

# 3

# Diagnosis and Assessment

## LEARNING GOALS

1. Be able to describe the purposes of diagnosis and assessment.
2. Be able to distinguish the different types of reliability and validity.
3. Be able to identify the basic features, historical changes, strengths, and weaknesses of the DSM.
4. Be able to describe the goals, strengths, and weaknesses of psychological and neurobiological approaches to assessment.
5. Be able to discuss the ways in which culture and ethnicity impact diagnosis and assessment.

## Clinical Case: Aaron

Hearing the sirens in the distance, Aaron realized that someone must have called the police. He didn't mean to get upset with the people sitting next to him at the bar, but he just knew that they were talking about him and plotting to have his special status with the CIA revoked. He could not let this happen again. The last time people conspired against him, he wound up in the hospital. He did not want to go to the hospital again and endure all of the evaluations. Different doctors would ask him all sorts of questions about his work with the CIA, which he simply was not at liberty to discuss. They asked other odd questions, such as whether he heard voices or believed others were putting thoughts into his head. He was never sure how they knew that he had those experiences, but he suspected that there were electronic bugging devices in his room at his parents' house, perhaps in the electrical outlets.

Just yesterday, Aaron began to suspect that someone was watching and listening to him through the electrical outlets. He decided that the safest thing to do was to stop speaking to his parents. Besides, they were constantly hounding him to take his medication. But when he took this medication, his vision got blurry and he had trouble sitting still. He reasoned that his parents must somehow be part of the group of people trying to remove him from the CIA. If he took this medication, he would lose his special powers that allowed him to spot terrorists in any setting, and the CIA would stop leaving messages for him in phone booths or in the commercials on Channel 2. Just the other day, he found a tattered paperback book in a phone booth, which he interpreted to mean that a new assignment was imminent. The voices in his head were giving him new clues about terrorist activity. They were currently telling him that he should be wary of people wearing the color purple, as this was a sign of a terrorist. If his parents were trying to sabotage his career with the

CIA, he needed to keep out of the house at all costs. That was what had led him to the bar in the first place. If only the people next to him wouldn't have laughed and looked toward the door. He knew this meant that they were about to expose him as a CIA operative. If he hadn't yelled at them to stop, his cover would have been blown.

**D**IAGNOSIS AND ASSESSMENT are the critically important “first steps” in the study and treatment of psychopathology. In the case of Aaron, a clinician may begin treatment by determining whether Aaron meets the diagnostic criteria for a mood disorder, schizophrenia, or perhaps a substance-related disorder. Diagnosis can be the first major step in good clinical care. Having a correct **diagnosis** will allow the clinician to describe base rates, causes, and treatment for Aaron and his family, all of which are important aspects of good clinical care. More broadly, imagine that your doctor told you, “There is no diagnosis for what you have.” Rather than this alarming scenario, hearing a diagnosis can provide relief in several different ways. Often, a diagnosis can help a person begin to understand why certain symptoms are occurring, which can be a huge relief. Many disorders are extremely common, such as depression, anxiety, and substance abuse—knowing that his or her diagnosis is common can also help a person feel less unusual.

Diagnosis enables clinicians and scientists to communicate accurately with one another about cases or research. Without agreed-on definitions and categories, our field would face a situation like the Tower of Babel (Hyman, 2002), in which different scientists and clinicians would be unable to understand each other.

Diagnosis is important for research on causes and treatments. Sometimes researchers discover unique causes and treatments associated with a certain set of symptoms. For example, autism was only recognized in the *Diagnostic and Statistical Manual* in 1980. Since that time, research on the causes and treatments of autism has grown exponentially.

To help make the correct diagnosis, clinicians and researchers use a variety of assessment procedures, beginning with a clinical interview. Broadly speaking, all clinical assessment procedures are more or less formal ways of finding out what is wrong with a person, what may have caused problems, and what can be done to improve the person's condition. Assessment procedures can help in making a diagnosis, and they can also provide information beyond a diagnosis. Indeed, a diagnosis is only a starting point. In the case of Aaron, for example, many other questions remain to be answered. Why does Aaron behave as he does? Why does he believe he is working for the CIA? What can be done to resolve his conflicts with his parents? Has he performed up to his intellectual potential in school and in his career? What obstacles might interfere with treatment? These are also the types of questions that mental health professionals address in their assessments.

In this chapter, we will describe the official diagnostic system used by many mental health professionals, as well as the strengths and weaknesses of this system. We will then turn to a discussion of the most widely used assessment techniques, including interviews, psychological assessment, and neurobiological assessment. We then conclude the chapter with an examination of a sometimes neglected aspect of assessment, the role of cultural bias. Before considering diagnosis and assessment in detail, however, we begin with a discussion of two concepts that play a key role in diagnosis and assessment: reliability and validity.

## Cornerstones of Diagnosis and Assessment

The concepts of reliability and validity are the cornerstones of any diagnostic or assessment procedure. Without them, the usefulness of our methods is seriously limited. That said, these two concepts are quite complex. There are several kinds of each, and an entire subfield of psychology—psychometrics—exists primarily for their study. Here, we provide a general overview.





Reliability is an essential property of all assessment procedures. One means of establishing reliability is to determine whether different judges agree, as happens when two umpires witness the same event in a baseball game. (Reuters/NewMedia Inc./Corbis Images.)

## Reliability

**Reliability** refers to consistency of measurement. An example of a reliable measure would be a wooden ruler, which produces the same value every time it is used to measure an object. In contrast, an unreliable measure would be a flexible, elastic-like ruler whose length changes every time it is used. Several types of reliability exist, and here we will discuss the types that are most central to assessment and diagnosis.

**Interrater reliability** refers to the degree to which two independent observers agree on what they have observed. To take an example from baseball, two umpires may or may not agree as to whether the ball is fair or foul.

**Test–retest reliability** measures the extent to which people being observed twice or taking the same test twice, perhaps several weeks or months apart, receive similar scores. This kind of reliability makes sense only when we can assume that the people will not change appreciably between test sessions on the underlying variable being measured; a prime example of a situation in which this type of reliability is typically high is in evaluating intelligence tests. On the other hand, we cannot expect people to be in the same mood at a baseline and a follow-up assessment 4 weeks later.

Sometimes psychologists use two forms of a test rather than giving the same test twice, perhaps when there is concern that test takers will remember their answers from the first round of taking the test and aim merely to be consistent. This approach enables the tester to determine **alternate-form reliability**, the extent to which scores on the two forms of the test are consistent.

Finally, **internal consistency reliability** assesses whether the items on a test are related to one another. For example, one would expect the items on an anxiety questionnaire to be interrelated, or to correlate with one another, if they truly tap anxiety. A person who reports a dry mouth in a threatening situation would be expected to report increases in muscle tension as well, since both are common characteristics of anxiety.

## Validity

**Validity** is a complex concept, generally related to whether a measure measures what it is supposed to measure. For example, if a questionnaire is supposed to measure a person's hostility, does it do so? Before we describe types of validity, it is important to note that validity is related to reliability—unreliable measures will not have good validity. Because an unreliable measure does not yield consistent results (recall our example of a ruler whose length is constantly changing), it will not relate very strongly to other measures. For example, an unreliable measure of coping is not likely to relate well to how a person adjusts to stressful life experiences. Reliability, however, does not guarantee validity. Height can be measured very reliably, but height would not be a valid measure of anxiety.

**Content validity** refers to whether a measure adequately samples the domain of interest. For example, later in this chapter we will describe an interview that is often used to make an Axis I diagnosis. It has excellent content validity because it contains questions about all the symptoms that are involved in most Axis I diagnoses. For certain uses, though, the interview might have poor content validity. The interview doesn't cover questions about kleptomania (a disorder characterized by a compulsive need to steal). If one were trying to assess kleptomania, this interview would have poor content validity.

**Criterion validity** is evaluated by determining whether a measure is associated in an expected way with some other measure (the criterion). If both variables are measured at the same point in time, the resulting validity is referred to as **concurrent validity**. For example, below we will describe a measure of the overly negative thoughts that are believed to play an important role in depression. Criterion validity for this measure of negative thoughts could be established by showing that people with depression score higher on the test than do people without depression. Alternatively, criterion validity can be assessed by evaluating the ability of

the measure to predict some other variable that is measured at some point in the future, often referred to as **predictive validity**. For example, IQ tests were originally developed to predict future school performance. Similarly, a measure of negative thinking could be used to predict the development of depression in the future. In summary, concurrent and predictive validity are both types of criterion validity.

**Construct validity** is a more complex concept. It is relevant when we want to interpret a test as a measure of some characteristic or construct that is not observed simply or overtly (Cronbach, 1955; Hyman, 2002). A construct is an inferred attribute, such as anxiousness or distorted cognition. Consider an anxiety-proneness questionnaire as an example. If the questionnaire has construct validity, people who obtain different scores on our test really will differ in anxiety proneness. Just because the items seem to be about the tendency to become anxious (“I find that I become anxious in many situations”), it is not certain that the test is a valid measure of the construct of anxiety proneness.

Construct validity is evaluated by looking at a wide variety of data from multiple sources (compare this to criterion validity, where a test is typically evaluated against just one other piece of data). For example, people diagnosed as having an anxiety disorder and people without such a diagnosis could be compared on their scores on our self-report measure of anxiety proneness. The self-report measure would achieve some construct validity if the people with anxiety disorders scored higher than the people without anxiety disorders. Greater construct validity would be achieved by showing that the self-report measure was related to other measures thought to reflect anxiety, such as observations of fidgeting and trembling, and physiological indicators, such as increased heart rate and rapid breathing. When the self-report measure is associated with these multiple measures (diagnosis, observational indicators, physiological measures), its construct validity is increased.

More broadly, construct validity is related to theory. For example, we might hypothesize that being prone to anxiety is in part caused by a family history of anxiety. We could then obtain further evidence for the construct validity of our questionnaire by showing that it relates to a family history of anxiety. At the same time, we would also have gathered support for our theory of anxiety proneness. Thus, construct validation is an important part of the process of theory testing.

Construct validity is also centrally important to diagnostic categories. Below, we consider in more detail the issue of construct validity and the DSM-5.

## Classification and Diagnosis

### Clinical Case: Roxanne

Roxanne is a middle-aged woman who was brought to the local psychiatric emergency room by the police. They had found her running through a crowded street, laughing loudly and running into people. Her clothes were dirty and torn. When they questioned her, she was speaking very rapidly, and she was hard to follow. At the ER, she wrestled free of the police and began running down the hallway. She knocked over two staff members during her flight, while bellowing at the top of her lungs, “I am the resurrection! Come follow me!” Police brought her back to the exam room, and the staff began to form hypotheses. Clearly, she was full of energy. Had she been through some trauma? She believed she had special religious powers—could this be a delusion? Unfortunately, the staff were unable to gain much

information from an interview due to her rapid and incoherent speech. Rather, Roxanne sat restlessly, occasionally laughing and shouting; treatment could not proceed without understanding the reason for her unusual behavior. When efforts to calm Roxanne failed, police helped the staff to contact family members, who were relieved to hear that Roxanne was safe. She had disappeared from home the day before. Family members described a long history of bipolar disorder (formerly known as manic depression), and they reported having been concerned for the past couple weeks because Roxanne had stopped taking medications for her bipolar disorder and for her high blood pressure. Treatment was able to proceed based on the idea that Roxanne was experiencing a new manic episode of her long-standing bipolar disorder.

## The Diagnostic System of the American Psychiatric Association: Toward DSM-5

In this section, we focus on the official diagnostic system used by mental health professionals, the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). The DSM is now in its fourth edition, commonly referred to as DSM-IV-TR. A draft of the DSM-5 is now available ([www.dsm5.org](http://www.dsm5.org)). This draft edition is being reviewed and tested, and the final release of DSM-5 is expected in 2013. Because the DSM-5 will be in use by the time most of you are working in clinical and research settings, we will focus on the likely DSM-5 in this book. We recognize, though, that this is an interesting time in the field—the DSM-IV-TR remains in use, even as the DSM-5 is anticipated soon. Throughout the chapters of this book, then, we will note major differences between the DSM-IV-TR and the proposed DSM-5. In this chapter, we will review the history of the DSM and the major features of the latest versions of the DSM, and then we will review some strengths and criticisms of the DSM as well as of diagnosis in general.

In 1952, the American Psychiatric Association published its *Diagnostic and Statistical Manual* (DSM). The publication of the DSM was informed by earlier systems of classification (for a review, see Focus on Discovery 3.1), and it has been revised five times since 1952. DSM-IV was published in 1994, and in June 2000, a “text revision,” DSM-IV-TR, followed. Almost no changes were made to the diagnostic categories and criteria in the 2000 revision. Rather, DSM-IV-TR provided a summary of new research findings on prevalence rates, course, and etiology (causes).

Each version of the DSM has included improvements. Beginning with the third edition of DSM and continuing today, an effort was made to create more reliable and valid diagnostic categories. Two major innovations were introduced in DSM-III that have been retained by each edition since.

1. Specific diagnostic criteria—the symptoms for a given diagnosis—are spelled out precisely, and clinical symptoms are defined in a glossary. Table 3.1 compares the descriptions of a manic episode given in DSM-II with the diagnostic criteria given in the likely DSM-IV-TR. Notice how DSM-IV-TR is much more detailed and concrete.
2. The characteristics of each diagnosis are described much more extensively than they were in DSM-II. For each disorder there is a description of essential features, then of associated

### FOCUS ON DISCOVERY 3.1

#### A History of Classification and Diagnosis

By the end of the nineteenth century, medical diagnostic procedures were improving as physicians began to understand the advantages of tailoring treatments to different illnesses. During the same period, other sciences, such as botany and chemistry, advanced after classification systems were developed. Impressed by these successes, investigators of mental disorders sought to develop classification schemes. Unfortunately, progress in classifying mental disorders did not come easily.

##### Early Efforts at Classification of Mental Illness

Emil Kraepelin (1856–1926) authored an influential early classification system in his textbook of psychiatry first published in 1883. His classification system attempted to definitively establish the biological nature of mental illnesses. Kraepelin noted that certain symptoms clustered together as a *syndrome*. He labeled a set of syndromes and hypothesized that each had its own biological cause, course, and outcome. Even though effective treatments had not been identified, at least the course of the disease could be predicted.

Kraepelin proposed two major groups of severe mental illnesses: dementia praecox (an early term for schizophrenia) and manic-depressive psychosis (an early term for bipolar disorder). He postulated a chemical imbalance as the cause of dementia praecox and an irregularity in metabolism as the explanation of manic-depressive psychosis. Though his theories about causes were not quite correct, Kraepelin’s classification scheme nonetheless influenced the current diagnostic categories.

##### Development of the WHO and DSM Systems

In 1939 the World Health Organization (WHO) added mental disorders to the International List of Causes of Death (ICD). In 1948 the list was expanded to become the International Statistical Classification of Diseases, Injuries, and Causes of Death, a comprehensive listing of all diseases, including a classification of abnormal behavior. Unfortunately, the mental disorders section was not widely accepted. Even though American psychiatrists had played a prominent role in the WHO effort,

features, such as laboratory findings (e.g., enlarged ventricles in schizophrenia) and results from physical exams (e.g., electrolyte imbalances in people who have eating disorders). Next, a summary of the research literature provides information about age of onset, course, prevalence and sex ratio, familial pattern, and differential diagnosis (i.e., how to distinguish similar diagnoses from each other).

The DSM-IV and DSM-IV-TR introduced more focus on cultural issues as well as separate dimensions, or axes, to rate people. As shown in Figure 3.2, DSM-IV-TR includes five axes. This **multiaxial classification system**, by requiring judgments on each of the five axes, forces the diagnostician to consider a broad range of information. The DSM-IV-TR Axis I includes all

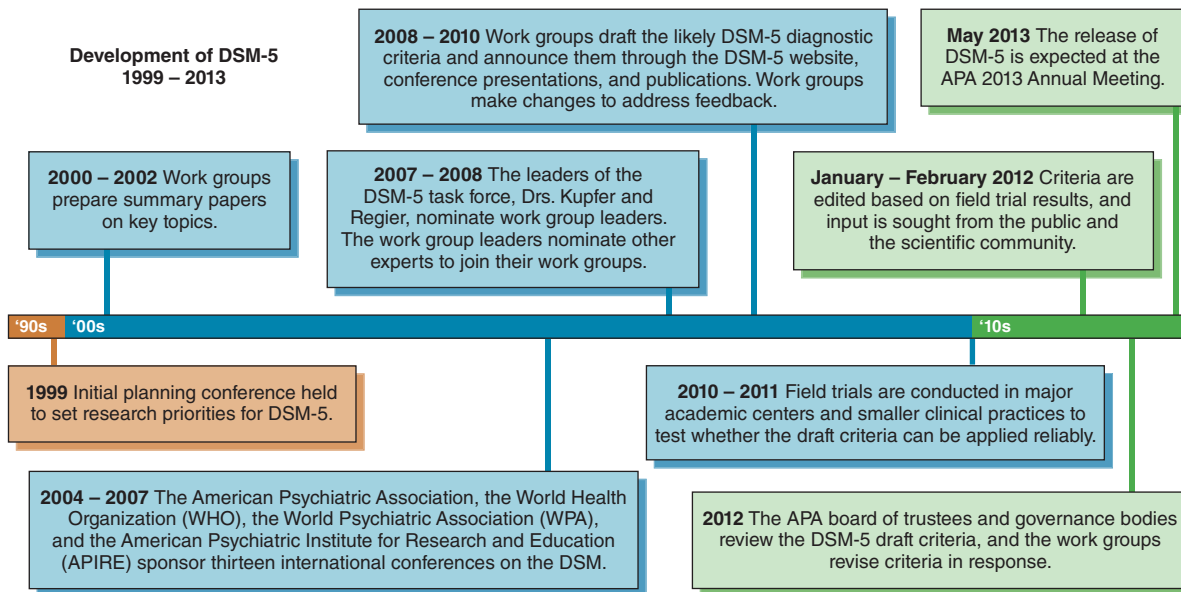


Figure 3.1 Timeline for the development of DSM-5.

the American Psychiatric Association published its own *Diagnostic and Statistical Manual* (DSM) in 1952.

In 1969 the WHO published a new classification system, which was more widely accepted. In the United Kingdom, a glossary of definitions was produced to accompany the WHO system (General Register Office, 1968). A second version of the American Psychiatric Association's DSM, DSM-II (1968), was similar to the WHO system. But true consensus still eluded the field. Even though DSM-II and the British Glossary of Mental Disorders specified some symptoms of diagnoses, the two systems defined different symptoms for a given disorder! Thus diagnostic practices still varied widely.

In 1980 the American Psychiatric Association published an extensively revised diagnostic manual, DSM-III, and a somewhat revised version, DSM-III-R, followed in 1987. In 1988 the American Psychiatric Association began work on DSM-IV, which was published in 1994. Thirteen working groups, which included many psychologists, were established to critique DSM-III-R, review literature, analyze previously collected data, and collect new data. Each work group tackled a different cluster of disorders. The committee adopted an important new approach—the reasons for changes in diagnoses would be explicitly

stated and supported by data. In previous versions of the DSM, the reasons for diagnostic changes had not always been explicit.

