuracy of clinical intuition by comparing clinical judgments with judgments made by laypersons or with judgments based on statistical or mechanical models. The clear result is that clinical intuition too often underperforms (Ægisdóttir et al., 2006; Goldberg, 1959; Grove, Zald, Lebow, Snitz, & Nelson 2000; Hanson & Morton-Bourgon, 2009; Meehl, 1957, 1965; Spengler et al., 2009).

Why do clinicians not make better clinical judgments, even after years of training and experience? Apparently, being human, clinicians are prone to the same cognitive habits and biases that can lead to error in anyone's information processing. For example, clinicians' judgments, like other humans' judgments, are more prone to error when they rely too heavily on experiences that are recent or remarkable enough to make them especially available to recall. The bias created by this *availability heuristic* can be seen, for example, when people are asked to judge which of the following is more common

7-letter English words ending in *-ing* _ _ _ _ _ i n g

or

7-letter English words with *n* in the sixth position _ _ _ _ _ n _

People often choose the first option, simply because *-ing* words are much easier to bring to mind than words with *n* in the sixth position, but in fact, words in the latter category are more common. They include all *-ing* words, as well as all 7-letter words that end in *nt* (e.g., *payment*), or *ne* (e.g., *biplane*), and so on. Consider the tragic and highly publicized mass shootings that took place in 2012 at an Aurora, Colorado movie theater and at the Sandy Hook Elementary School in Connecticut. Shocking incidents like these understandably lead people to think that multiple murders are quite common. The truth, though, is that while there have been 62 mass shootings since 1983, individual murders are far more frequent. FBI figures show that 8,583 people were shot to death in 2011 alone (Knox, 2013). Nevertheless, judgments about the probability of vivid, highly memorable events can easily become inflated in the minds of the public, clinicians included. In clinical situations, then, a psychologist may overemphasize the likelihood of diagnoses that come most easily to mind and thus misdiagnose (Garb, 1996).

Memorable clinical "folklore" can create *illusory correlations* (Chapman & Chapman, 1967) causing clinicians to draw false inferences from assessment data (Krol, DeBruyn, & van den Bercken, 1995; Lewis, 1991). Thus, some clinicians see paranoid tendencies in clients who draw large eyes on figure-drawing tests, even though there is no firm empirical evidence to support this association (Golding & Rorer, 1972). Like the first impressions all people form, clinicians also tend to display an *anchoring bias* in which they establish their views of a client more on the basis of the first few pieces of assessment information than on any subsequent information (Tutin, 1993). Anchoring bias can also influence clinicians to let assessment information coming from certain sources (e.g., a parent's report of a child's behavior) outweigh any other information they receive (McCoy, 1976). The problem with anchoring bias is that we hold too firmly to first impressions and do not make sufficient adjustments when new data warrants doing so. If anchoring bias combines with *confirmation bias*—the tendency to interpret new information in line with existing beliefs—the clinician may ignore contradictory evidence, discount its validity, or even distort it to fit initial impressions (Strohmer & Shivy, 1994).

Numerous studies also suggest that having larger amounts of assessment information may increase clinicians' confidence about their inferences, but it does not necessarily improve the

accuracy of those inferences (Einhorn & Hogarth, 1978; Garb, 1984; Kleinmuntz, 1984; Rock, Bransford, Maisto, & Morey, 1987). That's because additional pieces of information can contribute additional error—the less precise the predictor, the greater the error contributed. Clinicians' judgments may also be prone to error because they misremember information or fail to get important information. Like other people, clinicians tend to remember their successes more clearly than their failures and so may remain wedded to incorrect inference tendencies (or invalid assessment methods) simply because they think of them as valid (Garb, 1989). Indeed, clinicians may not get enough accurate feedback about their successes and failures.

CLINICAL AND STATISTICAL PREDICTION The sobering body of research that has challenged the image of clinicians as experts who can consistently predict future behavior has led some to question whether clinical predictions would be more accurate if they were based upon formal, statistical data-processing methods rather than on expert clinical judgment. *Statistical prediction* (also called *actuarial prediction* or *mechanical prediction*) involves inferences based on probability data and formal procedures for combining information, all usually derived from research. *Clinical prediction*, as we have seen, involves inferences based primarily on a practitioner's training, assumptions, and professional experiences. Which is better? As noted above, there is a large body of research on this question, including numerous meta-analyses that compare clinical and statistical prediction head to head.

When clinicians interpret assessment data informally, make recommendations based on these interpretations, and do research on such activities, it is usually because they believe that they, and perhaps most members of their profession, are good at these tasks. Thus it came as a shock when Paul Meehl's 1954 review of 20 studies comparing formal versus informal inference methods found that, in all but one case, the accuracy of the statistical approach equaled or surpassed that of the clinical approach. Later, even the sole exception to this surprising conclusion was called a tie, and as additional research became available, the superiority of the statistical method of prediction was more firmly established (Dawes, Faust, & Meehl, 1989; Meehl, 1957, 1965; see Table 3.5).

In the years since Meehl's review, numerous published responses have appeared, many of which pointed out methodological defects in some of the studies that could have biased results in favor of statistical procedures. Frederick C. Thorne (1972) put it this way: "The question must not be what naive judges do with inappropriate tasks under questionable conditions of comparability with actual clinical situations, but what the most sophisticated judges can do with appropriate methods under ideal conditions" p. 44. Still, the furor over Meehl's conclusions could not negate the fact that inference based on subjective, clinical methods is not as accurate as it was assumed to be, even on many clinically relevant tasks (Dawes, 1994; Hanson & Morton-Bourgon, 2009). Perhaps the strongest evidence comes

TABLE 3.5 Summary of Outcomes of Studies Comparing Clinical and Statistical Prediction

Source	Number of Studies Reviewed	Variables Predicted	Outcome		
			Clinical Better	Statistical Better	Tie
Meehl (1956)	20	Success in school or military; recidivism or parole violation; recovery from psychosis	1 ^a	11	8
Meehl (1957)	27	Same as above, plus personality description; therapy outcome	0	17	10
Meehl (1965)	51	Same as above, plus response to shock treatment; diagnosis label; job success and satisfaction; medical diagnosis	1 ^a	33	17
Grove Zald, Lebow, Snitz, and Nelson (2000)	136	Same as above, plus marital satisfaction; success on psychology internship; performance in medical school; and others	8	63	65
Ægisdóttir et al. (2006)	51	Brain impairment; personality; length of hospital stay; diagnosis; adjustment or prognosis; violence or offense; IQ; academic performance; if an MMPI profile was real or fictional; suicide attempt; sexual orientation	5	25	18

^aLater called a tie.

from meta-analyses comparing clinical and statistical prediction. For example, Grove and colleagues, (Grove, Zald, Lebow, Snitz, & Nelson, 2000) conducted a meta-analysis of 136 studies in which psychologists predicted criteria within their areas of expertise, including psychotherapy outcome, future criminal behavior, fitness for military service, marital satisfaction, psychiatric diagnoses, success on psychology internships, and several others. The results were consistent with previous studies: statistical/mechanical prediction outperformed clinical prediction overall, regardless of the type of judges, judges' experience, type of data being combined, or the design of the study. The advantage for statistical prediction was not large—roughly 10% on average—and in many studies mechanical and statistical prediction were essentially the same. In the small percentage of studies (6 to 16%) where clinical prediction outperformed mechanical prediction, the authors found no pattern of variables that reliably distinguished when or why clinical prediction was superior.

Similar results were obtained in two other meta-analyses, one which examined judgments in a variety of domains, and another which focused on predictions of recidivism among 45,398 sexual offenders (Ægisdóttir et al., 2006; Hanson & Morton-Bourgon, 2009). Both found a larger effect size for statistical prediction. In these studies, the difference in accuracy was not large, about 10–20%, and it varied by type of prediction (e.g., statistical models more clearly outperformed when predictions involved dangerousness; see Hilton, Harris, & Rice, 2006), setting, type of statistical formula used, and amount of information available to clinicians.