As shown in Figure 3.1, plans for DSM-5 began in 1999. As with the process for DSM-IV, 13 work groups were formed to review each set of diagnosis. A series of study groups were also formed to consider issues that cut across diagnostic categories, such as lifespan developmental approaches, gender and cross-cultural issues, general medical issues, impairment and disability, and diagnostic assessment instruments. These study groups conducted literature reviews and analyses and then provided feedback to the work groups regarding issues with specific diagnoses.

Work groups were asked to follow several principles in considering revisions. As with DSM-IV-TR, all changes were to be based on research data. The leaders emphasized, though, that the highest priority was to make the DSM-5 useful for clinicians. To protect the process from commercial interests, all work group members signed conflict-of-interest agreements, stating that they would limit their income to \$10,000 or less per year from pharmaceutical and technology companies and similar industry groups. The crafters of the DSM are striving to create a living document that will change as new research evidence emerges. New editions, then, will be on the horizon even as DSM-5 emerges.



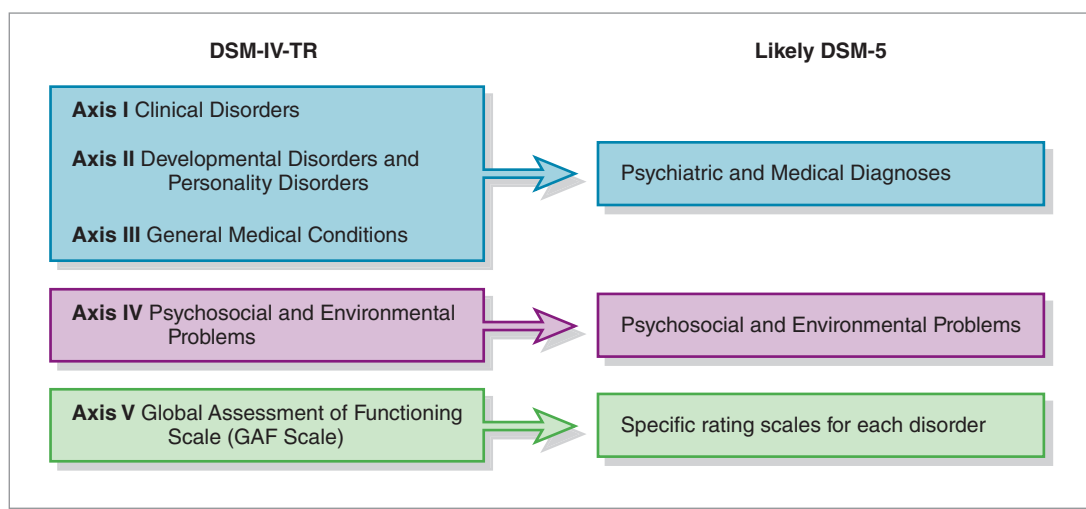
**Table 3.1 Description of Mania in DSM-II versus DSM-IV-TR****DSM-II (1968, p. 36)**

Manic-depressive illness, manic type. This disorder consists exclusively of manic episodes. These episodes are characterized by excessive elation, irritability, talkativeness, flight of ideas, and accelerated speech and motor activity. Brief periods of depression sometimes occur, but they are never true depressive episodes.

**DSM-IV-TR (2000, p. 362)****Diagnostic Criteria for a Manic Episode**

- A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary).
- B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:
1. inflated self-esteem or grandiosity
  2. decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
  3. more talkative than usual or pressure to keep talking
  4. flight of ideas or subjective experience that thoughts are racing
  5. distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
  6. increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
  7. excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)
- C. The symptoms do not meet criteria for a Mixed Episode.
- D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.
- E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

*Note:* DSM-IV-TR material reprinted with permission from the DSM-II, copyright 1968, and the DSM-IV-TR copyright 2000, American Psychiatric Association.

**Figure 3.2** Multiaxial classification system in DSM-IV-TR and proposed DSM-5.

diagnostic categories except the personality disorders and mental retardation, which make up Axis II. Thus Axes I and II cover the classification of mental disorders. Axis III covers general medical conditions. For many diagnoses, the DSM includes a provision for indicating that the

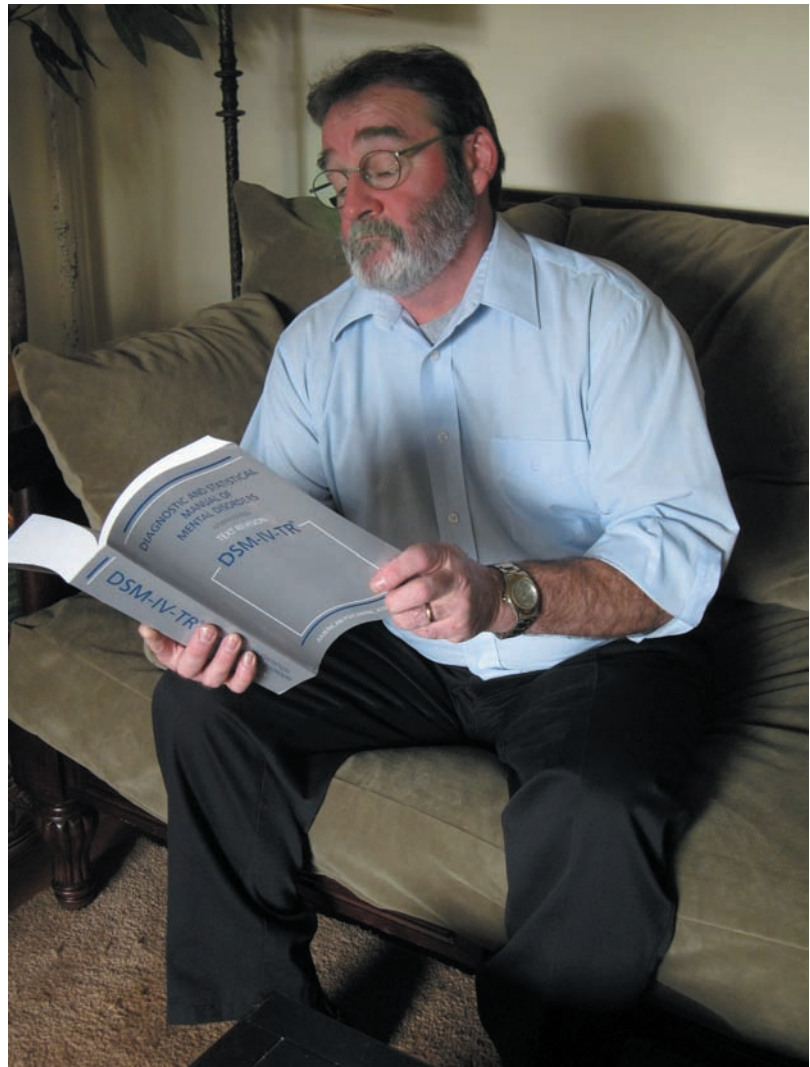
disorder is due to a medical condition or substance abuse. On Axis IV, the clinician codes psychosocial problems that may contribute to the disorder, including occupational problems, housing problems, economic problems, or interpersonal difficulties. Finally, on Axis V, the clinician indicates the person's current level of adaptive functioning, using ratings from 0 to 100 on the Global Assessment of Functioning (GAF) scale to consider social relationships, occupational functioning, and use of leisure time. As we discuss next, these axes are likely to change in DSM-5.

The DSM-5 will likely include many changes from DSM-IV-TR. Indeed, even conventions for labeling the edition have shifted—the Roman numerals used to denote the edition (e.g., DSM-IV) are replaced with Arabic numbers (i.e., DSM-5) to facilitate electronic printing. We will cover many of the changes as we discuss specific disorders in the chapters throughout this book. Here, then, we cover some of the major debates and changes that have implications across diagnoses.

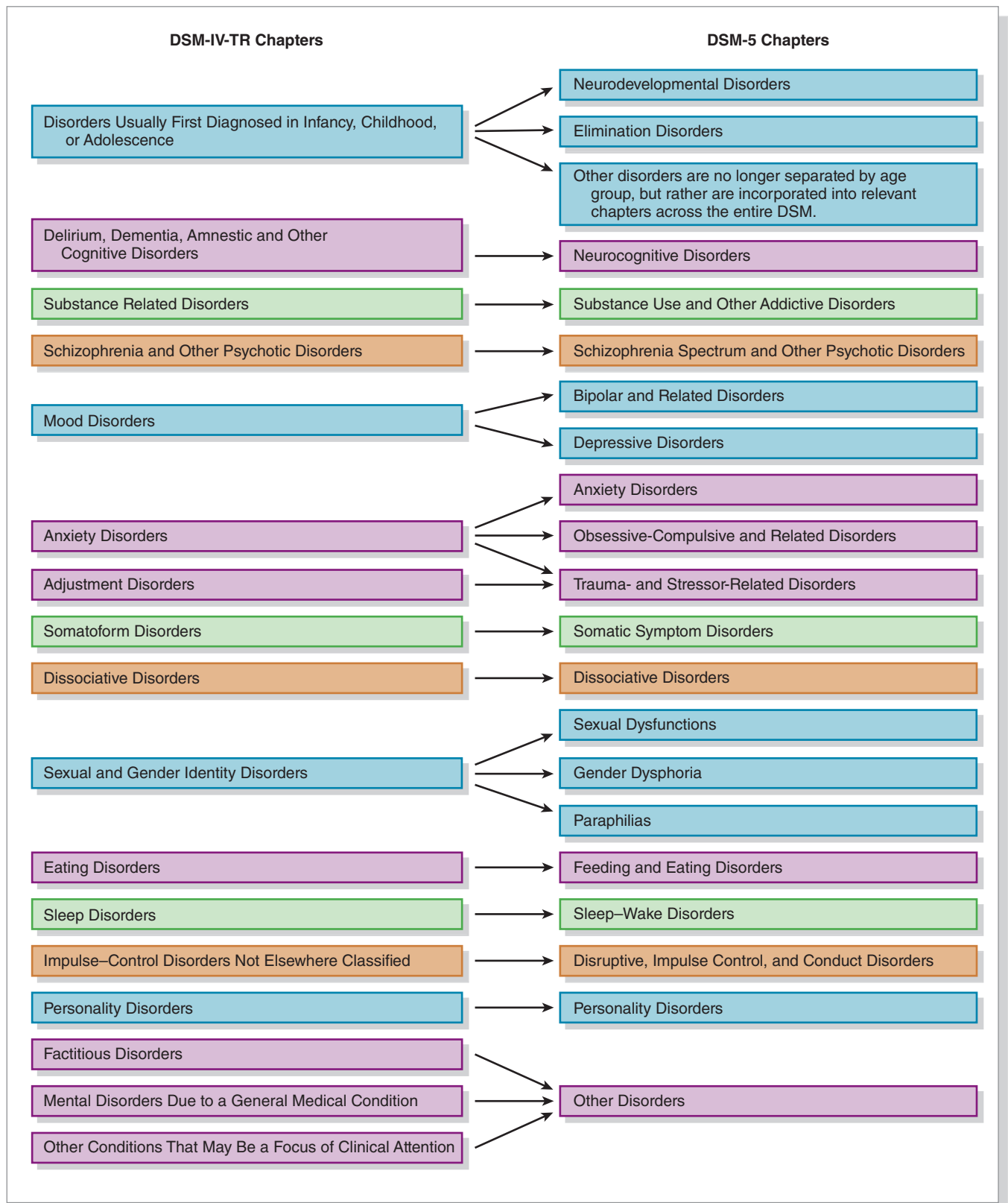
**Changes to the Multiaxial System** As shown in Figure 3.2, the multiaxial system developed for DSM-IV-TR is changed substantially in DSM-5. The five axes of DSM-IV-TR are reduced to one axis for clinical syndromes and one for psychosocial and environmental problems. The codes for the Psychosocial and Environmental Problems Axis are changed to be more similar to those used by the international community in the World Health Organization's (WHO) International Classification of Diseases (ICD). The DSM-IV-TR axis V is removed in DSM-5; instead, clinicians will be asked to rate severity along a continuum using scales developed specifically for each disorder.

**Organizing Diagnoses by Causes** DSM-IV-TR defines diagnoses entirely on the basis of symptoms. Some have argued that advances in our understanding of etiology (causes) could help us rethink this approach. For example, schizophrenia and schizotypal personality disorder share a great deal of genetic overlap. Could these ties be reflected in the diagnostic system? Others have proposed organizing diagnoses based on parallels in neurotransmitter activity, temperament, emotion dysregulation, or social triggers. After considerable review, it became clear that our knowledge base is not yet strong enough to organize diagnoses around etiology (Hyman, 2010). With the exception of IQ tests for intellectual developmental disorder (formerly known as mental retardation) or polysomnography for sleep disorders, we have no laboratory tests, neurobiological markers, or genetic indicators to use in making diagnoses. The DSM-5 will continue to use symptoms as the basis for diagnosis.

On the other hand, some changes have been made to reflect growing knowledge of etiology. The DSM-IV-TR diagnoses are clustered into chapters based on similarity of symptoms. In the DSM-5, the chapters are reorganized to reflect patterns of comorbidity and shared etiology (see Figure 3.3). For example, in DSM-IV-TR, obsessive-compulsive disorder is included as an anxiety disorder. The etiology of this disorder, though, seems to involve distinct genetic and neural influences compared to other anxiety disorders, as we discuss in Chapter 7. To reflect this, the DSM-5 includes a new chapter for obsessive-compulsive and other related disorders. This new chapter includes three disorders that often co-occur and share some risk factors: obsessive-compulsive disorder, hoarding disorder, and body dysmorphic disorder.



DSM-IV-TR is the current diagnostic system of the American Psychiatric Association. DSM-5 is expected in 2013. (Teri Stratford/Six-Cats Research Inc.)



**Figure 3.3** Chapters in DSM-IV-TR and proposed DSM-5.

### Including a Continuous Severity Rating to Supplement Categorical Classification

In the DSM-IV-TR clinical diagnoses are based on **categorical classification**. Does the patient have schizophrenia or not? This type of classification does not consider continuity between normal and abnormal behavior. For example, in Table 3.1 we see that the diagnosis of mania requires the presence of three symptoms from a list of seven, or four if the person's mood is irritable. But why require three symptoms rather than two or five? A categorical system forces clinicians to define one threshold as "diagnosable." There is often little research support for the threshold defined. Categorical diagnoses foster a false impression of discontinuity (Widiger, 2005).

It may be more helpful to know the severity of symptoms as well as whether they are present. In contrast to categorical classification, **dimensional** systems describe the *degree* of an entity that is present (e.g., a 1-to-10 scale of anxiety, where 1 represents minimal and 10, extreme). (See Figure 3.4 for an illustration of the difference between dimensional and categorical approaches.)

One reason categorical systems are popular is that they define a threshold for treatment. Consider high blood pressure (hypertension). Blood pressure measurements form a continuum, which clearly fits a dimensional approach; yet by defining a threshold for high blood pressure, doctors can feel more certain about when to offer treatment. Similarly, a threshold for clinical depression may help demarcate a point where treatment is recommended. Although the cutoffs are likely to be somewhat arbitrary, they can provide helpful guidance.

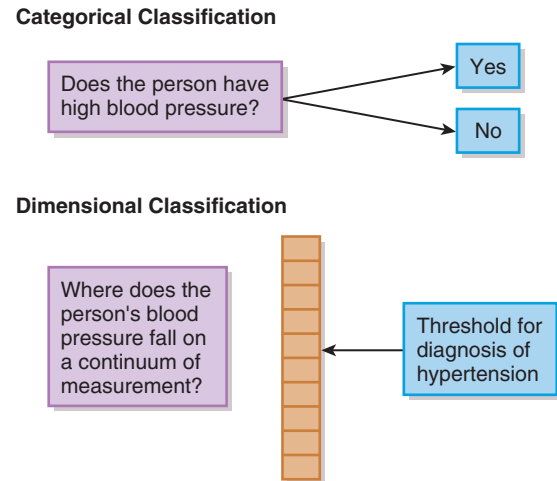
Despite some debate, DSM-5 preserves a categorical approach to diagnosis. The categories, though, are supplemented by a severity rating for each disorder. See Figure 3.5 for an example of one of the severity rating scales proposed for DSM-5. The severity rating provides a more precise estimate of how serious an illness is (Kraemer, 2007).

The severity ratings do not address all of the concerns raised about categorical diagnosis—severity ratings will not be considered unless a person is first diagnosed with a categorical diagnosis. Up to half of the people seeking treatment have mild symptoms that fall just below the threshold for a diagnosis (Helmuth, 2003). Many of these people with subthreshold symptoms of a diagnosis still receive extensive treatment (Johnson, 1992). As with DSM-IV-TR, the DSM-5 will likely include the category "not otherwise specified" to be used when a person meets many but not all of the criteria for a diagnosis. Just as with DSM-IV-TR, it is probable that far too many people will fit the "not otherwise specified" category.

**Changes in Personality Disorder Diagnoses** DSM-IV-TR includes 10 different personality disorder categories. The proposed DSM-5 will include criteria for deciding if a personality disorder is present and then will specify five types. Rating scales will be provided to assess how well a person's symptoms fit with the different types. We discuss this in more detail in Chapter 15.

**New Diagnoses** Several new diagnoses are proposed in the DSM-5. For example, disruptive mood dysregulation disorder is included to address the growing number of children and adolescents who are seen by clinicians due to severe mood changes and irritability as well as some of the symptoms of mania. Many of these youth do not meet the full criteria for mania (the defining feature of bipolar disorder) but were often falsely labeled with bipolar disorder because no other category seemed to fit their symptoms. It is hoped that by including this diagnosis, the overdiagnosis of bipolar disorder in children and adolescents will be lessened. Other new diagnoses include mixed anxiety depressive disorder, language impairment disorder, premenstrual dysphoric disorder, simple somatic symptom disorder, and illness anxiety disorder.

**Combining Diagnoses** Some of the DSM-IV-TR diagnoses have been combined because there is not enough evidence for differential etiology, course, or treatment response to justify labeling the conditions separately. For example, the DSM-IV-TR diagnoses of substance abuse and dependence are replaced with the DSM-5 diagnosis of substance use disorder. The DSM-IV-TR



**Figure 3.4** Categorical versus dimensional systems of diagnosis.



Name: John Q. Sample Date: \_\_\_\_\_

Over the *last 2 weeks*, how often have you been bothered by any of the following problems? (use “√” to indicate your answer)

	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2 ✓	3
2. Feeling down, depressed, or hopeless	0	1 ✓	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2 ✓	3
4. Feeling tired or having little energy	0	1	2	3 ✓
5. Poor appetite or overeating	0	1 ✓	2	3
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down	0	1	2 ✓	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2 ✓	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual	0	1	2 ✓	3
9. Thoughts that you would be better off dead, or of hurting yourself in some way	1 ✓	1	2	3

Add columns: 2 + 10 + 3

TOTAL: = 15

10. If you checked off <i>any</i> problems, how <i>difficult</i> have these problems made it for you to do your work, take care of things at home, or get along with other people?	Not difficult at all	_____
	Somewhat difficult	_____ ✓
	Very difficult	_____
	Extremely difficult	_____

**Figure 3.5** Example of a possible severity rating scale for major depression in the DSM-5. Developed by Drs. Robert Spitzer, Janet B.W. Williams, Kurt Kroenke, and colleagues, with an educational grant from Pfizer Inc. Drawn from Pfizer.com.

diagnoses of hypoactive sexual desire disorder and female sexual arousal disorder are replaced with the DSM-5 diagnosis of sexual interest/arousal disorder in women.

**Clearer Criteria** For many disorders, criteria have been rewritten to provide clearer guidance about thresholds for diagnosis. For example, duration and intensity rules have been added for some diagnoses. For some diagnoses, criteria have been changed to reflect new information. Across diagnoses, many criteria have been reworded for clarity.

**Ethnic and Cultural Considerations in Diagnosis** Mental illness is universal. There is not a single culture in which people are free of mental illness. But there are many different cultural influences on the risk factors for mental illness (e.g., social cohesion, poverty, access to

drugs of abuse, and stress), the types of symptoms experienced, the willingness to seek help, and the treatments available. Sometimes these differences across cultures are profound. For example, although mental health care is widely available in the United States, it is estimated that there is only one psychiatrist for every 2 million people living in sub-Saharan Africa (World Health Organization, 2001, p. 17).

Cultural differences do not always play out in the way one might expect. For example, even with the access to medical care in the United States, a major study found that outcomes for schizophrenia were more favorable in Nigeria, India, and Colombia than in more industrialized countries, including the United States (Sartorius, 1986). People who immigrate from Mexico to the United States are initially about half as likely to meet criteria for mental illness as native born citizens in the United States, but over time, they and their children begin to show an increase in certain disorders, such as substance abuse, such that their risk for disorder begins to approximate that of people born in the United States (Alegria, 2008). As shown in Table 3.2, rates of mental illnesses tend to be higher in the United States than in many other countries. If we hope to understand how culture influences risk, symptom expression, and outcomes, we need a diagnostic system that can be applied reliably and validly in different countries and cultures.

**Table 3.2 Twelve-Month Prevalence Rates of the Most Common DSM-IV-TR Diagnoses by Country**

Country	Anxiety Disorders	Mood Disorders	Substance Disorders	Any Psychological Disorder
<b>Americas</b>				
Colombia	10.0	6.8	2.8	17.8
Mexico	6.8	4.8	2.5	12.2
United States	18.2	9.6	3.8	26.4
<b>Europe</b>				
Belgium	6.9	6.2	1.2	12.0
France	12.0	8.5	0.7	18.4
Germany	6.2	3.6	1.1	9.1
Italy	5.8	3.8	0.1	8.2
Netherlands	8.8	6.9	3.0	14.9
Spain	5.9	4.9	0.3	9.2
<b>Middle East and Africa</b>				
Lebanon	11.2	6.6	1.3	16.9
Nigeria	3.3	0.8	0.8	4.7
<b>Asia</b>				
Japan	5.3	3.1	1.7	8.8
Beijing	3.2	2.5	2.6	9.1

Source: The WHO World Mental Health Survey Consortium (2004).

Anxiety disorders include agoraphobia, generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, social phobia, and specific phobia. Mood disorders include bipolar I and II disorders, dysthymia, and major depressive disorder. Substance disorders include alcohol or drug abuse or dependence. Diagnoses were assessed with the Composite International Diagnostic Interview. Values are percentages.

Note: In the European countries, bipolar disorders and non-alcohol-related substance use disorders were not assessed. Obsessive-compulsive disorder was not assessed in Asian countries.

Previous editions of the DSM were criticized for their lack of attention to cultural and ethnic variations in psychopathology. DSM-IV-TR enhanced cultural sensitivity in three ways: (1) by providing a general framework for evaluating the role of culture and ethnicity, (2) by



The core symptoms of depression appear to be similar cross-culturally. (Richard Nowitz/Photo Researchers, Inc.)

describing cultural factors and ethnicity for each disorder, and (3) by listing culture-bound syndromes in an appendix.

In the general framework, clinicians are cautioned not to diagnose symptoms unless they are atypical and problematic within a person's culture. People vary in the degree to which they identify with their cultural or ethnic group. Some value assimilation into the majority culture, whereas others wish to maintain close ties to their cultural background. In general, clinicians are advised to be constantly mindful of how culture and ethnicity influence diagnosis and treatment.

Attention is paid to how culture can shape the symptoms and expression of a given disorder. For example, the symptoms of both schizophrenia (e.g., delusions and hallucinations) and depression (e.g., depressed mood and loss of interest or pleasure in activities) are similar cross-culturally (Draguns, 1989). But as we will discuss in Chapter 6, it is more likely in Japan than in the United States for anxiety to be focused around fears of offending others (Kirmayer, 2001). In evaluating symptoms, clinicians also need to be aware that cultures may shape the language used to describe distress. In many cultures, for example, it is common to describe grief or anxiety in physical terms—"I am sick in my heart" or "My heart is heavy"—rather than in psychological terms.

The DSM includes 25 culture-bound syndromes in the appendix. Culture-bound syndromes are diagnoses that are likely to be seen within specific regions. It is important to note that these culture-bound syndromes are not just found in cultures outside the United States. For example, some argued for listing bulimia nervosa as a Western culture-bound syndrome, a topic we return to in more detail in Chapter 11. The following are some examples of syndromes listed in the DSM appendix.

- *Amok*. A dissociative episode in which there is a period of brooding followed by a violent and sometimes homicidal outburst. The episode tends to be triggered by an insult and is found primarily among men. Persecutory delusions are often present as well. The term is Malaysian and is defined by the dictionary as a murderous frenzy. You may have heard the phrase "run amok."
- *Ghost sickness*. An extreme preoccupation with death and those who have died, found among certain Native American tribes.
- *Drat*. A term used in India to refer to severe anxiety about the discharge of semen.
- *Koru*. Reported in South and East Asia, an episode of intense anxiety about the possibility that the penis or nipples will recede into the body, possibly leading to death.
- *Shenjing shuairuo* (neurasthenia). A common diagnosis in China, a syndrome characterized by fatigue, dizziness, headaches, pain, poor concentration, sleep problems, and memory loss.
- *Taijin kyofusho*. The fear that one could offend others through inappropriate eye contact, blushing, a perceived body deformation, or one's own foul body odor. This disorder is most common in Japan, but cases have been reported in the United States. Japanese cultural norms appear to prescribe more careful attention to social appropriateness and hierarchy, perhaps intensifying the risk of these symptoms (Fabrega, 2002).
- *Hikikomori* (withdrawal). This refers to a syndrome observed in Japan, Taiwan, and South Korea in which an individual, most often an adolescent boy or young adult man, shuts himself into a room (e.g., bedroom) for a period of 6 months or more and does not socialize with anyone outside the room.

Some have argued that we should try to identify broad syndromes that can be identified across cultures and, in this light, have argued against the inclusion of culture-bound syndromes (Lopez-Ibor, 2003). In support of this position, they point toward a number of culture-bound syndromes that are not so different from the main DSM diagnoses. For example, Kleinman (1986) interviewed 100 Chinese people who had been diagnosed with *shenjing shuairuo* and

## Clinical Case: Lola: An Example of Diagnosis

Lola is a 17-year-old high school junior. She moved to the United States from Mexico with her parents and brother when she was 14 years old. A few months after they arrived, Lola's father returned to Mexico to attend the funeral of his brother. He was denied reentry to the United States due to a problem with his visa, and he has been unable to reunite with the family for nearly 3 years. Lola's mother has found it difficult to make ends meet on her salary as a bookkeeper, and the family was forced to move to a rougher neighborhood a year ago. Lola's English was fairly good when she came to the United States, and she has picked up many of the nuances of the language since arriving in the country. For the past 2 years, she has been dating a boy in her school. They have been fairly constant companions, and she describes him as the one person she would turn to if she was feeling upset. If her mother had any previous concern about Lola, it was that she seemed to rely on her boyfriend too much—she asked for his advice with small and large decisions, and she seemed wary of social interactions when he wasn't present. Lola's mother stated, "It is as though she is afraid to think for herself." Lola's mother noted that she

had always been a bit shy and had tended to count on her brother a lot for decisions and social support when she was younger.

With little warning, her boyfriend announced that he wanted to break up with her. Lola was extremely distressed by this change and reported that almost immediately she was unable to sleep or eat. She lost weight rapidly and found herself unable to concentrate on her schoolwork. Friends complained that she no longer wanted to talk during lunch or by phone. After 2 weeks of steadily feeling worse, Lola left a suicide note and disappeared. Police found her the next day in an abandoned home, holding a bottle of medicines. She reported that she had been sitting there all night, considering ending her life. Lola's mother reported that she had never seen her this distressed but noted that a few other family members had struggled with periods of sadness. Still, these family members in Mexico had not made suicide attempts nor had they received any formal treatment. Instead, the family learned to give these family members support and time to heal on their own. After the police found Lola, she was hospitalized for intensive treatment.

### DSM-IV-TR Diagnosis

- Axis I Major depressive disorder
- Axis II Dependent personality disorder
- Axis III None
- Axis IV Problems with primary support group (father not with family); problems related to social environment (acculturation stress; relationship with boyfriend)
- Axis V GAF: 25

### Likely-DSM-5 Diagnosis

- Major depressive disorder
- Personality disorder trait specified
  - Level of functioning: 1
  - Traits of submissiveness and separation insecurity

found that 87 percent of them met criteria for major depressive disorder. Many of those responded to antidepressant medications. Suzuki and colleagues (2003) have pointed out that the symptoms of *taijin kyofusho* overlap with those of social phobia (excessive fear of social interaction and evaluation) and body dysmorphic disorders (the mistaken belief that one is deformed or ugly), which are more commonly diagnosed in the United States. Other syndromes may reflect the common concerns of anxiety and distress, with the content shaped by life circumstances and values (Lopez-Ibor, 2003). Hence, some researchers believe it is important to look for commonalities across cultures. In contrast, others believe that culture-bound syndromes are central, because local and personal meanings are a key issue in understanding mental illness (Gaw, 2001).

In the planning process for DSM-5, one study group was dedicated to considering gender and culture issues. They recommended ways to keep culture salient for clinicians. As one example, the DSM-IV-TR includes an appendix on culture and diagnosis. More than half of clinicians surveyed reported that they didn't realize the appendix existed (Kirmayer et al., 2008). In the proposed DSM-5, this material is included in the introductory material on diagnostic assessment (Alarcón et al., 2009).



A therapist must be mindful of the role of cultural differences in the ways in which patients describe their problems. (© Chris Schmidt/iStockphoto.)



## Returning to Clinical Case: Roxanne: A Second Example of a Diagnosis

Previously, we described the case of Roxanne, who was brought to the psychiatric emergency room by the police. The

DSM-IV-TR and likely DSM-5 diagnosis for Roxanne might look as follows.

### DSM-IV-TR Diagnosis

- Axis I Bipolar I disorder, manic
- Axis II None
- Axis III High blood pressure
- Axis IV Problems with housing (homeless)
- Axis V GAF: 20

### Likely DSM-5 Diagnosis

- Bipolar I disorder, current or most recent episode manic
- High blood pressure

## Quick Summary

Because diagnosis provides the first step in thinking about the causes of symptoms, it is the first step in planning treatment. Because psychopathology is diagnosed on the basis of symptoms, clinical interviews are used to make diagnoses.

With all assessments, the reliability (the consistency of measurement) and validity (whether an assessment measures what it is designed to measure) should be evaluated. Reliability can be estimated by examining how well raters agree, how consistent test scores are over time, how alternate forms of a test compare, or how well items correlate with each other. There are many different forms of validity, including content, criterion, and construct validity.

Diagnostic systems for mental illness have changed a great deal in the past 100 years. DSM-III introduced explicit rules for

diagnosis. The system in use currently, the DSM-IV-TR, introduced several features to improve cultural sensitivity, such as providing a framework for clinicians to evaluate the role of culture, detailing the ways in which culture might influence symptoms of disorder, and including an appendix to describe culture-bound symptoms. DSM-5 has been drafted and is in the process of being reviewed and tested. Key changes include severity ratings that are specific to each disorder, a new approach to diagnosing personality disorder, and a reduction in the number of axes to be rated. Drawing on research evidence, the DSM-5 chapters are reorganized to reflect current knowledge of etiology. Some disorders are added, some are removed, and others are combined. The release of the final DSM-5 is expected in 2013.

## Check Your Knowledge 3.1 (Answers are at the end of the chapter.)

Answer the questions.

1. Major changes in the likely DSM-5 include (circle all that apply):
  - a. more axes
  - b. inclusion of severity ratings
  - c. a greater number of personality disorder diagnoses
  - d. many fewer diagnoses
2. Which type of reliability or validity is tested with the following procedures?
  - \_\_\_\_\_ A group of high school students is given the same IQ test 2 years in a row.
  - \_\_\_\_\_ A group of high school students is given an IQ test, and their scores are correlated with a different IQ test they took the year before.

\_\_\_\_\_ A measure of the tendency to blame oneself is developed, and researchers then test whether it predicts depression, whether it is related to childhood abuse, and whether it is related to less assertiveness in the workplace.

\_\_\_\_\_ Patients are interviewed by two different doctors. Researchers examine whether the doctors agree about the diagnosis.

- a. interrater reliability
- b. test-retest reliability
- c. criterion validity
- d. construct validity

## Specific Criticisms of the DSM

Some specific questions and concerns have been raised about the DSM. We review some of these concerns in the following sections.

**Too Many Diagnoses?** DSM-IV-TR contains almost 300 different diagnoses. Some have critiqued the burgeoning number of diagnostic categories (see Table 3.3). As one example, the DSM-IV and likely DSM-5 include a category for acute stress disorder in order to capture symptoms in the first month after a severe trauma. Should these relatively common reactions to trauma be pathologized by diagnosing them as a mental disorder (Harvey & Bryant, 2002)? By expanding its coverage, the authors of the DSM seem to have made too many problems into psychiatric disorders, without good justification for doing so.

Others argue that the system includes too many minute distinctions based on small differences in symptoms. One side effect of the huge number of diagnostic categories is a phenomenon called **comorbidity**, which refers to the presence of a second diagnosis. Comorbidity is the norm rather than the exception. Among people who meet criteria for at least one DSM-IV-TR psychiatric diagnosis, 45 percent will meet criteria for at least one more psychiatric diagnosis (Kessler, 2005). Some argue that this overlap is a sign that we are dividing syndromes too finely (Hyman, 2010).

A more subtle issue about the large number of diagnoses is that many risk factors seem to trigger more than one disorder. For example, some genes increase the risk for externalizing disorders as a whole (Kendler et al., 2003). Early trauma, dysregulation of stress hormones, tendencies to attend to and remember negative information about the self, and neuroticism all seem to increase risk for a broad range of anxiety disorders as well as mood disorders (Harvey et al., 2004). Anxiety and mood disorder also seem to share overlap in genes (Kendler, 2003), diminished function of a brain region called the prefrontal cortex (Hyman, 2010), and low serotonin function (Carver, Johnson, & Joormann, 2008). Similarly, selective serotonin reuptake inhibitors (SSRIs), such as Prozac, often seem to relieve symptoms of anxiety as well as depression (Van Ameringen, 2001). Different diagnoses do not seem to be distinct in their etiology or treatment.

Does this mean that we should lump some of the disorders into one category? Beliefs about lumping versus splitting differ. Some think we should keep the finer distinctions, whereas others believe we should lump (Watson, 2005). Among people who think there are too many diagnostic categories, several researchers have considered ways to collapse into broader categories. To begin, some disorders seem to co-occur more frequently than do others. For example, a person with antisocial personality disorder is highly likely to meet diagnostic criteria for a substance use disorder. In the DSM, these are diagnosed as separate disorders. Some have argued that childhood conduct disorder, adult antisocial personality disorder, alcohol use disorder, and substance use disorder co-occur so often that they should be considered different manifestations of one underlying disease process or vulnerability (Krueger, 2005). These different types of problems could be jointly considered “externalizing disorders.”

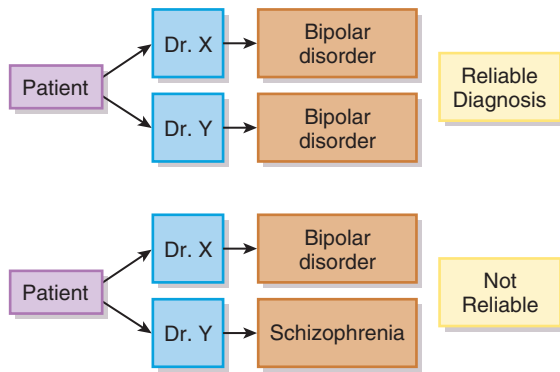
The authors of DSM-5 took modest steps toward addressing these concerns. In a few cases, two disorders were combined into one disorder. For example, as noted previously, the DSM-IV-TR diagnoses of substance abuse and dependence are replaced with the proposed DSM-5 diagnosis of substance use disorder. The new diagnosis of mixed anxiety depressive disorder is included in the likely DSM-5 because of the large number of people who present with both anxiety and depressive symptoms. The changes in DSM-5 are small, though. It includes more than 300 diagnoses, and comorbidity will remain the norm.

**Reliability of the DSM in Everyday Practice** Suppose you were concerned about your mental health, and you went to see two psychologists. Consider the distress you would feel if the two psychologists disagreed—one told you that you had schizophrenia, and the other told you that you had bipolar disorder. Diagnostic systems must have high interrater reliability to be useful. Before DSM-III, reliability for DSM diagnoses was poor, mainly because the criteria for making a diagnosis were not clear (see Figure 3.6 for an illustration of interrater reliability).

**Table 3.3 Number of Diagnostic Categories per Edition of DSM**

Edition of DSM	Number of Categories
DSM I	106
DSM-II	182
DSM-III	265
DSM-III-R	292
DSM-IV-TR	297
Proposed DSM-5	>300

Source: Pincus et al. (1992).