In short, there is considerable evidence that too few clinicians have appreciated the limits of clinical intuition. This makes them vulnerable to criticisms from those within and outside their profession (Faust & Ziskin, 1988; Ridley & Shaw-Ridley, 2009; Tavis & Aronson, 2007). Years ago, Thorne (1972) suggested that “clinicians must become much more critical of the types of judgments they attempt to make, the selection of cues upon which judgments are based, and their modes of collecting and combining data” (p. 44). That statement still applies today. Based on years of research on the issue of statistical versus clinical prediction, the following conclusions seem reasonable:

- Statistical/actuarial prediction generally outperforms clinical prediction.
- The superiority of statistical prediction is most evident in predicting violence and other low-base-rate events.
- The overall advantage for statistical prediction is modest, and occasionally clinical prediction does as well or slightly better, though not in any particular pattern of circumstances.
- Practicing clinicians typically underutilize and undervalue actuarial prediction methods.

Improving Clinical Judgment

Clinicians find it discouraging that their inferences, honed by training and experience, are often inferior to mechanical decisions that can be arrived at by anyone who can read and has the relevant assessment data to plug into a formula. Yet, if one takes a broader view, the generally superior performance of actuarial models, particularly for low-base-rate events such as violence, need not be seen as a sign of professional failure. A great deal of clinical experience and research typically goes into the development of useful actuarial models—we have always relied on research to separate the wheat from the chaff in clinicians' favored theories. Further, we don't denigrate meteorologists or stock brokers when they rely on statistical models and computer-generated forecasting. Nor do we distrust physicians who refer to databases and formal decision models to help them diagnose unusual medical conditions. Indeed, though some people are willing to accept intuitive, “from-the-hip” judgments from a doctor, we suspect that most appreciate the thoroughness and professionalism of physicians who take advantage of the latest research and technology in reaching a diagnosis.

At the same time, it is important to recognize that clinical experience and judgment are not invariably error prone. With more years of experience, clinicians' judgments tend to improve, though modestly at best. Spengler et al. (2009) conducted a meta-analysis of 75 judgment studies, involving 4,607 clinicians, and found that accuracy increased by about 13% with experience, regardless of other factors. Just *how* experience improves judgment is not entirely clear. We suspect it is not merely the repeated use of clinical intuition, but rather the repeated use of intuition

combined with timely feedback about the accuracy of predictions. More experience may also be associated with more time to learn about research findings on clinical decision processes, including the numerous factors that can bias clinical judgment, which in turn may make clinicians less vulnerable to those factors.

The preceding discussion suggests that clinical psychologists should be familiar with the extensive research literature on clinical judgment and on clinical versus statistical prediction, and that they should be guided by it. What determines whether clinicians incorporate statistical models in their clinical decisions? One survey of 491 members of the American Psychological Association's Division 12 (Clinical Psychology) found that some of them dismissed research on clinical judgment as conceptually misguided or irrelevant to real decisions (Vrieze & Grove, 2009)—an overgeneralization, given the diversity of studies, that risks “throwing the baby out with the bath water.” However, the main factor that determined whether clinicians made use of statistical models in their clinical decisions was simply the extent to which that topic was discussed when they were in graduate school.

A number of proposals have been made to make clinical reasoning a more explicit, rather than implicit, part of training in the hopes of teaching health professions to avoid the more common sources of inference errors (Ridley & Shaw-Ridley, 2009; Spengler et al., 2009). By most counts, this change is needed. In accredited clinical psychology programs, issues related to decision making are more likely to be covered in non required courses, such as cognitive psychology, and only 9% of programs required courses that contained a significant component on decision making (Harding, 2007).

Clinical decisions can be difficult. Whether they involve anticipating a client's response to treatment, predicting dangerousness, determining competency to stand trial, or something else, such judgments require envisioning distant events that can be affected by numerous and often unforeseen variables. With experience, clinicians might improve their judgment, but because clinicians are not blessed with special intuitive powers, and because they often receive limited formal instruction in clinical decision making, experience-based improvements alone are likely to be small. Greater improvements in judgment are likely to occur if clinicians develop, through graduate training and/or continuing education, an appreciation of the common errors of judgment to which humans are prone. The appropriate use of formal models can improve specific kinds of decision making.

SECTION SUMMARY

Clinicians are often asked to make inferences that go well beyond observable behavior. When they rely on their beliefs about unobserved phenomena associated with behavior, or about presumed underlying mechanisms or psychic structures that cause behavior, their inferences can appear rich and insightful. However, those same inferences are also more prone to bias and inaccuracy. Years of research suggest that clinicians should be especially aware of the limitations of clinical judgments and employ assessment instruments and decision-making procedures that have the best track records—doing so is the only reliable way to improve one's “clinical intuition.” Although prediction is often a difficult task, accuracy can be improved if clinicians seek out the best available evidence and follow established guidelines.

PSYCHOMETRIC PROPERTIES OF ASSESSMENT INSTRUMENTS

SECTION PREVIEW

The assessment goal is the most important factor in determining a clinician's choice of assessment tools, but a number of other factors also affect selection. Among the most important of these are the psychometric properties of instruments. Reliability, validity, standardization, utility, and the like are the “credentials” of assessment instruments. When deciding what tools to use in assessment or research, clinical psychologists must consider these credentials carefully.

Sound clinical judgments depend on the soundness of the measures used to help make those judgments (Ayearst & Bagby, 2010). Assessment instruments of higher quality are those that carefully conducted research has shown to be reliable, valid, and useful. Below we briefly consider the dimensions that determine an instrument's psychometric quality. In the next two chapters, we summarize psychometric information on several of the more popular assessment instruments.

Reliability

Reliability refers to consistency in measurement or to agreement among different judges or raters. It can be evaluated in several ways. If the results of repeated measurements of the same client are very similar, the assessment procedures are said to have high *test-retest* reliability, analogous to a bathroom scale that shows the same weight when someone steps off and back on again.

Another way to evaluate reliability is to examine internal consistency. If data from one part of an assessment, such as odd-numbered test items, are similar to data from other parts, such as even-numbered items, that assessment is said to be *internally consistent*. This dimension has sometimes been called split-half reliability. Finally, *interrater* reliability is measured by comparing the conclusions drawn by different clinicians using a particular assessment system to diagnose, rate, or observe the same client. When clinicians judging the same set of assessment results arrive at significantly differing conclusions, interrater reliability is low. The more they agree, the higher the interrater reliability of the instrument.

Interrater reliability tends to be higher when clinicians make judgments about a diagnosis that has relatively clear diagnostic criteria (e.g., panic disorder). In such cases, they are all using the same interpretive rules. Interrater reliability tends to be lower when clinicians make judgments about a diagnosis with less precise criteria (e.g., antisocial personality disorder) and use unstructured interviews and projective tests—assessments with ambiguity in the possible meanings of information obtained. What about the interrater reliability of *DSM-5* disorders? Preliminary studies suggest that, as with previous *DSMs*, some are very high (e.g., posttraumatic stress disorder) and some unacceptably low (e.g., generalized anxiety disorder) (Freedman et al., 2013).

To make the best use of reliability data, clinicians must examine the published research on specific assessment methods or instruments. They can also consult books or manuals that address assessment issues. In Chapters 4 and 5, we present reliability information on a number of the more common assessment instruments that clinicians use.

Validity

The *validity* of an assessment method reflects the degree to which it measures what it is supposed to measure. Like reliability, validity can be evaluated in several ways. The *content* validity of an assessment method is determined by how well it taps all the relevant dimensions of its target. An interview-based assessment of depression that includes questions about sad feelings but not about their duration or cause would have low content validity. *Predictive* validity is measured by evaluating how well an assessment forecasts events, such as violent behavior or suicide attempts. When two assessment devices agree about the measurement of the same quality, they are said to have *concurrent* validity. Predictive and concurrent validity are subtypes of *criterion* validity, which measures how strongly an assessment result correlates with important independent criteria of interest.

Finally, there is *construct* validity (Cronbach & Meehl, 1955). To oversimplify somewhat, an assessment device has good construct validity when its results are shown to be systematically related to the construct it is supposed to be measuring. Psychologists evaluate construct validity by determining whether a test or other assessment method yields results that make sense in light of some theory about human behavior and mental processes. For example, scores on a measure of anxiety should increase under circumstances thought to increase anxiety (e.g., facing major surgery). If no change occurs, the measure's construct validity is suspect. Fully evaluating construct validity requires numerous studies and an elaborate set of statistical analyses (Campbell & Fiske, 1959).

Validity is related to reliability because an assessment device cannot be any more valid than it is reliable. However, the validity of an instrument is not guaranteed just because the instrument is reliable. Consider a situation in which 50 people use their eyes to assess the gender of a man skilled in female impersonation. All 50 observers might agree on their judgment that the man is a female, but they would all be wrong. In this case, visual assessment had high interrater reliability but very poor criterion validity.

It is important to remember that reliability and validity are matters of degree, not all-or-none propositions. The question is always how much imprecision is tolerable in an assessment. For instance, if a bathroom scale varied by only a few ounces in test-retest procedures, we would likely consider the instrument reliable for our purposes, but if a postal scale varied by that much, we would consider it unreliable. How much error one tolerates depends on the goals and potential uses of the assessment, on the availability of alternative measurement instruments, and on practical constraints such as time and resources.

Remember, too, that the validity of an instrument must always be viewed in relation to the purposes for which the assessment instrument is to be used. For example, a test might be a valid measure of typing skill but an invalid measure of aggressiveness. Test validity can reasonably be assessed only when tests are used for the purposes for which they were designed. So clinicians should compare reliability and validity data only across tests designed for the same purpose and then select the better-performing instrument, all other considerations being equal.

Standardization

When we say that a test or other assessment instrument is standardized, we mean that the designers of the test have given it to a large, representative sample of persons and analyzed the scores. Doing so gives the designers information about what the average score is in a population—the average mathematics score for 7-year-old children, for instance. It also gives information about the variance of scores on individual items or subtests. Without that information, there would be no way to determine whether a particular score on a test was average, below average, or above average.

Key considerations for a clinician are whether the size of the standardization sample was large enough and representative enough. If it wasn't, clinicians should have limited confidence in a test. Another key consideration is whether the particular client being tested is similar enough to the sample on which that test was standardized. Suppose, for instance, a clinician suspects a client might have a somatoform disorder and wants to conduct a structured interview—a series of prescribed questions—to help make a diagnosis. However, the client is an immigrant from China. If the structured interview has not been validated on Chinese or other Asian populations, the client's responses might not be interpretable on the basis of the standardization of the test.

Bandwidth-Fidelity Issues

Clinicians' assessment choices are further guided by their attempts to resolve the *bandwidth-fidelity* dilemma (Shannon & Weaver, 1949). Just as greater bandwidth is associated with lower fidelity in broadcasting, clinicians have found that, given limited time and resources, the more extensively they explore a client's behavior, the less intensive each aspect of that exploration becomes (and vice versa). The breadth of an assessment device is thus referred to as its *bandwidth* and the depth or exhaustiveness of the device as its *fidelity* (Cronbach & Glesser, 1964). This issue is also called the depth-breadth issue. If, during a 2-hour interview, for example, a clinician tries to cover a long list of questions, the result would be superficial information about a wide range of topics (broad bandwidth, low fidelity). If the time is spent exploring the client's early childhood memories, the result would be a lot of detailed information about only one part of the client's life (narrow bandwidth, high fidelity).

Accordingly, clinicians must seek assessment strategies and measurement tools that result in an optimum balance of bandwidth and fidelity. The questions, levels of inquiry, and assessment techniques that will be useful in identifying stress-resistant executives differ substantially from those that will help detect brain damage in a 4-year-old child.

OTHER FACTORS AFFECTING ASSESSMENT CHOICES

SECTION PREVIEW

A variety of other factors affect assessment choices, including the clinician's theoretical orientation and experience with various assessment instruments, as well as the assessment context, and cultural factors.

Clinicians' Experience and Theoretical Orientation

Clinical psychologists may tend to use, or avoid, particular assessment methods because the methods were either emphasized or criticized by faculty in their graduate training program. Similarly, those who find certain measurement tactics tedious or unrewarding tend to seek answers to assessment questions through other procedures with which they are more comfortable. Ideally, a clinician's comfort with an instrument should not be the primary criteria on for selection, but sometimes clinicians become comfortable—too comfortable, probably—with certain methods of assessments that are not optimal for the task. Such clinician-specific factors help explain why some assessment methods continue to be used by some clinicians even when research evidence fails to support their reliability or validity.

As just described, clinicians' differing theoretical orientations direct them to pursue certain questions and concerns; in effect, orientation provides an outline for assessment. Thus, psychodynamically oriented case study outlines tend to include questions about unconscious motives and fantasies, ego functions, early developmental periods, object relations, and character structure (e.g., Gabbard, 2010). Cognitive-behavioral case study outlines (e.g., Beck, 2011) focus on client skills, habitual thought patterns, and the stimuli that precede and follow problematic behavior. Humanistically oriented clinicians are less likely to follow a specific assessment outline; indeed, they are likely to see assessment as a collaborative process in which they seek to understand with each client how that client perceives himself or herself and the world (Fischer, 2001; Gorske, 2008).

Ideally, a clinician's assessment outline will be broad enough to provide a general overview of the client yet focused enough to allow coverage of all the more specific questions that the clinician wishes to address. The outline guides production of the assessment report—the organized presentation of assessment results. Let's consider two sample assessment outlines representing two theoretical approaches to clinical psychology.

Psychodynamic Assessment Outline

Patient information

- Personal
- Demographic

Presenting problem(s)

History

- Attachment, early family relations
- Medical and psychiatric
- Interpersonal, social, cultural

Psychological testing

- Projective tests
- Objective tests

Mental status

DSM diagnosis

Ego strengths

Defense mechanisms

Relationship boundaries

Self-concept and self-esteem

Cognitive–Behavioral Assessment Outline

Client information

- Personal
- Demographic

Presenting problem(s)

History of presenting problems

Client understanding of problems

Coping responses

- Behavioral
- Cognitive

Psychiatric history

Family, social, educational history

Psychological testing

- Objective

Daily activities

- Mood
- Work and social functioning

Client goals, motivation, and expectations of treatment

Humanistically Oriented Assessment

Humanistically oriented clinicians have suggested assessment alternatives that differ substantially from those of the cognitive-behavioral approach (e.g., Fischer, 2001; Gorske, 2008). Some of them have argued against assessment on the grounds that such procedures are dehumanizing, take responsibility away from clients, and threaten the quality of clinician–client relations (Rogers, 1951). Other humanistic psychologists raise the possibility that assessment data collected through traditional means can be useful if they are processed in line with humanistic principles (Fischer, 1989). For example, test results can be viewed as clues to how a client looks at the world, and conducting those tests can provide opportunities for the clinician and client to build their relationship (Dana & Leech, 1974). Fischer, for instance, argues that assessments using standardized, empirically validated instruments can be both scientific and therapeutic when assessments are conducted interactively between therapist and client. She gives examples of discussing *DSM-IV* criteria and possible interpretations of MMPI-2 profiles with clients, arguing that such collaboration not only provides additional data for assessments but avoids objectifying the client (Fischer, 2001; see also Finn, 1996).

The Assessment Context

The assessment choices that clinicians make are dictated not only by their goals, the quality of their instruments, and the time and resources available, but also by the contexts, or settings, in which assessments are conducted. Common settings include general medical and psychiatric facilities, private or community psychological clinics, jails, prisons, forensic (legal) situations, schools and other educational institutions, and the like. Each type of setting influences the nature of the referral questions asked, the kinds of assessment instruments expected or preferred, and the style of reporting that is most appropriate or most often requested (Groth-Marnat, 2009). For instance, clinicians who are unfamiliar with legal settings may initially find it disconcerting to have their clinical judgments, and even their professional qualifications, challenged. The language used in legal contexts is also different from that used in psychological contexts, and it is incumbent on the clinician to translate psychological terms that may not be understood by non psychologists into terms that those in the legal profession can understand. Similarly, educational settings bring with them their own preferred theories, terms, and methods of practice. So, as mentioned earlier, clinicians must not only clarify the referral question before conducting assessment, they must also select context-appropriate instruments and present their conclusions in ways that are the most useful to the referral source.