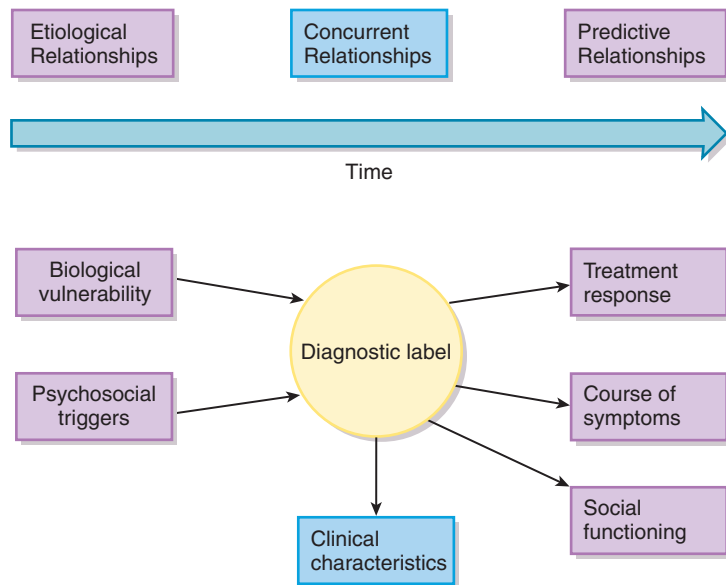
**Figure 3.6** Interrater reliability. In this example, the diagnosis of the first patient is reliable—both clinicians diagnose bipolar disorder—whereas the diagnosis of the second is not reliable.

The increased explicitness of the DSM criteria has improved reliability (see Table 3.1). Nonetheless, because clinicians might not rely on the criteria precisely, the reliability of the DSM in everyday usage may be lower than that seen in research studies. Even when following criteria, there is some room for disagreement in DSM-5. Consider again the criteria for mania in Table 3.1. What does it mean to say that mood is “abnormally” elevated . . . ? Or when is “involvement in pleasurable activities that have a high potential for painful consequences” excessive? Such judgments set the stage for the insertion of cultural biases as well as the clinician’s own personal ideas of what the average person should be doing. Because different clinicians may adopt different definitions for symptoms like “elevated mood,” achieving high reliability can be a challenge.

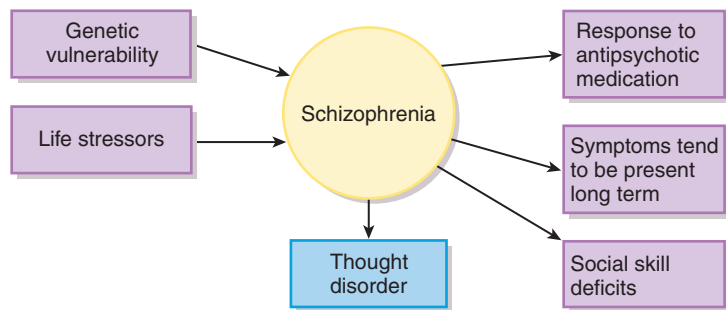
**How Valid Are Diagnostic Categories?** The DSM diagnoses are based on a pattern of symptoms. A diagnosis of schizophrenia, then, does not have the same status as a diagnosis of, say, diabetes, for which we have laboratory tests.

One way of thinking about diagnosis is to ask whether the system helps organize different observations (see Figure 3.7). Diagnoses have construct validity if they help make accurate predictions. What types of predictions should a good diagnostic category facilitate? One would hope that a diagnosis would inform us about related clinical characteristics and about functional impairments. The DSM specifies that impairment or distress must be present to meet criteria for a diagnosis, so perhaps it is not surprising that diagnoses are related to functional impairments such as marital distress and missed days at work (see Table 3.4). Beyond capturing the most common difficulties for a person with a diagnosis, one would hope that a diagnosis would inform us about what to expect next—the likely course of the disorder and response to different treatments. Perhaps most importantly, one would hope that the diagnosis relates to possible causes of the disorder, for example, a genetic predisposition or a biochemical imbalance. A diagnosis with strong construct validity should help predict a broad range of characteristics.

The central question, then, is whether diagnoses made with the DSM criteria reveal anything useful about patients. We have organized this book around the major DSM diagnostic categories because we believe that they do indeed possess some construct validity. Certain categories have less validity than others, however, and we will discuss some gaps in the validity of specific diagnostic categories in later chapters.



(a)



(b)

**Figure 3.7** Construct validity. An example of the types of information a diagnosis might help predict.

### General Criticisms of Diagnosing Mental Illness

Although we described many advantages of diagnosis in the beginning of this chapter, it is also clear that diagnoses can have negative effects on a person. Consider how your life might be changed by receiving the diagnosis of schizophrenia. You might become worried that someone will recognize your disorder. Or you might fear the onset of another episode. You might worry about your ability to deal with new challenges. The fact that you are a “former mental patient” could have a stigmatizing effect. Friends and loved ones might treat you differently, and employment might be hard to find.

There is little doubt that hearing a diagnosis can be difficult. Research shows that many view people with mental illness negatively, and patients and their families often encounter stigma against

**Table 3.4 Rates of Marital Distress and Missed Work Days among People with Mental Illness in the Past Year**

Disorder	Odds of Marital Distress for a Given Diagnosis Compared to No Mental Illness	Odds of Missed Work Days for a Given Diagnosis Compared to No Mental Illness
Panic disorder	1.28	3.32
Specific phobia	1.34	2.82
Social phobia	1.93	2.74
Generalized anxiety disorder	2.54	1.15
Posttraumatic stress disorder	2.30	2.05
Major depressive disorder	1.68	2.14
Bipolar I or II disorder	3.60	Not assessed
Alcohol use disorder	2.78	2.54

*Note:* Age, gender, education, and race/ethnicity are controlled for in marital distress analyses, and age and gender are controlled for in work-loss analyses. Diagnoses were based on the Composite International Diagnostic Interview. Marital distress was measured using a 14-item version of the Dyadic Adjustment Scale. Missed work days were measured during the month before the interview.

*Source:* Information on marital distress drawn from M. A. Whisman (2007). Information on work-loss days drawn from The ESEMeD/MHEDEA 2000 investigators (2004).

mental illness (Wahl, 1999), which, as we discussed in Chapter 1, remains a huge problem. Many have raised concerns that a diagnosis might contribute to stigma. To study this, researchers have given people brief written descriptions of a target individual. Beyond including a bit of information about the person's life and personality, the descriptions include either a mental health diagnosis (such as schizophrenia or bipolar disorder), a description of their symptoms (such as periods of high moods, decreased sleep, and restlessness), both (a diagnosis and symptoms), or neither. In this way, researchers can examine whether people tend to be more negative about labels or behavior. Research clearly shows that people tend to view the behaviors negatively. Sometimes labels may actually relieve stigma by providing an explanation for the symptomatic behavior (Lilienfeld et al., 2010). Of course, making a diagnosis is still a serious process that warrants sensitivity and privacy. But it may not be fair to presume that diagnostic labels are the major source of stigma.

Another concern is that when a diagnostic category is applied, we may lose sight of the uniqueness of that person. Because of this concern, the American Psychological Association recommends that people avoid using words like *schizophrenic* or *depressive* to describe people. Consider that we do not call people with medical illnesses by their disease (e.g., you aren't likely to hear someone with cancer described as the *canceric*). Rather, psychologists are encouraged to use phrases such as *a person with schizophrenia*.

Even with more careful language, some maintain that diagnosis leads us to focus on illnesses and, in doing so, to ignore important differences among people. Unfortunately, this criticism ignores a fundamental truth: it is human nature to categorize whenever we think about anything. Some would argue, then, that if we use categories anyway, it is best to systematically develop the categories. If one accepts this perspective, then the question is how well the current system does in grouping similar illnesses.

## Quick Summary

Despite the major improvements in the DSM, a number of problems remain. Some argue that there are too many diagnoses. Reliability is substantially higher than it was for DSM-II, but there is still some disagreement across clinicians regarding some diagnoses, and the

reliability achieved in practice may not be as high as the reliability achieved in research studies. Finally, the field as a whole faces a huge challenge; researchers are focused on validating this diagnostic system by trying to identify the causal patterns, symptom





patterns, and treatment that can be predicted by a given diagnosis. In sum, although the DSM is continually improving, it is far from perfect. Regardless of which diagnostic system is used, there are certain problems inherent in diagnosing people with mental illness. It is important to be aware of the tendency to ignore a person's strengths when focusing on diagnoses. The American Psychological

Association recommends using phrases such as *person with schizophrenia* rather than *schizophrenic* as one way to acknowledge that a person is much more than his or her diagnosis. Although many worry that stigma can be increased by applying labels, diagnoses can sometimes relieve stigma by providing a way of understanding symptoms.

## Check Your Knowledge 3.2

Answer the questions.

1. List three reasons why some think DSM should lump diagnoses.

2. What are three broad types of characteristics that a valid diagnosis should help predict?

## Psychological Assessment

To make a diagnosis, mental health professionals can use a variety of assessment measures and tools. Beyond helping to make a diagnosis, psychological assessment techniques are used in other important ways. For example, assessment methods are often used to identify appropriate therapeutic interventions. And repeated assessments are very useful in monitoring the effects of treatment over time. In addition, assessments are fundamental to conducting research on the causes of disorder.

We will see that beyond the basic interview, many of the assessment techniques stem from the paradigms presented in Chapter 2. Here we discuss clinical interviews; measures for assessing stress; personality tests, including objective and projective tests; intelligence tests; and behavioral and cognitive assessment techniques. Although we present these methods individually, a complete psychological assessment of a person will often entail combining several assessment techniques. The data from the various techniques complement each other and provide a more complete picture of the person. In short, there is no one best assessment measure. Rather, using multiple techniques and multiple sources of information will provide the best assessment.



Although it is illegal to discriminate based on mental illness, many employers do so. Stigma must be considered when giving a person a diagnosis of a mental disorder. (Ryan McVay/PhotoDisc, Inc./Getty Images.)

### Clinical Interviews

Most of us have probably been interviewed at one time or another, although the conversation may have been so informal that we did not regard it as an interview. For mental health professionals, both formal and structured as well as informal and less structured clinical interviews are used in psychopathological assessment.

**Characteristics of Clinical Interviews** One way in which a **clinical interview** is different from a casual conversation is the attention the interviewer pays to how the respondent answers questions—or does not answer them. For example, if a person is recounting marital conflicts, the clinician will generally be attentive to any emotion accompanying the comments. If the person does not seem upset about a difficult situation, the answers probably will be understood differently from how they would be interpreted if the person was crying or agitated while relating the story.

Great skill is necessary to carry out good clinical interviews. Clinicians, regardless of the paradigm adopted, recognize the importance of establishing rapport with the client. The interviewer must obtain the trust of the person; it is naive to assume that a client will easily reveal information to another, even to an authority figure with the title “Doctor.” Even a client who sincerely, perhaps desperately, wants to recount intensely personal problems to a professional may not be able to do so without help.

Most clinicians empathize with their clients in an effort to draw them out and to encourage them to elaborate on their concerns. An accurate summary statement of what the client has been saying can help sustain the momentum of talk about painful and possibly embarrassing events and feelings, and an accepting attitude toward personal disclosures dispels the fear that revealing “secrets of the heart” (London, 1964) to another human being will have disastrous consequences.

Interviews vary in the degree to which they are structured. In practice, most clinicians probably operate from only the vaguest outlines. Exactly how information is collected is left largely up to the particular interviewer and depends, too, on the responsiveness and responses of the interviewee. Through years of training and clinical experience, each clinician develops ways of asking questions that he or she is comfortable with and that seem to draw out the information that will be of maximum benefit to the client. Thus, to the extent that an interview is unstructured, the interviewer must rely on intuition and general experience. As a consequence, reliability for unstructured clinical interviews is probably lower than for structured interviews; that is, two interviewers may reach different conclusions about the same patient.

**Structured Interviews** At times, mental health professionals need to collect standardized information, particularly for making diagnostic judgments based on the DSM. To meet that need, investigators use a **structured interview**, in which the questions are set out in a prescribed fashion for the interviewer. One example of a commonly used structured interview is the Structured Clinical Interview (SCID) for Axis I of DSM-IV (Spitzer, Gibbon, & Williams, 1996). (Plans are under way to revise this and other structured interviews to cover DSM-5 criteria.)

The SCID is a branching interview; that is, the client’s response to one question determines the next question that is asked. It also contains detailed instructions to the interviewer concerning when and how to probe in detail and when to go on to questions about another diagnosis. Most symptoms are rated on a three-point scale of severity, with instructions in the interview schedule for directly translating the symptom ratings into diagnoses. The initial questions pertaining to obsessive-compulsive disorder (discussed in Chapter 7) are presented in Figure 3.8. The interviewer begins by asking about obsessions. If the responses elicit a rating of 1 (absent), the interviewer turns to questions about compulsions. If the patient’s responses again elicit a rating of 1, the interviewer is instructed to go to the questions for posttraumatic stress disorder. On the other hand, if positive responses (2 or 3) are elicited about obsessive-compulsive disorder, the interviewer continues with further questions about that problem.

Results of several studies demonstrate that the SCID achieves good interrater reliability for most diagnostic categories. As shown in Table 3.5, interrater reliability is a bit low for some of the anxiety disorders. Other structured interviews with good reliability have been developed for diagnosing personality disorders and for more specific disorders, such as the anxiety disorders, and for diagnosing disorders of childhood (DiNardo, 1993; Shaffer, 2000). With adequate training, interrater reliability for structured interviews is generally good (Blanchard & Brown, 1998).

In practice, most clinicians review the DSM symptoms in an informal manner without using a structured interview. Note, however, that clinicians using unstructured diagnostic interviews tend to miss comorbid diagnoses that often accompany a primary diagnosis (Zimmerman, 1999). When clinicians use an informal interview rather than a structured interview, the reliability of diagnoses also tends to be much lower (Garb, 2005).

**Table 3.5 Interrater Reliability of Selected DSM-IV-TR Diagnoses**

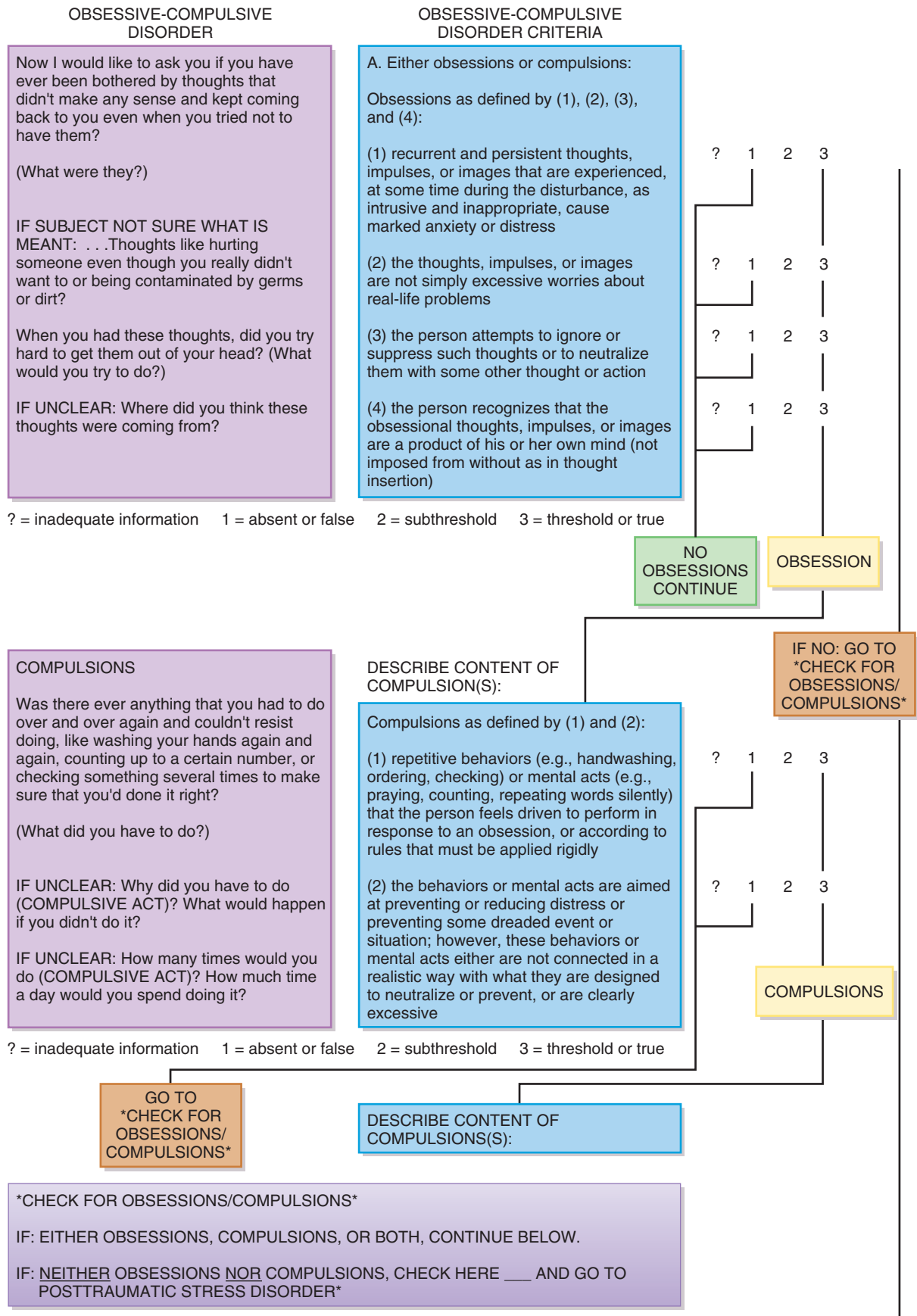
Diagnosis	Kappa
<b>Axis I disorders</b>	
Major depressive disorder	.80
Dysthymic disorder	.76
Bipolar disorder	.84
Schizophrenia	.79
Alcohol dependence/abuse	1.00
Other substance dependence/abuse	1.00
Panic disorder	.65
Social phobia	.63
Obsessive-compulsive disorder	.57
Generalized anxiety disorder	.63
Posttraumatic stress disorder	.88
Any eating disorder	.77
<b>Personality disorders</b>	
Avoidant	.97
Obsessive-compulsive	.83
Schizotypal	.91
Narcissistic	.98
Borderline	.91
Antisocial	.95

*Note:* The numbers here are a statistic called kappa, which measures the proportion of agreement over and above what would be expected by chance. Generally, kappas over 0.70 are considered good.

*Sources:* Estimates for bipolar disorder are based on a study using DSM-III-R criteria (Williams et al., 1992), which are largely comparable to DSM-IV-TR. Estimates for schizophrenia are drawn from Flaum et al. (1998). Other Axis I estimates are drawn from Zanarini et al. (2000), and Axis II estimates are based on Maffei et al. (1997).



Structured interviews are widely used to make reliable diagnoses. (© BSIP/Phototake.)