Cultural Factors

A European American clinician employed by a community mental health system greets a Mexican American family referred by a school counselor. The daughter has been having difficulties in school—social isolation, academic difficulties, and signs of depression—and there are concerns that the father might be suffering from a mental illness. The family was reluctant, but they agreed to an intake visit. They are now sitting somewhat nervously as the clinician asks them to describe their understanding of why they are there. Their English is poor, and the clinician, anxiously holding a standard multi page assessment form in her hands, does not speak Spanish.

What should the clinician do in a situation like this? What steps could have been taken ahead of time to make the situation less awkward and more productive? What should the goals of this assessment be and what assessment instruments would best accomplish those goals? Will the assessment instruments be understood by the clients? Will the results be valid?

Multicultural competence is increasingly necessary for mental health professionals (Rosenberg, Almeida, & McDonald, 2012). With ever more diversity in the U.S. population, there is a much greater chance that clinicians will encounter clients whose cultural backgrounds and world views are significantly different from their own. In such cases, the initial challenge is quite basic: establish lines of communication and trust (Comas-Díaz, 2012). In the case example above, even if the language barrier is overcome—either by hard work on the part of all participants or by referral to a clinician who speaks Spanish—there still may be problems related to trust. Members of minority groups may differ from other clients in their expectations and beliefs about the goals of assessment and the extent to which clinical interventions might help them. So clinicians must take care to assure that assessment goals are not socially or culturally biased in a way that would make the process less valid for minority clients (Mitchell, Patterson, & Boyd-Franklin, 2011).

A good deal of multicultural assessment research is devoted to exploring whether assessment instruments (questionnaires, tests, structured interviews, etc.) are equally valid for different populations. This research is important because many psychological assessments were originally developed and normed on U.S. samples. Psychological tests that are not applicable for other groups—a finding that can only be established by careful cross-cultural research—may not only be inappropriate, but may lead to court decisions prohibiting their use for educational placement and other purposes (see Lambert, 1985, for discussion of the landmark *Larry P. v. Wilson Riles* case in California). When conducting assessments on clients from different cultures, clinicians must therefore be knowledgeable about how sociocultural factors can affect assessment results, as well as diagnosis and treatment (Lopez & Guarnaccia, 2000). We discuss cultural factors and cultural competencies in assessment further in Chapter 4, “Interviewing and Observation in Clinical Psychology,” and in Chapter 5, “Testing in Clinical Psychology.”

Core Competencies in Clinical Psychology Assessment

There are so many types of assessment that no clinician can be expected to have mastered them all. However, every clinical psychologist should possess a set of core competencies in assessment. While no single formal set of competencies is universally agreed upon, those suggested by professional organizations, employers, educational institutions, and clinical internship sites have many similarities (Krishnamurthy et al., 2004). We have drawn upon those similarities to construct the following list. Some competencies relate to the empirical and theoretical foundations of assessment (e.g., an understanding of personality or cognitive variables that tests measure, an ability to judge reliability and validity), while others relate to the use of specific assessments (e.g., interview formats, intellectual tests, personality tests, diagnostic assessment). Psychologists trained in assessment should be able to

- understand the theoretical, empirical, and contextual bases of assessment.
- evaluate the psychometric properties of assessment instruments.
- successfully administer and interpret instruments designed to assess cognitive functioning, behavioral functioning, and personality.
- conduct and interpret clinical interviews and behavioral observations.
- formulate appropriate *DSM* diagnoses.
- recognize the limitations and appropriate uses of assessment instruments for special populations (cultural and linguistic groups, physically challenged, etc.).

- integrate data from multiple assessment sources into empirically grounded conclusions.
- effectively communicate the results of assessments to others in written and spoken reports.
- understand and follow APA Ethics Code guidelines for assessment.

SECTION SUMMARY

The psychometric properties of various instruments such as reliability, validity, and bandwidth can affect a clinician's assessment choices. The clinician's theoretical orientation and personal experiences with various instruments can also affect choices. While theoretical approaches can be biasing if too rigidly held, they also provide useful guidelines for clinicians by directing data collection efforts toward discovery of selected, clinically relevant information. The settings in which an assessment takes place can affect how assessment is conducted and perceived, and cultural factors can play a significant role in assessment, especially when the clinician and client(s) do not share the same set of assumptions and understandings. Assessment involves numerous skills, and clinical competency in assessment is an ongoing process.

COMMUNICATING ASSESSMENT RESULTS

SECTION PREVIEW

Assessment results are communicated in an assessment report. Although it seems obvious, it is important to remember that reports should be clear, relevant to the goals of the assessment, and conveyed using language that will make them maximally useful to the consumers of the report.

If assessment results are to have maximum value, they must be presented in reports that are clear, relevant to the assessment goals, and useful to the intended consumer. Accordingly, clinicians must guard against problems that can make reports vague, irrelevant, and useless. Table 3.6 illustrates how an assessment outline—in this case, a cognitive-behavioral outline—is translated into an assessment report. Notice that it is sufficiently problem oriented to be used with clients seeking help, while also reminding the assessor to consider broader and less problematic aspects of a person's life.

Report Clarity

The first criterion for an assessment report is clarity. Without this basic attribute, relevance and usefulness cannot be evaluated. Lack of clarity in psychological reports is troublesome because misinterpretation of a report can lead to misguided decisions. Here is a case in point:

A young girl, mentally defective, was seen for testing by the psychologist, who reported to the social agency that the girl's test performance indicated moderate success and happiness for her in "doing things with her hands." Three months later, however, the social agency reported to the psychologist that the girl was not responding well. Although the social agency had followed the psychologist's recommendation, the girl was neither happy nor successful doing things with her hands. When the psychologist inquired what kinds of things, specifically, the girl had been given to do, he was told "We gave her music lessons—on the saxophone." (Hammond & Allen, 1953, p. v)

A related problem exists when the assessor uses jargon that may be meaningless to the reader. Consider the following excerpt from a report on a 36-year-old man:

Test results emphasize a basically characterological problem with currently hysteroid defenses. Impairment of his ability to make adequate use of independent and creative fantasy, associated with emotional lability and naivete, are characteristic of him. Due to markedly passive-aggressive character make up, in which the infantile dependency needs are continually warring with his hostile tendencies, it is not difficult to understand this current conflict over sexual expression. (Mischel, 1968, p. 105)

TABLE 3.6 An Assessment Report Based on a Cognitive-Behavioral Outline**Behavior During Interview and Physical Description**

Phil is a 20-year-old college student. At the first interview he seemed shy, soft spoken, and obviously uncomfortable about seeking therapy. Still he seemed willing and able to describe his problems and his feelings.

Presenting Problem:

A. *Nature of problem:* Anxiety in social situations, especially those involving public speaking, in which he feels that others are evaluating him.

B. *History:* Phil reports "always" being shy when in social situations and overly concerned with performance. He attributes the latter problem to his father, whom he describes as being harshly critical of him and who would "grade" the quality of his homework before he turned it in and would sometimes make him redo it until it met his exacting standards. He described his mother, too, as overly controlling, though more overtly affectionate than his father. Both parents compared him unfavorably to his younger brother, who was always a good student.

C. *Specific problematic situations:* Interaction with his parents, academic examinations, family gatherings, class discussions, meeting new people.

D. *Cognitive factors:* Phil's descriptions suggest that he has adopted unreasonably high standards, expecting himself to achieve perfection in academic and social situations, and believing that he must attain approval from everyone in order to be worthwhile.

E. *Dimensions of problem:* The client's social and evaluative anxiety are long-standing and occur in a wide variety of day-to-day situations.

F. *Consequences of problem:* In addition to creating social isolation and underperformance in academic exams, Phil's anxiety may be related to a history of gastrointestinal distress that requires him to take antacids on a regular basis.

Other Problems:

A. *Assertiveness.* Although obviously shy, Phil said that lack of assertiveness is no longer a problem with him. He feels that he

can stand up for himself better now than in the past, but this is something to explore further to determine if his assertiveness skills are adequate.

B. *Forgetfulness:* Phil describes himself as "scatterbrained" because he often forgets appointments, loses keys and other items, and even misses exams.

Personal Strengths:

Phil is intelligent, and appeared in the interview to be sensitive, friendly, and to have a good sense of humor.

Targets for Treatment:

Irrational expectations and self-statements in social-evaluative situations; possibly assertiveness.

Recommended Treatment Approach:

Progressive relaxation training to reduce and control anxiety, followed by cognitive restructuring, role-playing, and homework assignments to practice new self-talk and other cognitive skills.

Motivation for Treatment:

High.

Prognosis:

Very good.

Priority for Treatment:

High.

Client's Expectancies:

Phil says he wants to learn to eliminate his anxiety, and it appears he is sincere about this. He will probably be cooperative with the treatment plan.

The writer may understand the client, but will the reader understand the writer? Anyone not well versed in psychoanalytic terminology would find such a report mystifying. Even professionals may not agree on the meaning of the terms employed. Factors such as excessive length (or cryptic brevity), excessively technical information (statistics or esoteric test scores), and lack of coherent organization also contribute to lack of clarity in assessment reports (Olive, 1972; Wright, 2011).

Relevance to Goals

Although far less common today than in the past, clinicians may still be asked for "psychologicals" (usually a standard test battery and interview) without being told why assessment is being done. Under such circumstances, the chances of writing a relevant report are minimal. Unfortunately, there are other cases in which a report's lack of relevance is due mainly to the clinician's failure to keep established assessment objectives in mind.

Usefulness of Reports

Finally, one must ask if an assessment report is useful. Does the information it contains add anything important to what we already know about the client? Reports that present clear, relevant information that is already available through other sources may appear useful but have little real value. Such reports tend to be written when the assessor has either failed to collect new information or has not made useful statements about new data. In the former case, the clinician may have employed techniques that have low incremental validity (Sechrest, 1963). For example, a clinician may use psychological tests to conclude that a client has strong hostile tendencies, but if police records show that the client repeatedly has been arrested for assault, this conclusion doesn't add much to the clinical picture. In other instances, the assessor's report may have limited usefulness because it says nothing beyond what would be expected on the basis of base-rate information, past experience, and common sense.

Consider the following edited version of a report written entirely on the basis of two pieces of information: (a) the client is a new admission to a Veterans Administration (VA) hospital, and (b) the case was to be discussed at a convention session entitled "A Case Study of Schizophrenia."

This veteran approached the testing situation with some reluctance. He was cooperative with the clinician but mildly evasive on some of the material. Both the tests and the past history suggest considerable inadequacy in interpersonal relations, particularly with members of his family. It is doubtful whether he has ever had very many close relationships with anyone. He has never been able to sink his roots deeply. He is immature, egocentric, and irritable, and often he misperceives the good intentions of the people around him. He tends to be basically passive and dependent, though there are occasional periods of resistance and rebellion against others. Vocationally, his adjustment has been very poor. Mostly he has drifted from one job to another. His interests are shallow, and he tends to have poor motivation for his work. Also, he has had a hard time keeping his jobs because of difficulty in getting along with fellow employees. Although he has had some relations with women, his sex life has been unsatisfactory to him. At present, he is mildly depressed. His intelligence is close to average, but he is functioning below his potential. Test results and case history suggest the diagnosis of schizophrenic reaction, chronic undifferentiated type. Prognosis for response to treatment appears to be poor. (Sundberg, Tyler, & Taplin, 1973, pp. 577-579)

In generating this impressive but utterly generic report, the clinician relied heavily on knowledge of VA hospital residents and familiarity with hospital procedures. For example, as the case was to be discussed at a meeting on schizophrenia, and since schizophrenia diagnoses are common for VA residents, it was easy to surmise the correct diagnosis. Also, because it fits the "average" VA resident, the report was likely to be at least partially accurate. This bogus document exemplifies a feature of assessment reports that reduces their usefulness: overgenerality, or the tendency to write in terms that are so ambiguous they can be true of almost anyone. Documents laden with overly general statements have been dubbed "Barnum reports" (in honor of P. T. Barnum's maxim that there is a sucker born every minute), "Aunt Fanny reports" (because the statements could also be true of "my aunt Fanny"), or "Madison Avenue reports" (given that they "sell" well) (Klopfer, 1983; Meehl, 1956; Tallent, 1992). Such overly general material has the dual disadvantages of spuriously increasing the impressiveness of a report while actually decreasing its usefulness.

SECTION SUMMARY

While there is no universally "right" way to organize assessment data, the criteria of clarity, relevance, and usefulness may be more easily achieved by using an outline organized around the goals of assessment, the clinician's theoretical approach, and the assessment context. It is also worth repeating that the information sought in an outline, and the instruments used to gain that information, should be based on empirical research.

ETHICAL CONSIDERATIONS IN ASSESSMENT

SECTION PREVIEW

A number of ethical issues confront clinicians who perform assessments, especially when they become embroiled in acrimonious divorce or child custody cases, when they help determine whether clients are eligible for disability based on mental illness, or when they must decide what information to reveal about a client to third-party payers (e.g., insurance companies). Clinicians must know the limitations of the assessments they perform, and they must be clear in advance how those assessments are to be used. They should also be knowledgeable about how federal and state laws and the APA Ethical Principles of Psychology and Code of Conduct govern their behavior.

The process of collecting, processing, and communicating assessment data obviously gives clinicians access to sensitive information that the client might not ordinarily reveal to others. This places a heavy responsibility on the assessor to use and report this privileged information in a fashion that safeguards the client's welfare and dignity and shows concern for (a) how psychological assessment data are being used; (b) who should have access to confidential material; and (c) the possibility that improper or irresponsible interpretation of assessment information will have negative consequences for clients.

With these concerns in mind, clinicians must first be sure that their inquiries do not constitute an unauthorized invasion of a client's privacy. In order to know what constitutes authorized and unauthorized disclosure of information, clinicians must be familiar with more than the details of individual cases. They must also wrestle with the problem of who may have access to assessment data if they do not maintain sole control over them. When test scores, conclusions, predictions, and other information are communicated in a report, they may be misused by persons who see the report but are not qualified to interpret it. In such cases, not only is the client's privacy invaded, but the assessment data may create harmful outcomes for the client. Minimizing these problems is a major concern of public officials, government agencies, citizens groups, and private individuals.

Perhaps the best way to ensure that clinicians follow ethical assessment practices is for them to know the American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* (APA, 2000b), particularly Section 9, which deals with assessment. In addition, the APA offers guidelines for certain specialized assessment practices, such as the *Guidelines for Psychological Evaluations in Child Protection Matters* (APA, 1998). Psychologists should also know the *General Guidelines for Providers of Psychological Services* (APA, 1987) and *Standards for Educational and Psychological Testing* (APA, 1985). These guidelines reflect federal legislation, including the Equal Employment Opportunity Act (part of the Civil Rights Act of 1964), which prohibits discriminatory use of tests that have adverse impact on the selection of minority group job candidates, and the Civil Rights Act of 1991, which bans adjustment of test scores on the basis of race, color, religion, sex, or national origin (Sackett & Wilk, 1994). The guidelines must also be implemented in accordance with the regulations of the Individuals with Disabilities Education Act and the Americans with Disabilities Act.