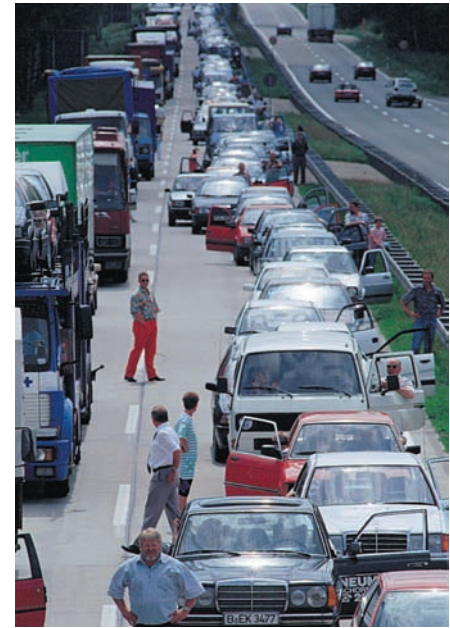
**Figure 3.8** Sample item from the SCID. Reprinted by permission of New York State Psychiatric Institute Biometrics Research Division. Copyright © 2004 Biometrics Research/New York State Psychiatric Institute.



## Assessment of Stress

Given its centrality to nearly all the disorders we consider in this book, measuring stress is clearly important in the total assessment picture. To understand the role of stress, we must first be able to define and measure it. Neither task is simple, as stress has been defined in many ways. See Focus on Discovery 3.2 for influential antecedents to our current conceptualizations of stress. Broadly, **stress** can be conceptualized as the subjective experience of distress in response to perceived environmental problems. Life stressors can be defined as the environmental problems that trigger the subjective sense of stress. Various scales and methods have been developed to measure life stress. Here we examine the most comprehensive measure of life stress: the Life Events and Difficulties Schedule (LEDS) as well as self-report checklist measures of stress.

**The Bedford College Life Events and Difficulties Schedule** This assessment is widely used to study life stressors (Brown & Harris, 1978). The LEDS includes an interview that covers over 200 different kinds of stressors. Because the interview is only semistructured, the interviewer can tailor questions to cover stressors that might only occur to a small number of people. The interviewer and the interviewee work collaboratively to produce a calendar of each of the major events within a given time period (see Figure 3.10 for an example). After the interview, raters evaluate the severity and several other dimensions of each stressor. The LEDS was designed to address a number of problems in life stress assessment, including the need to evaluate the importance of any given life event in the context of a person's life circumstances. For example, pregnancy might have quite a different meaning for an unmarried 14-year-old girl compared to a 38-year-old woman who has been trying to conceive for a long time. A second goal of the LEDS is to exclude life events that might just be consequences of symptoms. For example,



Stress can include major life events or daily hassles. (Herve Donnezan Photo Researchers, Inc.)

## FOCUS ON DISCOVERY 3.2

### A Brief History of Stress

The pioneering work by the physician Hans Selye set the stage for our current conceptualizations of stress. He introduced the term *general adaptation syndrome* (GAS) to describe the biological response to sustained and high levels of stress (see Figure 3.9). In Selye's model there are three phases of the response:

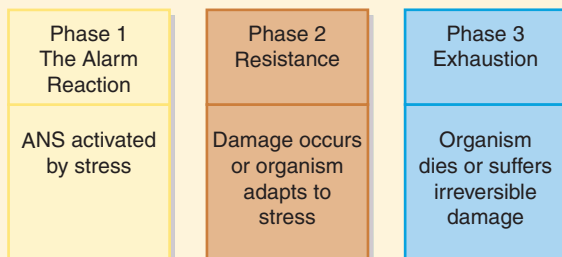


Figure 3.9 Selye's general adaptation syndrome.

1. During the first phase, the alarm reaction, the autonomic nervous system is activated by the stress.
2. During the second phase, resistance, the organism tries to adapt to the stress through available coping mechanisms.
3. If the stressor persists or the organism is unable to adapt effectively, the third phase, exhaustion, follows, and the organism dies or suffers irreversible damage (Selye, 1950).

In Selye's syndrome, the emphasis was on the body's response, not the environmental events that trigger that response. Psychological researchers later broadened Selye's concept to account for the diverse stress responses that people exhibited, including emotional upset, deterioration of performance, or physiological changes such as increases in the levels of certain hormones. The problem with these response-focused definitions of stress is that the criteria are not clear-cut. Physiological changes in the body can occur in response to a number of things that we would not consider stressful (e.g., anticipating a pleasurable event).

Other researchers defined stress as a stimulus, often referred to as a stressor, rather than a response, and identified stress with a long list of environmental conditions, such as electric shock, boredom, catastrophic life events, daily hassles, and sleep deprivation. Stimuli that are considered stressors can be major (the death of a loved one), minor (daily hassles, such as being stuck in traffic), acute (failing an exam), or chronic (a persistently unpleasant work environment). For the most part, they are experiences that people regard as unpleasant, but they can also be pleasant events.

Like response-based definitions of stress, stimulus-based definitions present problems. It is important to acknowledge that people vary widely in how they respond to life's challenges. A given event does not elicit the same amount of stress in everyone. For example, a family that has lost its home in a flood but has money enough to rebuild and strong social support from a network of friends nearby will experience less stress from this event than will a family that has neither adequate money to rebuild nor a network of friends to provide social support.

Current conceptualizations of stress emphasize that how we perceive, or *appraise*, the environment determines whether a stressor is present. Stress is perhaps most completely conceptualized as the subjective experience of

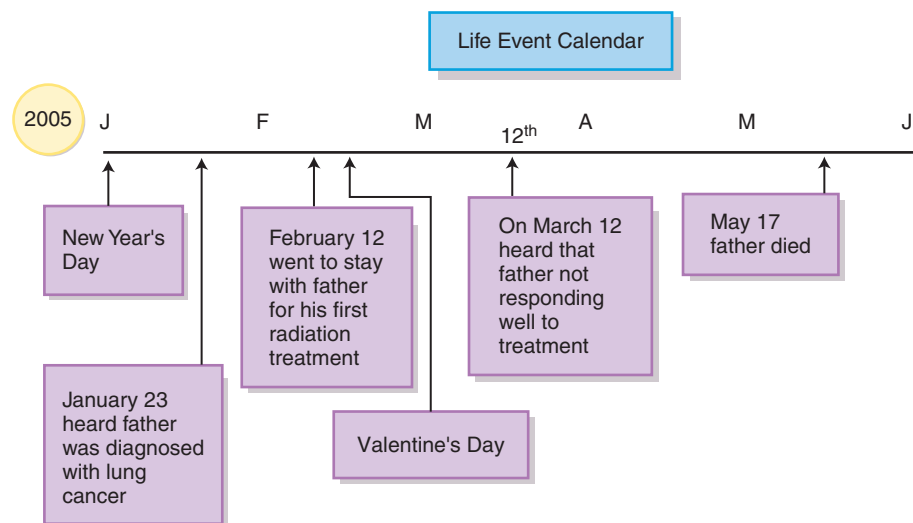
distress in response to perceived environmental problems. A final exam that is merely challenging to some students may be highly stressful to others who do not feel prepared to take it (whether their concerns are realistic or not).



The LEDES focuses on major stressors, such as deaths, job losses, and romantic breakups. (Bob Falcetti Reportage/Getty Images News and Sport Services.)

if a person misses work because he or she is too depressed to get out of bed, any consequent job problems should really be seen as symptoms of the disorder rather than a triggering life event. Finally, the LEDES includes a set of strategies to carefully date when a life stressor occurred. Using this more careful assessment method, researchers have found that life stressors are robust predictors of episodes of anxiety, depression, schizophrenia, and even the common cold (Brown & Harris, 1989b; Cohen et al., 1998).

**Self-Report Stress Checklists** Because intensive interview measures like the LEDES are so comprehensive, they take a good deal of time to administer. Often clinicians and researchers want a quicker way to assess stress and thus may turn to self-report checklists, such as the List of Threatening Experiences (LTE; Brugha & Cragg, 1990) or the Psychiatric Epidemiological Research Interview Life Events Scale (PERL; Dohrenwend et al., 1978). These checklists typically list different life events (e.g., death of a spouse, serious physical illness, major financial crisis), and participants are asked to indicate whether or not these events happened to them in a specified period of time. One difficulty associated with these types of measures is that there is a great deal of variability in how people view these events (Dohrenwend, 2006).



**Figure 3.10** Example of a life events timeline. The LEDES interview is designed to capture the major stressors a person has encountered in the past year.



For example, the death of a spouse could be the most horrible event ever for someone in a loving relationship. However, for someone in an abusive relationship, it might be the source of relief rather than stress. Other problems with such self-report checklists include difficulties with recall (Dohrenwend, 2006). For example, people may forget about some events. There is also evidence that people who are feeling depressed or anxious when they complete the measure may be biased in their responses. Perhaps because of these various issues influencing recall, test–retest reliability for life stress checklists can be low (McQuaid et al., 1992).

## Personality Tests

**Psychological tests** further structure the process of assessment. The two most common types of psychological tests are personality tests and intelligence tests. Here we will examine the two types of personality tests: self-report personality inventories and projective personality tests.

**Self-Report Personality Inventories** In a **personality inventory**, the person is asked to complete a self-report questionnaire indicating whether statements assessing habitual tendencies apply to him or her. When these tests are developed, they are typically administered to many people to analyze how certain kinds of people tend to respond. Statistical norms for the test can thereby be established. This process is called **standardization**. The responses of a particular person can then be compared with the statistical norms.

Perhaps the best known of these tests is the **Minnesota Multiphasic Personality Inventory (MMPI)**, developed in the early 1940s by Hathaway and McKinley (1943) and revised in 1989 (Butcher et al., 1989). The MMPI is called multiphasic because it was designed to detect a number of psychological problems. Over the years, the MMPI has been widely used to screen large groups of people for whom clinical interviews are not feasible.

In developing the test, the investigators used several steps. First, many clinicians provided statements that they considered indicative of various mental problems. Second, patients diagnosed with particular disorders and people with no diagnoses were asked to rate whether hundreds of statements described them. Items were selected for the final version of the test if patients in one clinical group responded to them more often in a certain way than did those in other groups.

With additional refinements, sets of these items were established as scales for determining whether a respondent should be diagnosed in a particular way. If a person answered many of the items in a scale in the same way as had a certain diagnostic group, his or her behavior was expected to resemble that of the particular diagnostic group. The 10 scales are described in Table 3.6.

The revised MMPI-2 (Butcher et al., 1989) was designed to improve validity and acceptability. The original sample assessed 65 years ago was composed mainly of white people from Minnesota and lacked representation of ethnic minorities. The new version was standardized using a sample that was much larger and more representative of 1980 U.S. census figures. Several items containing allusions to sexual adjustment, bowel and bladder functions, and excessive religiosity were removed because they were judged in some testing contexts to be needlessly intrusive and objectionable. Sexist wording was eliminated, along with outmoded idioms. New scales deal with substance abuse, emotions, and marital problems.

Aside from these differences, the MMPI-2 is otherwise quite similar to the original, having the same format, yielding the same scale scores and profiles (Ben-Porath & Butcher, 1989; Graham, 1988), and in general providing continuity with the vast literature already existing on the original MMPI (Graham, 1990). An extensive research literature shows that the MMPI-2 is reliable and has adequate criterion validity when it is related to diagnoses made by clinicians and to ratings made by spouses (Ganellan, 1996; Vacha-Hasse et al., 2001).

Like many other personality inventories, the MMPI-2 is typically administered and scored by computer. Many available computer programs even provide narratives about the respondent. Of course, the validity of the computer analysis is only as good as the program, which in turn is only as good as the competency and experience of the psychologist who wrote it. Figure 3.11 shows a hypothetical profile. Such profiles can be used in conjunction with a therapist's evaluation to help diagnose a client, assess personality functioning and coping style, and identify likely obstacles to treatment.

**Table 3.6 Typical Clinical Interpretations of Items Similar to Those on the MMPI-2**

Scale	Sample Item	Interpretation
? (Cannot say)	This is merely the number of items left unanswered or marked both true and false.	A high score indicates evasiveness, reading difficulties, or other problems that could invalidate results of the test. A very high score could also suggest severe depression or obsessional tendencies.
L (Lie)	I approve of every person I meet. (True)	Person is trying to look good, to present self as someone with an ideal personality.
F (Infrequency)	Everything tastes sweet. (True)	Person is trying to look abnormal, perhaps to ensure getting special attention from the clinician.
K (Correction)	Things couldn't be going any better for me. (True)	Person is guarded, defensive in taking the test, wishes to avoid appearing incompetent or poorly adjusted.
1. Hs (Hypochondriasis)	I am seldom aware of tingling feelings in my body. (False)	Person is overly sensitive to and concerned about bodily sensations as signs of possible physical illness.
2. D (Depression)	Life usually feels worthwhile to me. (False)	Person is discouraged, pessimistic, sad, self-deprecating, feeling inadequate.
3. Hy (Hysteria)	My muscles often twitch for no apparent reason. (True)	Person has somatic complaints unlikely to be due to physical problems; also tends to be demanding and histrionic.
4. Pd (Psychopathy)	I don't care about what people think of me. (True)	Person expresses little concern for social mores, is irresponsible, has only superficial relationships.
5. Mf (Masculinity–Femininity)	I like taking care of plants and flowers. (True, female)	Person shows nontraditional gender characteristics (e.g., men with high scores tend to be artistic and sensitive).
6. Pa (Paranoia)	If they were not afraid of being caught, most people would lie and cheat. (True)	Person tends to misinterpret the motives of others, is suspicious and jealous, vengeful and brooding.
7. Pt (Psychasthenia)	I am not as competent as most other people I know. (True)	Person is overanxious, full of self-doubts, moralistic, and generally obsessive-compulsive.
8. Sc (Schizophrenia)	I sometimes smell things others don't sense. (True)	Person has bizarre sensory experiences and beliefs, is socially reclusive.
9. Ma (Hypomania)	Sometimes I have a strong impulse to do something that others will find appalling. (True)	Person has overly ambitious aspirations and can be hyperactive, impatient, and irritable.
10. Si (Social Introversion)	Rather than spend time alone, I prefer to be around other people. (False)	Person is very modest and shy, preferring solitary activities.

*Note:* The first four scales assess the validity of the test; the numbered scales are the clinical or content scales.  
*Sources:* Hathaway & McKinley (1943); revised by Butcher et al. (1989).

You may wonder whether it would be easy to fake answers that suggest no psychopathology. For example, a superficial knowledge of contemporary psychopathology research could alert someone that to be regarded as psychologically healthy, he or she must not admit to worrying a great deal about receiving messages from television.

As shown in Table 3.6, the MMPI-2 includes several “validity scales” designed to detect deliberately faked responses. In one of these, the lie scale, a series of statements sets a trap for the person who is trying to look too good. An item on the lie scale might be, “I read the newspaper editorials every day.” The assumption is that few people would be able to endorse such a statement honestly. Persons who endorse a large number of the statements in the lie scale might be attempting to present themselves in a good light. High scores on the F scale also discriminate between people trying to fake psychopathology and real patients (Bagby et al., 2002). If a person obtains high scores on the lie or F scale, his or her profile might be viewed with skepticism. People who are aware of these validity scales, however, can effectively fake a normal profile (Baer & Sekirnjak, 1997; Walters & Clopton, 2000). In most testing circumstances, however, people do not want to falsify their responses because they want to be helped. Focus on Discovery 3.3 discusses other issues surrounding the validity of self-report questionnaires.

**Projective Personality Tests** A **projective test** is a psychological assessment tool in which a set of standard stimuli—inkblots or drawings—ambiguous enough to allow variation in responses is presented to the person. The assumption is that because the stimulus materials are unstructured and ambiguous, the person’s responses will be determined primarily by

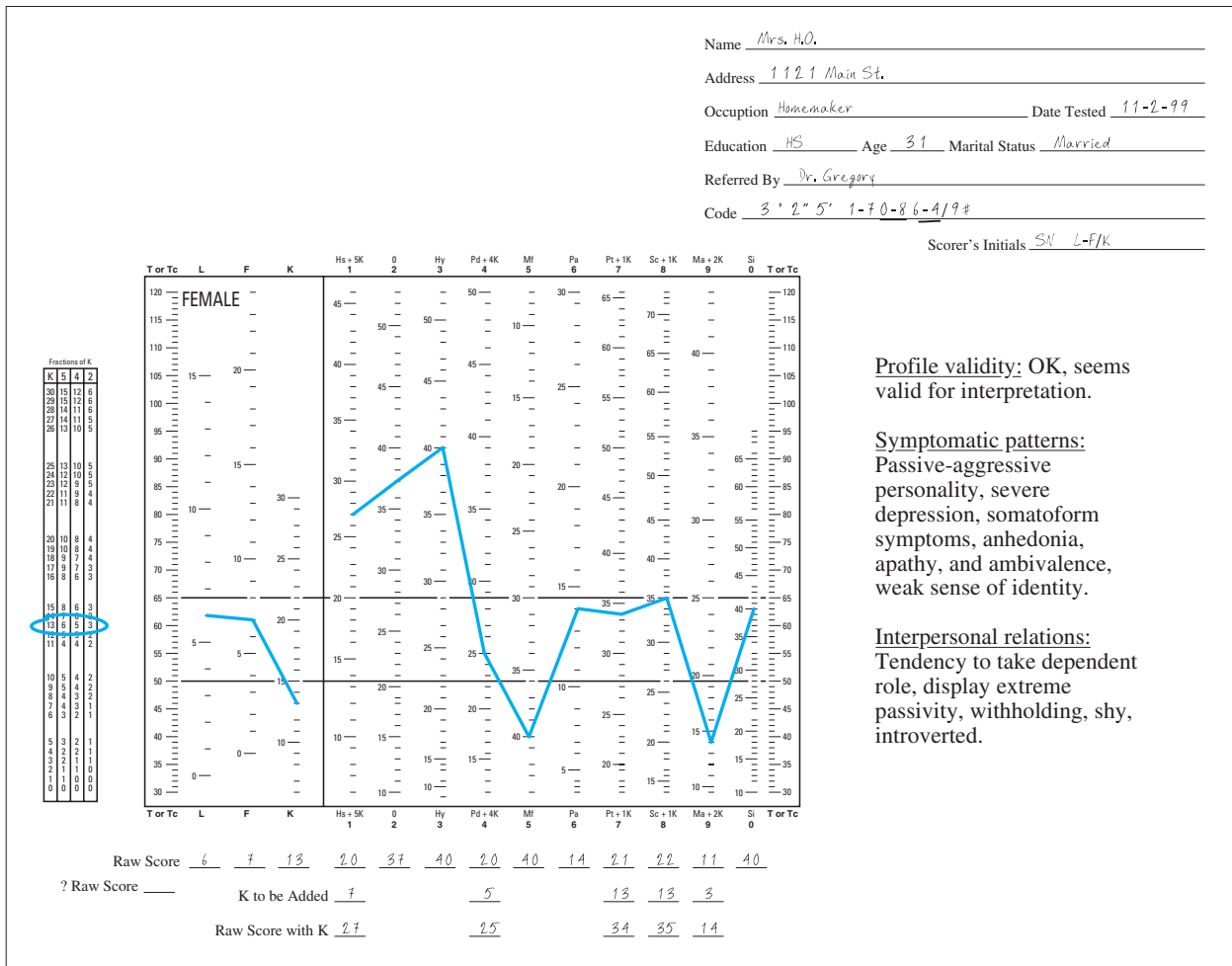


Figure 3.11 Hypothetical MMPI-2 profile.