Care should also be taken to assure that assessment goals are not socially or culturally biased such that certain clients (e.g., members of ethnic or racial minorities) are placed at a disadvantage (Malgady, 1996). For example, some psychological tests are alleged to be inappropriate for use with minority groups, leading to court decisions prohibiting their use for educational placement and other purposes (see Lambert, 1985, for discussion of the landmark *Larry P. v. Wilson Riles* case in California). When conducting assessments on clients from different cultures, clinicians must be knowledgeable about how sociocultural factors can affect diagnosis, assessment, and treatment (Lopez & Guarnaccia, 2000).

The guidelines for assessors are often just that—guidelines; they do not clearly tell psychologists what they should or should not do. Ethical decision making often involves taking into account various federal, state, and local laws as well as professional codes and individual concerns. Clinicians faced with difficult decisions can also consult with colleagues and seek guidance from the American Psychological Association about the best course of action. Ethical problems and standards associated with clinical psychology are considered in greater detail in Chapter 15, "Professional Issues in Clinical Psychology."

Chapter Summary

Clinical assessment is the process of collecting information to be used as the basis for informed decisions by the assessor or by those to whom results are communicated. Interviews, tests, observations, and life records serve as the main sources of assessment data in clinical psychology. The clinical assessment process includes five stages: clarifying the referral, planning data collection, collecting data, processing data, and communicating results. The methods and levels of inquiry in assessment tend to follow a case study guide that is shaped by assessment goals, clinicians' theoretical preferences and experience, and contexts. Selection of assessment methods is also guided by research on their reliability (consistency) and validity (ability to measure what they are supposed to measure), and depth versus breadth.

The goals of clinical assessment typically involve diagnostic classification, description, treatment planning, and prediction. Diagnostic classification normally employs the *DSM*. Description involves broader assessments of clients' personalities by looking at person-environment interactions. Assessment for treatment planning involves collecting information about how clients might respond to various treatment approaches. Predictions often involve personnel selection but sometimes focus on a client's potential for violence or suicide.

Unfortunately, clinicians have no unique intuitive power or special information-processing capacity, so the quality of their judgments and decisions about clients can be threatened by the same cognitive biases and errors that affect all human beings. Indeed, research on clinical judgment suggests that in many situations, clinicians can make their greatest contribution to assessment as collectors of information that is then processed by computer-based statistical formulae.

Assessments are driven primarily by the assessment goals, but other factors can also influence how assessments are conducted. The reliability, validity, and generality of assessment instruments should play an important role in a clinician's selection. The clinician's theoretical orientation and prior experience with instruments affect assessment choices, too, as does the context in which assessments are made (i.e., legal, educational, psychiatric hospital, psychological clinic).

The results of clinical assessment are presented in an organized assessment report, which should be clear, relevant to assessment goals, and useful to the intended consumer. These reports often reflect the theoretical approach taken by each clinician, but they should be constructed in a way that is maximally useful to the referral source and consistent with ethical practices.

Study Questions

1. What core competencies in assessment should clinical psychologists possess?
2. What are the steps involved in a psychological assessment?
3. What are the most common data collection techniques?
4. What are the general goals of assessment?
5. How do clinical psychologists make diagnoses?
6. What is the basic structure of the *DSM-IV-R* and how might it differ in *DSM-5*?
7. Discuss the major changes to the *DSM-5*.
8. What are some criticisms of the *DSM* and alternate proposals for clinical diagnosis?
9. What has research revealed about clinical psychologists' abilities to predict future violence?
10. What are clinical judgment and clinical intuition, and how can they be improved?
11. Why are reliability, validity, and bandwidth important considerations in selecting assessment instruments?
12. How does a clinician's theoretical orientation influence assessment?
13. How does the clinical context or setting influence assessment data collection and reporting?
14. What are the basic requirements for a good psychological report?
15. How might clients' cultural or ethnic backgrounds influence the results of assessments?
16. What types of ethical concerns and dilemmas can be raised when clinicians conduct assessments?

Web Sites

- APA Rights and Responsibilities of Test Takers: <http://www.apa.org/science/programs/testing/rights.aspx>
- APA Report of the Task Force on Test User Qualifications: <http://www.apa.org/science/programs/testing/qualifications.pdf>
- APA Strategies for Private Practitioners Coping With Subpoenas Compelling Testimony for Client Records or Test Data: <http://www.apa.org/about/offices/ogc/private-practitioners.pdf>
- For official updates on *DSM-5*: <http://www.dsm5.org/Pages/Default.aspx>

MOVIES

A Beautiful Mind (2001): Based on the real life of a gifted mathematician, this film shows the challenges of distinguishing between brilliance and mental illness.

Monster (2003): In a graphic film that depicts a prostitute who tortures and kills many of her clients, this film highlights the questions of dangerousness, responses to childhood trauma, and the fine line between psychological issues versus forensic issues.

MEMOIRS

Hurry Down Sunshine: A Father's Story of Love and Madness by Michael Greenberg (2008; New York: Vintage Books). This father's powerful account of his daughter's symptoms of schizophrenia brings the *DSM* alive.

Manic: A Memoir by Terri Cheney (2009; New York: Harper). In her detailed account of bipolar disorder, this author shows that many other comorbid disorders, such as substance abuse, can compound the challenges of treatment.

References

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugrman, A. S., Anderson, L., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist*, 34, 341–382.
- American Psychiatric Association. (1983). *Statement on prediction of dangerousness*. Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed.): Text revision (DSM-IV-TR)*. Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.): DSM-5*. Washington, DC: Author.
- American Psychological Association. (1985). *Standards for educational and psychological tests*. Washington, DC: Author.
- American Psychological Association. (1987). *General guidelines for providers of psychological services*. Washington, DC: Author.
- American Psychological Association. (1998). *Guidelines for psychological evaluations in child protection matters*. Washington, DC: Author.
- American Psychological Association. (2000b). *Current major field of APA members by membership status* (Table 3). Retrieved Dec 13, 2001, from <http://research.apa.org/member>.
- Antony, M. M., & Barlow, D. H. (Eds.). (2011). *Handbook of assessment and treatment planning for psychological disorders* (2nd ed.). New York, NY: Guilford.
- Ayearst, L. E., & Bagby, R. M. (2010). Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 23–61). New York, NY: Guilford.
- Baca-Garcia, E., Perez-Rodriguez, M. M., Oquendo, M. A., Keyes, M., Hasin, D. S., et al. (2011). "Estimating risk for suicide attempt: Are we asking the right questions?" Passive suicidal ideation as a marker for suicidal behavior. *Journal of Affective Disorders*, 134, 327–332.
- Beck, J. S. (2011). *Cognitive behavior therapy* (2nd ed.). New York, NY: Guilford Press.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100.
- Bertelsen, A. (1999). Reflections on the clinical utility of the ICD-10 and DSM-IV classifications and their diagnostic criteria. *Australian & New Zealand Journal of Psychiatry*, 32, 166–173.
- Beutler, L. E., & Malik, M. L. (2002). *Rethinking the DSM: A psychological perspective*. Washington, DC: American Psychological Association.
- Black, D. W. (2013). DSM-5 is approved, but personality disorders criteria have not changed. *Annals of Clinical Psychiatry*, 25, 1.
- Bornstein, R. F. (2006). A psychoanalytic construct lost and reclaimed: The psychodynamics of personality psychology. *Psychoanalytic Psychology*, 23, 339–353.
- Brabender, V., & Whitehead, M. L. (2011). Using the Psychodynamic Diagnostic Manual in the training of the competent assessor. *Journal of Personality Assessment*, 93, 185–193.
- Bracha, H. S. (2006). Human brain evolution and the "neuroevolutionary time-depth principle": Implications for the reclassification of fear-circuitry-related traits in DSM-V and for studying resilience to warzone-related posttraumatic stress disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 30, 827–853.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cassel, E., & Bernstein, D. A. (2007). *Criminal behavior* (2nd ed.). Mahwah, NJ: Erlbaum.
- Chapman, L. J., & Chapman, J. P. (1967). The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193–204.
- Clay, R. A. (2012). Protesting proposed changes to the DSM. *Monitor on Psychology*, Jan. 42–43.
- Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., et al. (2011). Most items in structured risk assessment instruments do not predict violence. *The Journal of Forensic Psychiatry and Psychology*, 22, 3–21.
- Colmen, J. G., Kaplan, S. J., & Boulger, J. R. (1964, August). *Selection and selecting research in the Peace Corps*. (Peace Corps Research Note No. 7).
- Comas-Diaz, L. (2012). *Multicultural care: A clinician's guide to cultural competence*. Washington, DC: American Psychological Association.
- Cronbach, L. J., & Glesser, G. C. (1964). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychology tests. *Psychological Bulletin*, 52, 281–302.
- Cunningham, M. D., Sorensen, J. R., & Reidy, T. J. (2009). Capital jury decision making: The limitations of predictions of future violence. *Psychology, Public Policy, and Law*, 15, 223–256.
- Dana, R. H., & Leech, S. (1974). Existential assessment. *Journal of Personality Assessment*, 38, 428–435.
- Dawes, R. M. (1994). *House of cards*. New York, NY: The Free Press.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668–1674.
- Duckworth, A. L., Steen, T. A., & Seligman, M. E. P. (2005). Positive psychology in clinical practice. *Annual Review of Clinical Psychology*, 1, 629–651.
- Ehde, D. M. (2010). Application of positive psychology to rehabilitation psychology. *Handbook of rehabilitation psychology* (2nd ed., pp. 417–424). Washington, DC: American Psychological Association.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85, 395–416.
- Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science*, 242, 31–35.
- Fazel, S., Singh, J. P., Doll, H., Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people: Systematic review and meta-analysis. *BMJ*, 345, e4692. doi: 10.1136/bmj.e4692.
- Finn, S. E. (1996). *Manual for using the MMPI-2 for a therapeutic intervention*. Minneapolis, MN: University of Minnesota Press.
- Finn, S. E., & Hale, M. (2013). Therapeutic assessment: Using psychological testing as brief therapy. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, et al. (Eds.), *APA handbook of testing and assessment in psychology*, (Vol. 2, pp. 453–465). Washington, DC: American Psychological Association.
- First, M. B. (2006). Relational processes in the DSM-V revision process: Comment on the special section. *Journal of Family Psychology*, 20, 356–358.
- Fischer, C. T. (1989). A life-centered approach to psychodiagnostics: Attending to lifeworld, ambiguity, and possibility. *Person-Centered Review*, 4, 163–170.

- Fischer, C. T. (2001). Psychological assessment: From objectification back to the life world. In B. D. Slife, R. N. Williams, & S. H. Barlow (Eds.), *Critical issues in psychotherapy* (pp. 29–44). Thousand Oaks, CA: Sage.
- Follette, W. C. (1996). Introduction to the special section on the development of theoretically coherent alternatives to the DSM system. *Journal of Consulting and Clinical Psychology*, 64, 1117–1119.
- Frances, A. J., & Widiger, T. (2011). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, 8, 109–130.
- Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., et al. (2013). The initial field trials of DSM-5: New blooms and old thorns. *American Journal of Psychiatry*, 170, 1–5.
- Frish, M. B. (1998). Quality of life therapy and assessment in health care. *Clinical Psychology: Science and Practice*, 5, 19–40.
- Gabbard, G. O. (2010). *Long-term psychodynamic psychotherapy: A basic text*. Washington, DC: American Psychiatric Publishers.
- Gabbard, G. O., Litowitz, B. E., & Williams, P. (Eds.). (2012). *Textbook of psychoanalysis*. Arlington, VA: American Psychiatric Publishing.
- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4, 641–656.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, 105, 387–396.
- Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychology: Theory and Practice*, 27, 272–277.
- Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt test. *Journal of Consulting Psychology*, 23, 25–33.
- Goldfried, M. R., & Davison, G. C. (1976). *Clinical behavior therapy* (pp. 52–63). New York, NY: Holt, Rinehart, and Winston.
- Goldfried, M. R., & Sprafkin, J. N. (1974). *Behavioral personality assessment*. Morristown, NJ: General Learning Press.
- Golding, S. L., & Rorer, L. G. (1972). Illusory correlation and subjective judgment. *Journal of Abnormal Psychology*, 80, 249–260.
- Gorske, T. T. (2008). Therapeutic neuropsychological assessment: A humanistic model and case example. *Journal of Humanistic Psychology*, 48, 320–339.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). New York, NY: Wiley.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.
- Hall, G. C. N. (2005). Introduction to the special section on multicultural and community psychology: Clinical psychology in context. *Journal of Consulting and Clinical Psychology*, 73, 787–789.
- Hammond, K. R., & Allen, J. M. (1953). *Writing clinical reports*. Englewood Cliffs, NJ: Prentice-Hall.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21, 1–21.
- Harding, T. P. (2007). Clinical decision-making: How prepared are we? *Training and Education in Professional Psychology*, 1, 95–104.
- Hart, S. D., & Cooke, D. J. (2013). Another look at the (im-)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Science and the Law*, 31, 81–102. doi: 10.1002/bsl.2049.
- Harwood, T. M., Beutler, L. E., & Groth-Marnat, G. (Eds.). (2011). *Integrative assessment of adult personality*. New York, NY: Guilford Press.
- Hays, D. G. (2013). *Assessment in counseling: A guide to the use of psychological assessment procedures* (5th ed.). Alexandria, VA: American Counseling Association.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on clinical versus actuarial prediction of violence. *Counseling Psychologist*, 34, 400–409.
- Holt, R. R., & Luborsky, L. (1958). *Personality patterns of psychiatrists: A study of methods for selecting residents* (Vol. 1). New York, NY: Basic Books.
- Houts, A. C. (2004). Discovery, invention, and the expansion of the modern Diagnostic and Statistical Manuals of Mental Disorders. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 17–68). Washington, DC: American Psychological Association.
- Hundt, N. E., Armento, M. E. A., Porter, B., Cully, J. A., Kunik, M. E., et al. (2013). Predictors of treatment satisfaction among older adults with anxiety in a primary care psychology program. *Evaluation and Program Planning*, 37, 58–63.
- Hunsley, J., & Mash, E. J. (2010). The role of assessment in evidence-based practice. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 3–22). New York, NY: Guilford.
- Huprich, S. K., & Meyer, G. J. (2011). Introduction to the JPA special issue: Can the Psychodynamic Diagnostic Manual put the complex person back at the center-stage of personality assessment? *Journal of Personality Assessment*, 93, 109–111.
- Institute of Personality Assessment and Research. (1970). *Annual report: 1969–1970*. Berkeley, CA: University of California.
- Joseph, S., & Wood, A. (2010). Assessment of positive functioning in clinical psychology: Theoretical and practical issues. *Clinical Psychology Review*, 30, 830–838.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed., p. 322). Boston, MA: Allyn & Bacon.
- Kazdin, A. E. (2006). Assessment and evaluation in clinical practice. In C. D. Goodhart, A. E. Kazdin, & R. J. Sternberg (Eds.), *Evidence-based psychotherapy* (pp. 153–177). Washington, DC: American Psychological Association.
- Kelly, E. L., & Fiske, D. W. (1951). *The prediction of performance in clinical psychology*. Ann Arbor, MI: University of Michigan Press.
- Kleinmuntz, B. (1984). The scientific study of clinical judgment in psychology and medicine. *Clinical Psychology Review*, 4, 111–126.
- Klopfer, W. G. (1983). Writing psychological reports. In C. E. Walker (Ed.), *The handbook of clinical psychology* (Vol. 1, pp. 501–527). Homewood, IL: Dow Jones-Irwin.
- Knox, O. (2013). 547 dead, 475 hurt in mass shootings since 1983. Retrieved March 20 from Yahoo! News at <http://news.yahoo.com/blogs/ticket/547-dead-476-hurt-mass-shootings-since-1983-182605303--politics.html>.
- Krishnamurthy, R., VandeCreek, L., Kaslow, N. J., Tazeau, Y. N., Miville, N. L., et al. (2004). Achieving competency in psychological assessment: Directions for education and training. *Journal of Clinical Psychology*, 60, 725–739.
- Krol, N., DeBruyn, E., & van den Bercken, J. (1995). Intuitive and empirical prototypes in childhood psychopathology. *Psychological Assessment*, 7, 533–537.
- Lambert, D. (1985). Political and economic determinants of mental health regulations. Unpublished doctoral dissertation, Brandeis University.
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkings, L., et al. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior*, 34, 37–59.
- Lewis, G. (1991). Observer bias in the assessment of anxiety and depression. *Social Psychiatry and Psychiatric Epidemiology*, 26, 265–272.
- Lopez, S., & Snyder, C. R. (Eds.). (2003). *Handbook of positive psychology assessment*. Washington, DC: American Psychological Association.