**Profile validity:** OK, seems valid for interpretation.

**Symptomatic patterns:** Passive-aggressive personality, severe depression, somatoform symptoms, anhedonia, apathy, and ambivalence, weak sense of identity.

**Interpersonal relations:** Tendency to take dependent role, display extreme passivity, withholding, shy, introverted.

unconscious processes and will reveal his or her true attitudes, motivations, and modes of behavior. This notion is referred to as the **projective hypothesis**.

If a patient reports seeing eyes in an ambiguous inkblot, for example, the projective hypothesis might be that the patient tends toward paranoia. The use of projective tests assumes that the respondent would be either unable or unwilling to express his or her true feelings if asked directly. As you might have guessed, projective techniques are derived from the work of Freud and his followers (see Chapter 1).

### FOCUS ON DISCOVERY 3.3

#### Underreporting of Stigmatized Behaviors

A survey of self-reported drug use, sexual behavior, and violence highlights the importance of the setting in establishing the validity of what people will tell about their actions and attitudes (Turner et al., 1998). Findings from self-report questionnaires were compared with results from a novel self-report method—boys and young men (ages 15 to 19) listened by themselves through headphones to questions probing risky, often stigmatized behavioral practices and then indicated whether they had engaged in those behaviors by pressing keys on a computer keyboard labeled Yes and No.

Compared to a matched control group who responded to the same items on a paper-and-pencil questionnaire, many more of the computer respondents admitted to having engaged in a range of high-risk behaviors. For example, they were almost 14 times more likely to report having had sex with an intravenous drug user (2.8 percent versus 0.2 percent), more than twice as likely to report having been paid for sex (3.8 percent versus 1.6 percent), and almost twice as likely to report having used cocaine (6.0 percent versus 3.3 percent). (One can safely



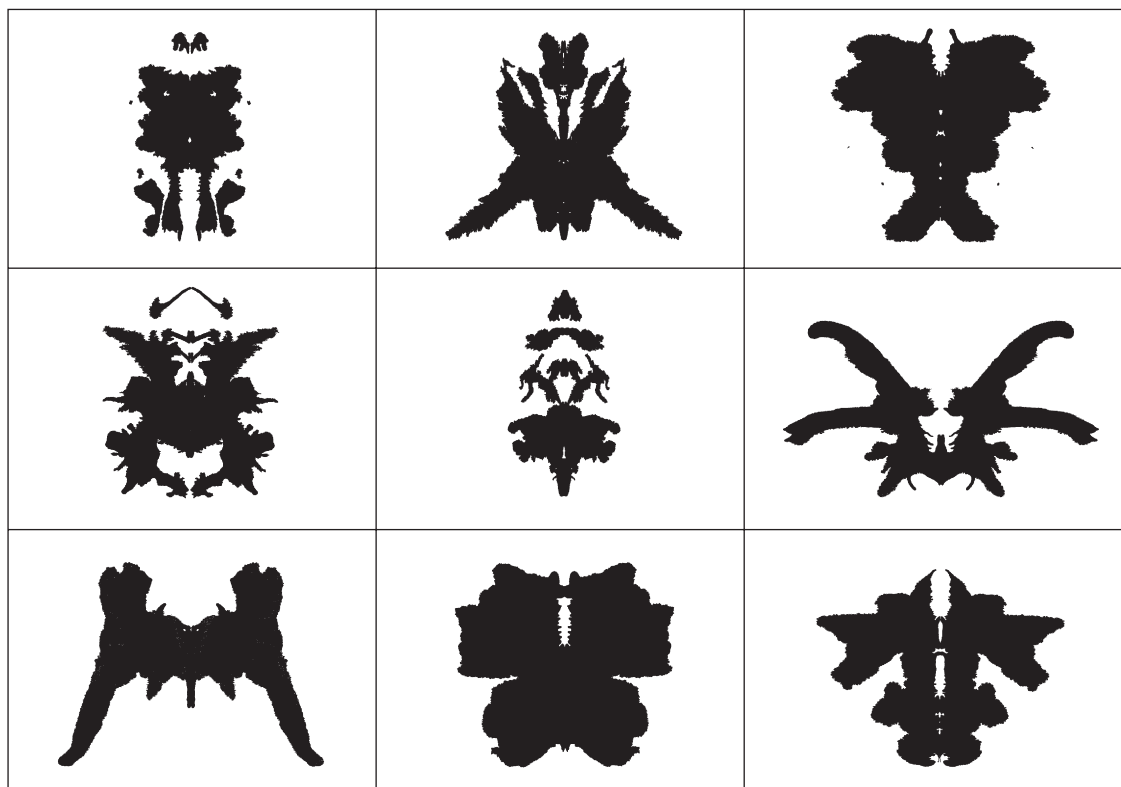
assume that the differences would have been even greater if the boys had been interviewed by an adult researcher facing them across a table, another method that has been used to collect such survey data.) No differences showed up on questions directed at nonstigmatized or legal behaviors such as having had sex with a female in the preceding year (47.8 percent for computer users versus 49.6 percent for paper-and-pencil questionnaires) or having drunk alcohol in the past year (69.2 percent versus 65.9 percent).

If these findings show nothing else, they strongly suggest that the frequencies of problematic behavior as determined by questionnaire or

interview studies may be underestimates and that social problems such as needle sharing and unsafe sex may be considerably more common than most people believe.

In an effort to obtain more accurate reports about stigmatized, sensitive, risky, or even illegal behaviors, investigators may apply for a Certificate of Confidentiality from the U.S. Department of Health and Human Services. These certificates provides additional protection for research participants by ensuring that sensitive information can be revealed during the research study without fear that the researchers will report their responses to legal or other authorities.

The **Thematic Apperception Test (TAT)** is a projective test. In this test a person is shown a series of black-and-white pictures one-by-one and asked to tell a story related to each. For example, a patient seeing a picture of a boy observing a youth baseball game from behind a fence may tell a story that contains angry references to the boy's parents. The clinician may, through the projective hypothesis, infer that the patient harbors resentment toward his or her parents. There are few reliable scoring methods for this test, and the norms are based on a small and limited sample (i.e., few norms for people of different ethnic or cultural backgrounds). The construct validity of the TAT is also limited (Lilienfeld, Wood, & Garb, 2000). The **Rorschach Inkblot Test** is perhaps the best-known projective technique. In the Rorschach test, a person is shown 10 inkblots (for similar inkblots, see Figure 3.12), one at a time, and asked to tell what the blots look like. Half the inkblots are in black, white, and shades of gray; two also have red splotches; and three are in pastel colors.



**Figure 3.12** In the Rorschach test, the client is shown a series of inkblots and is asked what the blots look like.

Exner (1978) designed the most commonly used system for scoring the Rorschach test. The Exner scoring system concentrates on the perceptual and cognitive patterns in a person's responses. The person's responses are viewed as a sample of how he or she perceptually and cognitively organizes real-life situations (Exner, 1986). For example, Erdberg and Exner (1984) concluded from the research literature that respondents who express a great deal of human movement in their Rorschach responses (e.g., "The man is running to catch a plane") tend to use inner resources when coping with their needs, whereas those whose Rorschach responses involve color ("The red spot is a kidney") are more likely to seek interaction with the environment. Rorschach suggested this approach to scoring in his original manual, *Psychodiagnostics: A Diagnostic Test Based on Perception* (1921), but he died only 8 months after publishing his 10 inkblots, and his immediate followers devised other methods of interpreting the test.

The Exner scoring system has norms, although the sample on which they are based was rather small and did not represent different ethnicities and cultures well. Regarding its reliability and validity, this work has enthusiastic supporters as well as equally harsh critics (e.g., Hunsley & Bailey, 1999; Lilienfeld et al., 2000; Meyer & Archer, 2001). Perhaps trying to make a blanket statement about the validity of the Rorschach (or the MMPI-2) is not the right approach. The test appears to have more validity in assessing some issues more than others. For example, limited evidence suggests that the Rorschach may have validity in identifying schizophrenia, borderline personality disorder, and dependent personality traits, but it remains unclear whether it does so better than other assessment techniques (Lilienfeld et al., 2000). In other words, it is unclear whether the Rorschach provides information that could not be obtained more simply—for example, through an interview.

## Intelligence Tests

Alfred Binet, a French psychologist, originally constructed tests to help the Parisian school board predict which children were in need of special schooling. Intelligence testing has since developed into one of the largest psychological industries. An **intelligence test**, often referred to as an IQ test, is a way of assessing a person's current mental ability. IQ tests are based on the assumption that a detailed sample of a person's current intellectual functioning can predict how well he or she will perform in school, and most are individually administered. The most commonly administered tests include the Wechsler Adult Intelligence Scale, 4th edition (WAIS-IV, 2008); the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV, 2003); the Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III, 2002); and the Stanford–Binet, 5th edition (SB5, 2003); IQ tests are regularly updated, and, like personality inventories, they are standardized.

Beyond predicting school performance, intelligence tests are also used in other ways:

- In conjunction with achievement tests, to diagnose learning disorders and to identify areas of strengths and weaknesses for academic planning
- To help determine whether a person has intellectual developmental disorder (formerly known as mental retardation; see Chapter 13)
- To identify intellectually gifted children so that appropriate instruction can be provided them in school
- As part of neuropsychological evaluations; for example, periodically testing a person believed to be suffering from dementia so that deterioration of mental ability can be followed over time

IQ tests tap several functions believed to constitute intelligence, including language skills, abstract thinking, nonverbal reasoning, visual-spatial skills, attention and concentration, and speed of processing. Scores on most IQ tests are standardized so that 100 is the mean (i.e., the average score) and 15 or 16 is the standard deviation (a measure of how scores are dispersed above and below the average). Approximately 65 percent of the population receives scores between 85 and 115. Approximately 2.5 percent of the population falls below 70 or above 130 (i.e., 2 standard deviations below or above the mean score of 100). In Chapter 14 we discuss people whose IQ falls at the low end of the distribution.



During a ride in the country with his two children, Hermann Rorschach (1884–1922), a Swiss psychiatrist, noticed that what they saw in the clouds reflected their personalities. From this observation came the famous inkblot test. (Courtesy National Library of Medicine.)





The French psychologist Alfred Binet developed the first IQ test to predict how well children would do in school. (Archives of the History of American Psychology, The Center for the History of Psychology-The University of Akron.)

IQ tests are highly reliable (e.g., Canivez & Watkins, 1998) and have good criterion validity. For example, they distinguish between people who are intellectually gifted and those with intellectual developmental disorder and between people with different occupations or educational attainment (Reynolds et al., 1997). They also predict educational attainment and occupational success (Hanson, Hunsley, & Parker, 1988), at least among Caucasians (see below for a discussion of cultural bias in assessment). Although IQ and educational attainment are positively correlated (see Chapter 4 for a discussion of correlational methods), what remains less clear is whether more education causes an increase in IQ or whether IQ causes one to attain more education (Deary & Johnson, 2010). Furthermore, although correlations between IQ scores and school performance are statistically significant, IQ tests explain only a small part of school performance; much more is unexplained by IQ test scores than is explained.

Of interest to the subject matter of this hook, IQ is also correlated with mental health. In one study of over one million Scandinavian men, lower IQ scores at age 20 were associated with a greater risk of hospitalization for schizophrenia, mood disorders, or substance dependence some 20 years later, even after controlling for other possible contributing factors, such as the participants' families' socioeconomic status (Gale et al., 2010). A recent meta-analysis of 16 prospective, longitudinal studies (see Chapter 4 for a description of these methods) found that lower IQ scores in early adulthood were associated with greater mortality risk (i.e., death) later in life, even after controlling for other variables such as socioeconomic status and educational attainment (Calvin et al., 2010).

Regarding construct validity, it is important to keep in mind that IQ tests measure only what psychologists consider intelligence. Factors other than what we think of as intelligence, however, also play an important role in how people will do in school, such as family and circumstances, motivation to do well, expectations, performance anxiety, and difficulty of the curriculum. Another factor relevant to IQ test performance is called stereotype threat. It suggests that the social stigma of poor intellectual performance borne by some groups (e.g.,

African Americans do poorly on IQ tests; women perform more poorly than men on mathematics tests) actually interferes with their performance on these tests. In one study demonstrating this phenomenon, groups of men and women were given a difficult mathematics test. In one condition the participants were told that men scored higher than women on the test they were going to take (stereotype threat condition), while in the other condition they were told there were no gender differences in performance on the test. Only when the test was described as yielding gender differences did the women perform more poorly than the men (Spencer, Steele, & Quinn, 1999).

Unfortunately, awareness of these stereotypes develops early. For example, a study revealed that children develop awareness of stereotypes regarding ethnicity and ability between the ages of 6 and 10, with 93 percent of children being aware of such stereotypes by age 10 (McKown & Weinstein, 2003). This awareness seems to influence stereotype threat (and performance). In the McKown and Weinstein (2003) study, children were asked to complete a puzzle task. Half of the children received instructions that the task reflected their ability (stereotype threat condition), and half the children received instructions that the test did not reflect their ability. African American children who were aware of the stereotype about ethnicity and ability showed evidence of stereotype threat. Specifically, among African American children, those who received the ability instructions performed more poorly on the puzzle task than the children who did not, suggesting that the instructions activated the stereotype and thus influenced their performance.

### Behavioral and Cognitive Assessment

Thus far, we have discussed assessment methods that measure personality traits and intellectual ability. Other types of assessment focus on behavioral and cognitive characteristics, including the following:



IQ tests have many subtests, including this test to assess spatial ability. (Bob Daemmrch/The Image Works.)

- Aspects of the environment that might contribute to symptoms (e.g., an office location next to a noisy hallway might contribute to concentration problems)
- Characteristics of the person (e.g., a client's fatigue may be caused in part by a cognitive tendency toward self-deprecation manifested in such statements as "I never do anything right, so what's the point in trying?")
- The frequency and form of problematic behaviors (e.g., procrastination taking the form of missing important deadlines)
- Consequences of problem behaviors (e.g., when a client avoids a feared situation, his or her partner offers sympathy and excuses, thereby unwittingly keeping the client from facing up to his or her fears)

The hope is that understanding these aspects of cognition and behavior will guide the clinician toward more effective intervention targets.

The information necessary for a behavioral or cognitive assessment is gathered by several methods, including direct observation of behavior in real life as well as in laboratory or office settings, interviews and self-report measures, and various other methods of cognitive assessment (Bellack & Hersen, 1998). We turn to these now.

**Direct Observation of Behavior** It is not surprising that cognitive behavior therapists have paid considerable attention to careful observation of behavior in a variety of settings, but it should not be assumed that they simply go out and observe. Like other scientists, they try to fit events into a framework consistent with their points of view. In formal behavioral observation, the observer divides the sequence of behavior into various parts that make sense within a learning framework, including such things as the antecedents and consequences of particular behaviors. Behavioral observation is also often linked to intervention (O'Brien & Haynes, 1995). The cognitive behavioral clinician's way of conceptualizing a situation typically implies a way to try to change it.

It is difficult to observe most behavior as it actually takes place, and little control can be exercised over where and when it may occur. For this reason, many therapists contrive artificial situations in their consulting rooms or in a laboratory so they can observe how a client or a family acts under certain conditions. For example, Barkley (1981) had a mother and her child spend time together in a laboratory living room, complete with sofas and a television set. The mother was given a list of tasks for the child to complete, such as picking up toys or doing arithmetic problems. Observers behind a one-way mirror watched the proceedings and reliably coded the child's reactions to the mother's efforts to control as well as the mother's reactions to the child's compliant or noncompliant responses. These **behavioral assessment** procedures yielded data that could be used to measure the effects of treatment.

**Self-Observation** Cognitive behavior therapists and researchers have also asked people to observe and track their own behavior and responses, an approach called **self-monitoring**. Self-monitoring is used to collect a wide variety of data of interest to both clinicians and researchers, including moods, stressful experiences, coping behaviors, and thoughts (Hurlburt, 1979; Stone et al., 1998).

Another method of self-observation is called **ecological momentary assessment**, or **EMA**. EMA involves the collection of data in real time as opposed to the more usual methods of having people reflect back over some time period and report on recently experienced thoughts, moods, or stressors. The methods for implementing EMA range from having people complete diaries at specified times during the day (perhaps signaled by a wristwatch that beeps at those times) to supplying them with smartphones that not only signal when reports



Behavioral assessment often involves direct observation of behavior, as in this case, where the observer is behind a one-way mirror. (© Spencer Grant/Alamy Limited.)





Self-monitoring generally leads to increases in desirable behaviors and decreases in undesirable ones. (ANDREW GOMBERT/ EPA/Landov LLC.)

are to be made but also allow them to enter their responses directly into the device (Stone & Shiffman, 1994).

Given the problems in retrospective recall, some theories in the field of psychopathology can best be tested using EMA. For example, current theories of both anxiety disorders and depression propose that emotional reactions to a life event are triggered in part by the thoughts that the event elicits. It is unlikely, however, that these thoughts can be recalled accurately in retrospect.

EMA may also be useful in clinical settings, revealing information that traditional assessment procedures might miss. For example, Hurlburt (1997) describes a case of a man with severe attacks of anxiety. In clinical interviews, the patient reported that his life was going very well, that he loved his wife and children, and that his work was both financially and personally rewarding. No cause of the anxiety attacks could be discerned. The man was asked to record his thoughts as he went about his daily routine. Surprisingly, about a third of his thoughts were concerned with annoyance with his children (e.g., “He left the fence gate open again and the dog got out”).

*Once the high frequency of annoyance thoughts was pointed out to him, he . . . accepted that he was in fact often annoyed with his children. However, he believed that anger at his children was sinful and felt unfit as a father for having such thoughts and feelings. . . . [He] entered into brief therapy that focused on the normality of being annoyed by one’s children and on the important distinction between being annoyed and acting out aggressively. Almost immediately, his anxiety attacks disappeared. (Hurlburt, 1997, p. 944)*

Although some research indicates that self-monitoring or EMA can provide accurate measurement of such behavior, considerable research indicates that behavior may be altered by the very fact that it is being self-monitored—that is, the self-consciousness required for self-monitoring affects the behavior (Haynes & Horn, 1982). The phenomenon wherein behavior changes because it is being observed is called **reactivity**. In general, desirable behavior, such as engaging in social conversation, often increases in frequency when it is self-monitored (Nelson, Lipinski, & Black, 1976), whereas behavior the person wishes to reduce, such as cigarette smoking, diminishes (McFall & Hammen, 1971). Therapeutic interventions can take advantage of the reactivity that is a natural by-product of self-monitoring. Smoking, anxiety, depression, and health problems have all undergone beneficial changes in self-monitoring studies (Febbraro & Clum, 1998). Beyond reactivity, self-monitoring with

portable electronic devices like smart phones has also been included effectively in cognitive behavior therapy for different anxiety disorders (Przeworski & Newman, 2006).