- Lopez, S. R., & Guarnaccia, P. J. (2000). Cultural psychopathology: Uncovering the social world of mental illness. *Annual Review of Psychology*, 51, 571-598.
- Malgady, R. G. (1996). The question of cultural bias in assessment and diagnosis of ethnic minority clients: Let's reject the null hypothesis. *Professional Psychology: Research and Practice*, 27, 73-77.
- McCoy, S. A. (1976). Clinical judgments of normal childhood behavior. *Journal of Consulting and Clinical Psychology*, 44, 710-714.
- McReynolds, P. (1975). Historical antecedents of personality assessment. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 3, pp. 477-532). San Francisco, CA: Jossey-Bass.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.
- Meehl, P. E. (1956). Wanted—A good cookbook. *American Psychologist*, 11, 263-272.
- Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Consulting Psychology*, 4, 268-273.
- Meehl, P. E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality*, 1, 27-32.
- Miller, G. (2012). Criticism continues to dog psychiatric manual as deadline approaches. *Science*, 336, 1088-1089.
- Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.
- Mitchell, M. J., Patterson, C. A., & Boyd-Franklin, N. (2011). Commentary: Increasing cultural diversity in pediatric psychology family assessment research. *Journal of Pediatric Psychology*, 36, 634-641.
- Mulvey, E. P., & Cauffman, E. (2001). The inherent limits of predicting school violence. *American Psychologist*, 56, 797-802.
- Nathan, P. E., & Langenbucher, J. W. (1999). Psychopathology: Description and classification. *Annual Review of Psychology*, 50, 79-107.
- Nietzel, M. T., & Bernstein, D. A. (1976). The effects of instructionally mediated demand upon the behavioral assessment of assertiveness. *Journal of Consulting and Clinical Psychology*, 44, 500.
- Norcross, J. C., & Lambert, M. J. (2011). Psychotherapy relationships that work II. *Psychotherapy*, 48, 4-8.
- Office of Strategic Services Assessment Staff. (1948). *Assessment of men*. New York, NY: Rinehart.
- Olive, H. (1972). Psychoanalysts' opinions of psychologists' reports: 1952 and 1970. *Journal of Clinical Psychology*, 28, 50-54.
- Ougrin, D., Zundel, T., Kyriakopoulos, M., Banarsee, R., Stahl, D., & Taylor, E. (2011). Adolescents with suicidal and nonsuicidal self-harm: Clinical characteristics and response to therapeutic assessment. *Psychological Assessment*. Advance online publication. doi: 10.1037/a0025043.
- Packard, E. (2007). A new tool for psychotherapists. *Monitor on Psychology*, 38, 30-31.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31, 109-118.
- Peterson, C. (2006). The Values in Action (VIA) Classification of Strengths: The un-DSM and the real DSM. In M. Csikszentmihalyi & I. Csikszentmihalyi (Eds.), *A life worth living: Contributions to positive psychology* (pp. 29-48). New York, NY: Oxford University Press.
- Peterson, D. R. (1968). *The clinical style of social behavior*. New York, NY: Appleton-Century-Crofts.
- Peterson, J., Skeem, J., & Manchak, S. (2011). If you want to know, consider asking: How likely is it that patients will hurt themselves in the future? *Psychological Assessment*. doi: 10.1037/a0022971
- Pompili, M., & Taraelli, R. (Eds.). (2011). *Evidence-based suicidology: A source book*. Cambridge, MA: Hogrefe Publishing.
- Ridley, C. R., & Shaw-Ridley, M. (2009). Clinical judgment accuracy: From meta-analysis to metatheory. *The Counseling Psychologist*, 37, 400-409.
- Rock, D. L., Bransford, J. D., Maisto, S. A., & Morey, L. (1987). A study of clinical judgment: An ecological approach. *Clinical Psychology Review*, 7, 645-661.
- Rogers, C. R. (1951). *Client-centered therapy*. Boston, MA: Houghton Mifflin.
- Rosenberg, A., Almeida, A., & Macdonald, H. (2012). Crossing the cultural divide: Issues in translation, mistrust, and co-creation of meaning in cross-cultural therapeutic assessment. *Journal of Personality Assessment*. Advance online publication. doi: 10.1080/00223891.2011.648293.
- Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49, 929-954.
- Sartorius, N., Kaelber, C. T., Cooper, J. E., Roper, M. T., Rae, D. S., et al. (1996). Progress toward achieving a common language in psychiatry: Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders ICD-10. *Archives of General Psychiatry*, 50, 115-124.
- Secrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153-158.
- Shannon, D., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Singer, J. A. (2013). Lost in translation? Finding the person in the emerging paradigm of clinical science: Introduction to a special issue on personality psychology and psychotherapy. *Journal of Personality*, online, doi: 10.1111/jopy.12017.
- Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., et al. (2009). The meta-analysis of clinical judgment accuracy: Effects of experience on judgment accuracy. *The Counseling Psychologist*, 37, 350-399.
- Strohmer, D. C., & Shivy, V. A. (1994). Bias in counselor hypothesis testing: Testing the robustness of counselor confirmatory bias. *Journal of Counseling and Development*, 73, 191-197.
- Sundberg, N. D. (1977). *Assessment of persons* (pp. 97-98). New York, NY: Prentice Hall.
- Sundberg, N. D., Tyler, L. E., & Taplin, J. R. (1973). *Clinical psychology: Expanding horizons* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Tallent, N. (1992). *The practice of psychological assessment*. Englewood Cliffs, NJ: Prentice-Hall.
- Tavris, C., & Aronson, E. (2007). *Mistakes were made (but not by me)*. New York, NY: Harcourt.
- Thorne, F. C. (1972). Clinical judgment. In R. H. Woody & J. D. Woody (Eds.), *Clinical assessment in counseling and psychotherapy* (pp. 30-85). Englewood Cliffs, NJ: Prentice-Hall.
- Tutin, J. (1993). The persistence of initial beliefs in clinical judgment. *Journal of Social and Clinical Psychology*, 12, 319-335.
- Vincent, G. M. (2006). Psychopathy and violence risk assessment in youth. *Child and Adolescent Psychiatric Clinics of North America*, 15, 407-428.
- Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and Practice*, 40, 525-531.
- Westen, D. (2012). Prototype diagnosis of psychiatric syndromes. *World Psychiatry*, 11, 16-21.
- Widiger, T. A., Frances, A. J., Pincus, H. A., Davis, W. W., & First, M. B. (1991). Toward an empirical classification for the DSM-IV. *Journal of Abnormal Psychology*, 100, 280-288.
- Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, 62, 71-83.
- Wright, A. J. (2011). *Conducting psychological assessment*. New York, NY: John Wiley & Sons.

- Wright, J. W. (2010). *Conducting psychological assessment*. Hoboken, NJ: John Wiley & Sons.
- Wulfert, E., Greenway, D. E., & Dougher, M. J. (1996). A logical functional analysis of reinforcement-based disorders: Alcoholism and pedophilia. *Journal of Consulting and Clinical Psychology, 64*, 1140–1151.
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740–767.
- Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child and Adolescent Psychology, 42*, 139–159.
- Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., & Sanislow, C., et al. (2000). The collaborative longitudinal personality disorders study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders, 14*, 291–299.