Cognitive assessment focuses on the person’s perception of a situation, realizing that the same event can be perceived differently. For example, moving could be regarded as a very negative event or a very positive one, resulting in very different levels of stress. (Fuse/Getty Images, Inc.)

**Cognitive-Style Questionnaires** Cognitive questionnaires tend to be used to help plan targets for treatment as well as to determine whether clinical interventions are helping to change overly negative thought patterns. In format, some of these questionnaires are similar to the personality tests we have already described.

One self-report questionnaire that was developed based on Beck’s theory (see Chapters 2 and 8) is the Dysfunctional Attitude Scale (DAS). The DAS contains items such as “People will probably think less of me if I make a mistake” (Weissman & Beck, 1978). Supporting construct validity, researchers have shown that they can differentiate between depressed and nondepressed people on the basis of their scores on this scale and that scores decrease (i.e., improve) after interventions that relieve depression. Furthermore, the DAS relates to other aspects of cognition in ways consistent with Beck’s theory (Glass & Arnkoff, 1997).

**Table 3.7 Psychological Assessment Methods**

Interviews	Clinical interviews	The clinician learns about the patient's problems through conversation. The paradigm of the interviewer shapes the content of the interview.
	Structured interviews	Questions to be asked are spelled out in detail in a booklet. The Structured Clinical Interview for Axis I Disorders is a structured interview that is commonly used to make a diagnosis.
Stress measures		Self-report scales or interviews that assess stressful events and responses to these events.
Psychological tests	Personality tests	Self-report questionnaires, used to assess either a broad range of characteristics, as in the MMPI-2, or a single characteristic, such as dysfunctional attitudes.
	Projective tests	Ambiguous stimuli, such as inkblots (Rorschach test), are presented and responses are thought to be determined by unconscious processes.
	Intelligence tests	Assessments of current mental functioning. Used to predict school performance and identify cognitive strengths and weaknesses.
Direct observation		Used by clinicians to identify problem behaviors as well as antecedents and consequences.
Self-observation		People monitor and keep records of their own behavior, as in ecological momentary assessment.

## Quick Summary

The psychological assessments we have described are summarized in Table 3.7. A comprehensive psychological assessment draws on many different methods and tests. Interviews can be structured, with the questions predetermined and followed in a certain order, or unstructured to follow more closely what the client tells the interviewer. Structured interviews are more reliable. Rapport is important to establish regardless of the type of interview.

Stress is best assessed via a semistructured interview that captures the importance of any given life event in the context of a person's life circumstances, as in the LEADS. Self-report checklists are also used to assess stress, but they have poorer reliability and validity than the LEADS.

The MMPI-2 is a standardized and objective personality inventory. The test has good reliability and validity and is widely used. Projective personality tests, like the Rorschach or TAT, are not as widely used today, likely due to their poor validity. Reliability can be achieved using scoring systems such as Exner's. Intelligence tests have been used for a number of years and are quite reliable. Like any test, there are limits to what an IQ test can tell a clinician or researcher.

Direct observation of behavior can be very useful in assessment, though it can take more time than a self-report inventory. Other behavioral and cognitive assessment methods include ecological momentary assessment (EMA).

## Check Your Knowledge 3.3

True or false?

1. If conducted properly, a psychological assessment typically includes just one measure most appropriate to the client.
2. Unstructured interviews may have poor reliability, but they can still be quite valuable in a psychological assessment.
3. The MMPI-2 contains scales to detect whether someone is faking answers.
4. The projective hypothesis is based on the idea that a person does not really know what is bothering him or her; thus, a subtler means of assessment is needed.
5. Intelligence tests are highly reliable.
6. EMA is a method to assess unwanted impulses.

## Neurobiological Assessment

Recall from Chapters 1 and 2 that throughout history people interested in psychopathology have assumed, quite reasonably, that some symptoms are likely to be due to or at least reflected in malfunctions of the brain or other parts of the nervous system. We turn now to contemporary work in neurobiological assessment. We'll look at four areas in particular: brain imaging, neurotransmitter assessment, neuropsychological assessment, and psychophysiological assessment (see Table 3.8 for a summary of these methods).

**Table 3.8 Neurobiological Assessment Methods**

Brain imaging	CT and MRI scans reveal the structure of the brain. PET reveals brain function and, to a lesser extent, brain structure. fMRI is used to assess both brain structure and brain function.
Neurotransmitter assessment	Includes postmortem analysis of neurotransmitters and receptors, assays of metabolites of neurotransmitters, and PET scans of receptors.
Neuropsychological assessment	Behavioral tests such as the Halstead–Reitan and Luria–Nebraska assess abilities such as motor speed, memory, and spatial ability. Deficits on particular tests help point to an area of brain dysfunction.
Psychophysiological assessment	Includes measures of electrical activity in the autonomic nervous system, such as skin conductance, or in the central nervous system, such as EEG.

### Brain Imaging: “Seeing” the Brain

Because many behavioral problems can be brought on by brain dysfunction, neurological tests—such as checking the reflexes, examining the retina for any indication of blood vessel damage, and evaluating motor coordination and perception—have been used for many years to identify brain dysfunction. Today, devices have become available that allow clinicians and researchers a much more direct look at both the structure and functioning of the brain.

Computerized axial tomography, the **CT** or **CAT scan**, helps to assess structural brain abnormalities (and is able to image other parts of the body for medical purposes). A moving beam of X-rays passes into a horizontal cross section of the person's brain, scanning it through 360 degrees; the moving X-ray detector on the other side measures the amount of radioactivity that penetrates, thus detecting subtle differences in tissue density. A computer uses the information to construct a two-dimensional, detailed image of the cross section, giving it optimal contrasts. Then the machine scans another cross section of the brain. The resulting images can show the enlargement of ventricles (which can be a sign of brain tissue degeneration) and the locations of tumors and blood clots.

Other devices for seeing the living brain include **magnetic resonance imaging**, also known as **MRI**, which is superior to the CT scan because it produces pictures of higher quality and does not rely on even the small amount of radiation required by a CT scan. In MRI the person is placed inside a large, circular magnet, which causes the hydrogen atoms in the body to move. When the magnetic force is turned off, the atoms return to their original positions and thereby produce an electromagnetic signal. These signals are then read by the computer and translated into pictures of brain tissue. This technique provides an enormous advance. For example, it has allowed physicians to locate delicate brain tumors that would have been considered inoperable without such sophisticated methods of viewing brain structures.



An fMRI scanner is a long tubelike structure. (age fotostock/SuperStock, Inc.)



An even greater advance has been a technique called **functional MRI (fMRI)**, which allows researchers to measure both brain structure and brain function. This technique takes MRI pictures so quickly that metabolic changes can be measured, providing a picture of the brain at work rather than of its structure alone. fMRI measures blood flow in the brain, and this is called the **BOLD** signal, which stands for blood oxygenation level dependent. As neurons fire, blood flow increases to that area. Therefore, blood flow in a particular region of the brain is a reasonable proxy for neural activity in that brain region.

Positron emission tomography, the **PET scan**, a more expensive and invasive procedure, also allows measurement of both brain structure and brain function, although the measurement of brain structure is not as precise as with MRI or fMRI. A substance used by the brain is labeled with a short-lived radioactive isotope and injected into the bloodstream. The radioactive molecules of the substance emit a particle called a positron, which quickly collides with an electron. A pair of high-energy light particles shoot out from the skull in opposite directions and are detected by the scanner. The computer analyzes millions of such recordings and converts them into a picture of the functioning brain. The images are in color; fuzzy spots of lighter and warmer colors are areas in which metabolic rates for the substance are higher. Because this is more invasive than fMRI, it is now used less often as a measure of brain function.

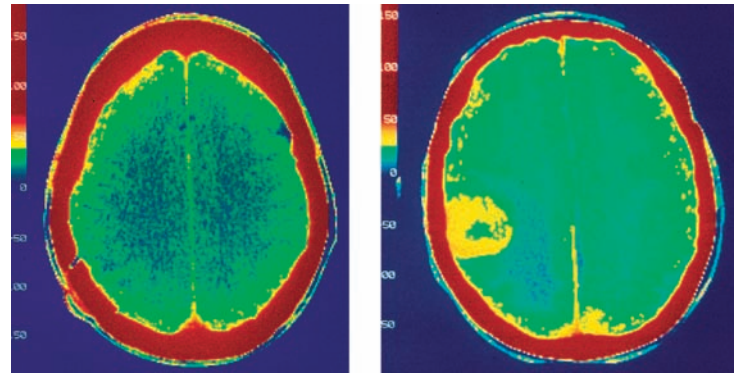
Visual images of the working brain can indicate sites of seizures, brain tumors, strokes, and trauma from head injuries, as well as the distribution of psychoactive drugs in the brain. fMRI and to a lesser extent PET are being used to study possible abnormal brain processes that are linked to various disorders, such as the failure of the prefrontal cortex of patients with schizophrenia to become activated while they attempt to perform a cognitive task. Current neuroimaging studies in psychopathology are attempting to identify not only areas of the brain that may be dysfunctional (e.g., the prefrontal cortex) but also deficits in the ways in which different areas of the brain communicate with one another. This type of inquiry is often referred to as functional connectivity analysis since it aims to identify how different areas of the brain are connected with one another.

## Neurotransmitter Assessment

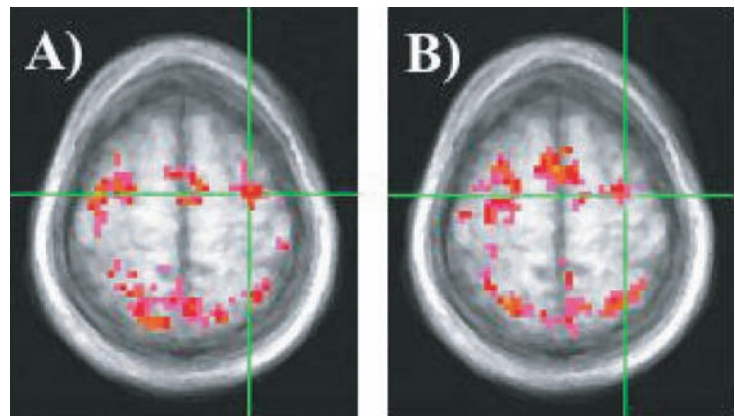
It might seem that assessing the amount of a particular neurotransmitter or the quantity of its receptors in the brain would be straightforward. But as we began to discuss in Chapter 2, it is not. Most of the research on neurotransmitters and psychopathology has relied on indirect assessments.

In postmortem studies, the brains of deceased patients are removed and the amount of specific neurotransmitters in particular brain areas can then be directly measured. Different brain areas can be infused with substances that bind to receptors, and the amount of binding can then be quantified; more binding indicates more receptors.

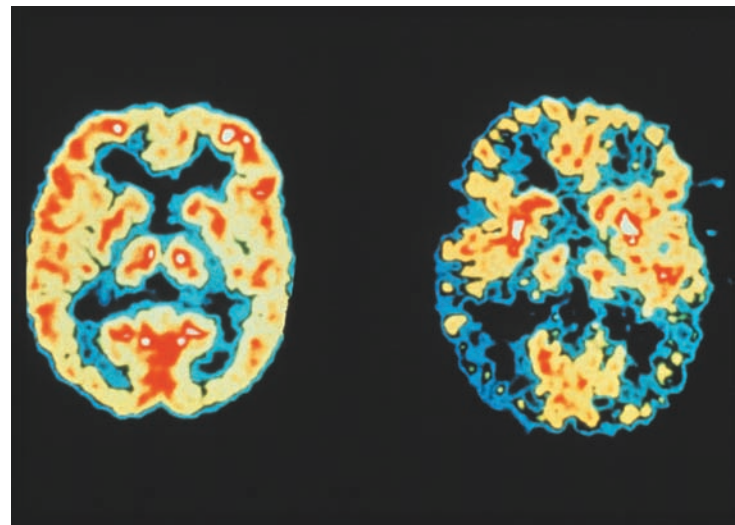
In studies of participants who are alive, one common method of neurotransmitter assessment involves analyzing the metabolites of neurotransmitters that have been broken down by enzymes. A **metabolite**, typically an acid, is produced when a neurotransmitter is deactivated. These by-products of the breakdown of



These two CT scans show a horizontal “slice” through the brain. The one on the left is normal; the one on the right has a tumor on the left side. (Dan McCoy/Rainbow.)



Functional magnetic resonance images (fMRI). With this method, researchers can measure how brain activity changes while a person is doing different tasks, such as viewing an emotional film, completing a memory test, looking at a visual puzzle, or hearing and learning a list of words. (Reprinted from J. E. McDowell et al., Neural correlates of refixation saccades and antisaccades in normal and schizophrenia subjects. *Biological Psychiatry*, 51, 216–223 2002 with permission from Elsevier.)



The PET scan on the left shows a normal brain; the one on the right shows the brain of a patient with Alzheimer’s disease. (Dr. Robert Friedland/Photo Researchers, Inc.)

neurotransmitters, such as norepinephrine, dopamine, and serotonin, are found in urine, blood serum, and cerebrospinal fluid (CSF; the fluid in the spinal column and in the brain's ventricles). For example, a major metabolite of dopamine is homovanillic acid; of serotonin, 5-hydroxyindoleacetic acid. A high level of a particular metabolite presumably indicates a high level of a neurotransmitter, and a low level indicates a low level of the transmitter.

But there is a problem with measuring metabolites from blood or urine: such measures are not direct reflections of levels of neurotransmitters in the brain; metabolites measured in this way could reflect neurotransmitters anywhere in the body. A more specific measure can be taken of metabolites in the CSF fluid drawn from a person's spinal cord. Even with CSF fluid, however, metabolites reflect activity throughout the brain and spinal cord, rather than regions that are directly involved in psychopathology. We will see in Chapter 5 that some people with depression have low CSF levels of the main metabolite of serotonin—a fact that has played an important role in the serotonin theory of depression.

Another problem with metabolite studies is that they are correlational. In Chapter 4, we discuss the limits of correlational research, including the fact that causation cannot be determined from a correlational study. That is, when researchers find that neurotransmitter levels are low among people with a particular disorder, such as depression, this could be because neurotransmitter levels cause depression, because depression causes neurotransmitter changes, or because a third variable causes shifts in both neurotransmitters and depression. For example, dopamine, norepinephrine, and serotonin levels change in response to stress. To test whether neurotransmitter levels could cause symptoms, experimental evidence is needed.

To provide more experimental data on whether these neurotransmitter systems actually help cause psychopathology, one strategy is to administer drugs that increase or decrease levels of neurotransmitters. For example, a drug that raises the level of serotonin should alleviate depression; one reducing it should trigger depressive symptoms. This strategy also has its

problems, though. One might wonder about whether it is ethical to do these studies if the goal of an experiment is to produce symptoms. On this front, it is reassuring that most studies find very temporary effects of these medications; neurotransmitter systems quickly return to normal levels, allowing for recovery from these brief mood episodes. Another issue is that drugs that change levels of one neurotransmitter often tend to influence other neurotransmitter systems. We will see examples of these types of studies throughout this book.

Clinicians and researchers in many disciplines are currently using brain imaging and neurotransmitter assessment techniques both to discover previously undetectable brain problems and to conduct inquiries into the neurobiological contributions to thought, emotion, and behavior. It is a very lively and exciting area of research and application. Indeed, one might reasonably assume that researchers and clinicians, with the help of such procedures and technological devices as fMRI, could observe the brain and its functions more or less directly and thus assess all brain abnormalities. Results to date, however, are not strong enough for these methods to be used in diagnosing psychopathology. Moreover, many brain abnormalities involve alterations in structure so subtle or slight in extent that they have thus far eluded direct examination. Furthermore, the problems in some disorders are so widespread that finding the contributing brain dysfunction is a daunting task. Take, for example, schizophrenia, which affects thinking, feeling, and behavior. Where in the brain might there be dysfunction? Looking for areas that influence thinking, feeling, and behavior requires looking at just about the entire brain.

### Neuropsychological Assessment

It is important at this point to note a distinction between neurologists and neuropsychologists, even though both specialists are concerned with the study of the central nervous system. A **neurologist** is a physician who specializes in diseases or problems that affect the nervous system, such as stroke, muscular dystrophy, cerebral palsy, or Alzheimer's disease. A **neuropsychologist** is a psychologist who



Measures of neurotransmitter metabolites in blood or urine levels do not provide a very accurate index of neurotransmitter levels in the brain. (Spencer Grant/Photo Researchers, Inc.)



studies how dysfunctions of the brain affect the way we think, feel, and behave. Both kinds of specialists contribute much to each other as they work in different ways, often collaboratively, to learn how the nervous system functions and how to ameliorate problems caused by disease or injury to the brain.

**Neuropsychological tests** are often used in conjunction with the brain imaging techniques just described, both to detect brain dysfunction and to help pinpoint specific areas of behavior that are impacted by problems in the brain. Neuropsychological tests are based on the idea that different psychological functions (e.g., motor speed, memory, language) rely on different areas of the brain. Thus, for example, neuropsychological testing might help identify the extent of brain damage suffered during a stroke, and it can provide clues about where in the brain the damage may exist that can then be confirmed with more expensive brain imaging techniques. There are numerous neuropsychological tests used in psychopathology assessment. Here, we highlight two widely used batteries of tests.

One neuropsychological test is Reitan's modification of a battery, or group, of tests previously developed by Halstead, called the Halstead–Reitan neuropsychological test battery. The following are three of the Halstead–Reitan tests.

- 1. Tactile Performance Test—Time.** While blindfolded, the patient tries to fit variously shaped blocks into spaces of a form board, first using the preferred hand, then the other, and finally both.
- 2. Tactile Performance Test—Memory.** After completing the timed test, the participant is asked to draw the form board from memory, showing the blocks in their proper location. Both this and the timed test are sensitive to damage in the right parietal lobe.
- 3. Speech Sounds Perception Test.** Participants listen to a series of nonsense words, each comprising two consonants with a long-*e* sound in the middle. They then select the “word” they heard from a set of alternatives. This test measures left-hemisphere function, especially temporal and parietal areas.

Extensive research has demonstrated that the battery is valid for detecting behavior changes linked to brain dysfunction resulting from a variety of conditions, such as tumors, stroke, and head injury (Horton, 2008).

The Luria–Nebraska battery (Golden, Hammeke, & Purisch, 1978), based on the work of the Russian psychologist Aleksandr Luria (1902–1977), is also widely used (Moses & Purisch, 1997). The battery includes 269 items divided into 11 sections designed to determine basic and complex motor skills, rhythm and pitch abilities, tactile and kinesthetic skills, verbal and spatial skills, receptive speech ability, expressive speech ability, writing, reading, arithmetic skills, memory, and intellectual processes. The pattern of scores on these sections, as well as on the 32 items found to be the most discriminating and indicative of overall impairment, helps reveal potential damage to the frontal, temporal, sensorimotor, or parietal-occipital area of the right or left hemisphere.

The Luria–Nebraska battery can be administered in 2½ hours and can be scored in a highly reliable manner (e.g., Kashden & Franzen, 1996). Criterion validity has been established by findings that test scores can correctly distinguish 86 percent of neurological patients and controls (Moses et al., 1992). A particular advantage of the Luria–Nebraska tests is that one can control for educational level so that a less educated person will not receive a lower score solely because of limited educational experience (Brickman et al., 1984). Finally, a version for children ages 8 to 12 (Golden, 1981a, 1981b) has been found useful in helping to pinpoint brain damage and in evaluating the educational strengths and weaknesses of children (Sweet et al., 1986).



Neuropsychological tests assess various performance deficits in the hope of detecting a specific area of brain malfunction. Shown here is the Tactile Performance Test. (Richard Nowitz/Photo Researchers, Inc.)

## Psychophysiological Assessment

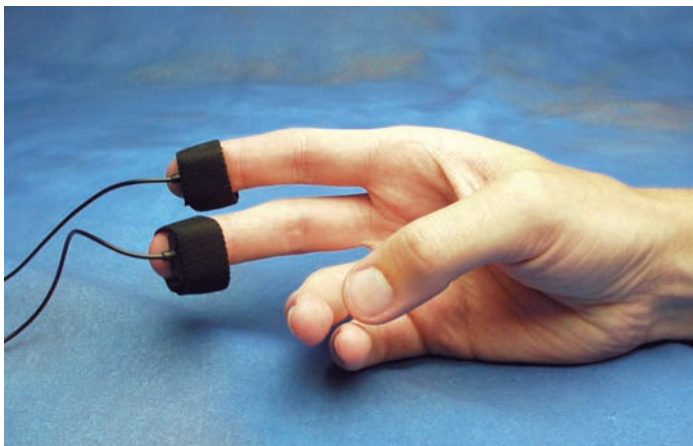
The discipline of **psychophysiology** is concerned with the bodily changes that are associated with psychological events. Experimenters have used measures such as heart rate, tension in the muscles, blood flow in various parts of the body, and electrical activity in the brain (so-called brain waves) to study physiological changes when people are afraid, depressed, asleep, imagining, solving problems, and so on. Like the brain-imaging methods we have already discussed, the assessments we describe here are not sensitive enough to be used for diagnosis. They can, however, provide important information about a person's reactivity and can also be used to compare individuals. For example, in using exposure to treat a patient with an anxiety disorder, it would be useful to know the extent to which the patient shows physiological reactivity when exposed to the stimuli that create anxiety. Patients who show more physiological reactivity may be experiencing more fear, which predicts more benefit from the therapy (Foa et al., 1995).

The activities of the autonomic nervous system (also discussed in Chapter 2) are often assessed by electrical and chemical measurements to understand aspects of emotion. One important measure is heart rate. Each heartbeat generates electrical changes, which can be recorded by electrodes placed on the chest that convey signals to an electrocardiograph or a polygraph. The signal is graphically depicted in an **electrocardiogram (EKG)**, which may be seen as waves on a computer screen or on a roll of graph paper.

A second measure of autonomic nervous system activity is **electrodermal responding**, or skin conductance. Anxiety, fear, anger, and other emotions increase activity in the sympathetic nervous system, which then boosts sweat-gland activity. Increased sweat-gland activity increases the electrical conductance of the skin. Conductance is typically measured by determining the current that flows through the skin as a small voltage is passed between two electrodes on the hand. When the sweat glands are activated, this current shows a pronounced increase. Since the sweat glands are activated by the sympathetic nervous system, increased sweat-gland activity indicates sympathetic autonomic excitation and is often taken as a measure of emotional arousal. These measures are widely used in research in psychopathology.

Brain activity can be measured by an **electroencephalogram (EEG)**. Electrodes placed on the scalp record electrical activity in the underlying brain area. Abnormal patterns of electrical activity can indicate seizure activity in the brain or help in locating brain lesions or tumors. EEG indices are also used to measure attention and alertness.

As with the brain-imaging techniques reviewed earlier, a more complete picture of a human being is obtained when physiological functioning is assessed while the person is engaging in some form of behavior or cognitive activity. If experimenters are interested in psychophysiological responding in patients with obsessive-compulsive disorder, for example, they would likely study the patients while presenting stimuli, such as dirt, that would elicit the problematic behaviors.



In psychophysiological assessment, physical changes in the body are measured. Skin conductance can be measured with sensors on two fingers. (Courtesy of BIOPAC Systems, Inc. (biopac.com).)

### A Cautionary Note about Neurobiological Assessment

A cautionary note regarding neurobiological assessment methods is in order here. Inasmuch as psychophysiology and brain imaging employ highly sophisticated electronic machinery, and many psychologists aspire to be as scientific as possible, researchers and clinicians sometimes believe uncritically in these apparently objective assessment devices without appreciating their real limitations and complications. Many of the measurements do not differentiate clearly among emotional states. Skin conductance, for example, increases not only with anxiety but also with other emotions—among them, happiness. In addition, being in a scanner is often a threatening experience. Thus, the investigator interested in measuring brain changes associated with emotion using fMRI must also take the scanning environment into account. It is also important to keep in mind that brain-imaging techniques do not allow us to manipulate brain activity and then measure a change in behavior (Feldman Barrett, 2003). In a typical study, we

show people a list of emotionally evocative words and then measure blood flow in the brain. Does a person who fails to show the same level of activation in emotion regions during this task have a brain-based emotion deficit? Not necessarily. The person might not have paid attention, might not have understood the words, or might be focused on the loud clanging noises that the fMRI machine is making. It is important to be extremely careful in considering alternative explanations for the effects found in these studies.

Neither is there a one-to-one relationship between a score on a given neuropsychological test or a finding on an fMRI scan on the one hand and psychological dysfunction on the other. The reasons for these sometimes loose relationships have to do with such factors as how the person has, over time, reacted to and coped with the losses brought about by the brain dysfunction. And the success of coping, in turn, has to do with the social environment in which the person has lived, for example, how understanding parents and associates have been or how well the school system has provided for the special educational needs of the person. Furthermore, the brain changes in response to these psychological and socioenvironmental factors over time. Therefore, in addition to the imperfect nature of the neurobiological assessment instruments themselves and our incomplete understanding of how the brain actually functions, clinicians and researchers must consider these environmental factors that operate over time to contribute to the clinical picture. In other words, a complete assessment must include multiple methods (clinical interviews, psychological and neurobiological methods).

A final caution is reflected in the simple yet often unappreciated fact that in attempting to understand the neurocognitive consequences of any brain dysfunction, one must understand the preexisting abilities that the patient had prior to diagnosis with a mental disorder. This straightforward truth brings to mind the story of the man who, recovering from an accident that has broken all the fingers in both hands, earnestly asks the surgeon whether he will be able to play the piano when his wounds heal. “Yes, I’m sure you will,” says the doctor reassuringly. “That’s wonderful,” exclaims the man, “I’ve always wanted to be able to play the piano.”

## Quick Summary

Advances in technology have allowed clinicians and researchers to “see” the living brain. Different imaging techniques, such as CT, MRI, and fMRI, have the potential to show areas of the brain that might not be working optimally. Direct assessment of neurotransmitters is not done often. Rather, examinations of the metabolites of neurotransmitters provide a rough way to estimate how neurotransmitters are functioning. Another approach is to administer drugs that increase or decrease the levels of a neurotransmitter. Postmortem exams also allow for measurements of neurotransmitters, particularly receptors.

Neuropsychological tests are tests that have been developed to show how changes in behavior may reflect damage or disturbance in particular areas of the brain. Psychophysiological assessment methods can show how behaviors and cognitions are linked to changes in nervous system activity, such as heart rate, skin conductance, or brain activity. These methods have as many or more limitations as other assessment measures, and the key concepts of reliability and validity are just as relevant with neurobiological assessment as with other forms of assessment.

## Check Your Knowledge 3.4

True or false?

1. MRI is a technique that shows both the structure and function of the brain.
2. Neurotransmitter assessment is most often done using indirect methods.
3. A neuropsychologist is a psychologist who studies how dysfunctions of the brain affect the way we think, feel, and behave.
4. Brain activity can be measured with the psychophysiological method called EKG.





## Cultural and Ethnic Diversity and Assessment

Studies of the influences of culture and ethnicity on psychopathology and its assessment have proliferated in recent years. As you read about some of this research, it is critical to keep in mind that there are typically more differences within cultural, ethnic, and racial groups than there are between them. Remembering this important point can help avoid the dangers of stereotyping members of a culture.

We should also note that the reliability and validity of various forms of psychological assessment have been questioned on the grounds that their content and scoring procedures reflect the culture of white European Americans and so may not accurately assess people from other cultures. In this section we discuss problems of cultural bias and what can be done about them.

### Cultural Bias in Assessment

The issue of cultural bias in assessment refers to the notion that a measure developed for one culture or ethnic group may not be equally reliable and valid with a different cultural or ethnic group. Some tests that were developed in the United States, however, have been translated into different languages and used in different cultures successfully. For example, a Spanish-language version of the WAIS has been available for over 40 years (Wechsler, 1968) and can be useful in assessing the intellectual functioning of people from Hispanic or Latino cultures (Gomez, Piedmont, & Fleming, 1992). Additionally, the MMPI-2 has been translated into more than two dozen languages (Tsai et al., 2001).

Simply translating words into a different language, however, does not ensure that the meaning of those words will be the same across different cultures. Several steps in the translation process, including working with multiple translators, back-translating, and testing with multiple native speakers, can help to ensure that the test is similar in different languages. This approach has been successful in achieving equivalence across different cultures and ethnic groups for some instruments, such as the MMPI-2 (Arbisi, Ben-Porath, & McNulty, 2002). Even with the MMPI-2, however, there are cultural differences that are not likely attributable to differences in psychopathology. For example, among Asian Americans who are not heavily assimilated into American culture, scores on most MMPI-2 scales are higher than those of Caucasians (Tsai & Pike, 2000). This is unlikely to reflect truly higher emotional disturbance among Asians. For children, the latest version of the WISC has not only been translated into Spanish (WISC-IV Spanish); it also has a complete set of norms for Spanish-speaking children in the United States, and the items have been designed explicitly to minimize cultural bias.

Despite these efforts, the field has a way to go in reducing cultural and ethnic bias in clinical assessment. These cultural assumptions or biases may cause clinicians to over- or underestimate psychological problems in members of other cultures (Lopez, 1989, 1996). African American children are overrepresented in special education classes, which may be a result of subtle biases in the tests used to determine such placement (Artiles & Trent, 1994). At least since the 1970s, studies have found that African Americans are more likely to receive a diagnosis of schizophrenia than are Caucasian Americans, but it is still unclear whether this reflects an actual difference or a form of ethnic bias on the part of clinicians (Arnold et al., 2004; Trierweiler et al., 2000). Yet take the example of an Asian American man who is very emotionally withdrawn. Should the clinician



Assessment must take the person's cultural background into account. Believing in possession by spirits is common in some cultures and thus should not always be taken to mean that the believer is psychotic. (Tony Savino/The Image Works.)

consider that lower emotional expressiveness in men is viewed more positively in Asian cultures than in European American culture? A clinician who quickly attributes the behavior to a cultural difference may overlook an emotional problem that he or she would be likely to diagnose if the patient were a white male.

How do such biases come about? Cultural factors may affect assessment in various ways. Language differences, differing religious and spiritual beliefs, the alienation or timidity of members of ethnic groups when being assessed by clinicians of the European American culture—all these factors can play a role. For example, clinicians who encounter clients claiming to be surrounded by spirits might view this belief as a sign of schizophrenia. Yet in Puerto Rican cultures, such a belief is common; therefore, believing that one is surrounded by spirits should probably not be taken as a sign of schizophrenia in a Puerto Rican person (Rogler & Hollingshead, 1985).

Cultural and ethnic differences in psychopathology must be examined more closely. Unfortunately, the cultural and ethnic biases that can creep into clinical assessment do not necessarily yield to efforts to compensate for them. There is no simple answer. The DSM-5's emphasis on cultural factors in the discussion of every category of disorder may well sensitize clinicians to the issue, a necessary first step. When practitioners were surveyed, they overwhelmingly reported taking culture into account in their clinical work (Lopez, 1994), so it appears that the problem, if not the solution, is clearly in focus.

### Strategies for Avoiding Cultural Bias in Assessment

Clinicians can—and do—use various methods to minimize the negative effects of cultural biases when assessing patients. Perhaps the place to begin is with graduate training programs. Lopez (2002) has noted three important issues that should be taught to graduate students in clinical psychology programs. First, students must learn about basic issues in assessment, such as reliability and validity. Second, students must become informed about the specific ways in which culture or ethnicity may impact assessment rather than relying on more global stereotypes about a particular cultural or ethnic group. Third, students must consider that culture or ethnicity may not impact assessment in every individual case.

Assessment procedures can also be modified to ensure that the person truly understands the requirements of the task. For example, suppose that a Native American child performed poorly on a test measuring psychomotor speed. The examiner's hunch is that the child did not understand the importance of working quickly and was overly concerned with accuracy instead. The test could be administered again after a more thorough explanation of the importance of working quickly without worrying about mistakes. If the child's performance improves, the examiner has gained an important understanding of the child's test-taking strategy and avoids diagnosing psychomotor speed deficits.

Finally, when the examiner and client have different ethnic backgrounds, the examiner may need to make an extra effort to establish a rapport that will result in the person's best performance. For example, when testing a shy Hispanic preschooler, one of the authors was unable to obtain a verbal response to test questions. However, the boy was overheard talking in an animated and articulate manner to his mother in the waiting room, leading to a judgment that the test results did not represent a valid assessment of the child's language skills. When testing was repeated in the child's home with his mother present, advanced verbal abilities were observed.

As Lopez (1994) points out, however, “the distance between cultural responsiveness and cultural stereotyping can be short” (p. 123). To minimize such problems, clinicians are encouraged to be particularly tentative about drawing conclusions regarding patients from different cultural and ethnic backgrounds. Rather, they are advised to make hypotheses about the influence of culture on a particular client, entertain alternative hypotheses, and then test those hypotheses.



Cultural differences can lead to different results on an aptitude or IQ test. For example, Native American children may lack interest in the individualistic, competitive nature of IQ tests because of the cooperative, group-oriented values instilled by their culture. (© Gabe Palmer/Alamy Limited.)

Training in cultural awareness is truly important, as a clinician's biases can influence diagnosis. As an example, schizophrenia is often overdiagnosed among African Americans, leading to high dosages of antipsychotic medications and too many hospitalizations (Alarcón et al., 2009). One way to combat these biases is to use structured diagnostic interviews, like the SCID described above. When clinicians use structured interviews, they are less likely to overdiagnose minority patients (Garb, 2005).

## Summary

- In gathering diagnosis and assessment information, clinicians and researchers must be concerned with both reliability and validity. Reliability refers to whether measurements are consistent and replicable; validity, to whether assessments are tapping into what they are meant to measure. Assessment procedures vary greatly in their reliability and validity. Certain diagnostic categories are more reliable than others.

## Diagnosis

- Diagnosis is the process of assessing whether a person meets criteria for a mental disorder. Having an agreed-on diagnostic system allows clinicians to communicate effectively with each other and facilitates the search for causes and treatments. Clinically, diagnosis provides the foundation for treatment planning.

- *The Diagnostic and Statistical Manual of Mental Disorders* (DSM), published by the American Psychiatric Association, is an official diagnostic system widely used by mental health professionals. The last edition of the manual, referred to as DSM-IV-TR, was published in 2000, and the publication of DSM-5 is expected in 2013.

- Reliability of diagnosis has been improved dramatically by including specific criteria for each diagnosis. Criticisms of the DSM include the proliferation of diagnoses that are often related to the same risk factors and tend to co-occur; the fact that reliability in practice may be lower than that achieved in research studies; and the ongoing need to validate diagnoses against etiology, course, and treatment. Most researchers and clinicians, though, recognize that the DSM is an enormous advance compared to historical systems.

- Some critics of the DSM argue against diagnosis in general. They point out that diagnostic classifications may ignore important information. Although many worry that diagnostic labels will increase stigma, there is some data that a diagnosis can reduce stigma by providing an explanation for worrisome behavior.

## Assessment

- Clinicians rely on several modes of psychological and neurobiological assessment in trying to find out how best to describe an individual, search for the reasons the person is troubled, arrive at an accurate diagnosis, and

design effective treatments. The best assessment involves multiple types of methods.

- Psychological assessments include clinical interviews, assessments of stress, psychological tests, and behavioral and cognitive assessments.

- Clinical interviews are structured or relatively unstructured conversations in which the clinician probes the patient for information about his or her problems. Assessing stress is key to the field of psychopathology. A number of useful methods for assessing stress have been developed, including the LEDS.

- Psychological tests are standardized procedures designed to assess personality or measure performance. Personality assessments range from empirically derived self-report questionnaires, such as the Minnesota Multiphasic Personality Inventory, to projective tests in which the patient interprets ambiguous stimuli, such as the Rorschach test. Intelligence tests, such as the Wechsler Adult Intelligence Scale, evaluate a person's intellectual ability and predict how well he or she will perform academically.

- Behavioral and cognitive assessment is concerned with how people act, feel, and think in particular situations. Approaches include direct observation of behavior, interviews, and self-report measures that are situational in their focus.

- Neurobiological assessments include brain-imaging techniques, such as fMRI, that enable clinicians and researchers to see various structures and access functions of the living brain; neurochemical assays that allow clinicians to make inferences about levels of neurotransmitters; neuropsychological tests, such as the Luria-Nebraska battery, that seek to identify brain defects based on variations in responses to psychological tests; and psychophysiological measurements, such as heart rate and electrodermal responding, that are associated with certain psychological events or characteristics.

- Cultural and ethnic factors play a role in clinical assessment. Assessment techniques developed on the basis of research with Caucasian populations may be inaccurate when used with clients of differing ethnic or cultural backgrounds, for example. Clinicians can have biases when evaluating ethnic minority patients, which can lead to minimizing or exaggerating a patient's psychopathology. Clinicians use various methods to guard against the negative effects of cultural biases in assessment.

## Answers to Check Your Knowledge Questions

**3.1** 1. b; 2. b, c, d, a

**3.2** 1. high comorbidity, many different diagnoses are related to the same causes, symptoms of many different diagnoses respond to the same treatments; 2. any three of the following: etiology, course, social functioning, treatment

**3.3** 1. F; 2. T; 3. T; 4. T; 5. T; 6. F

**3.4** 1. F; 2. T; 3. T; 4. F

## Key Terms

alternate-form reliability	dimensional diagnostic system	Minnesota Multiphasic	reactivity
behavioral assessment	ecological momentary	Personality Inventory	reliability
BOLD	assessment (EMA)	(MMPI)	Rorschach Inkblot Test
categorical classification	electrocardiogram (EKG)	multiaxial classification system	self-monitoring
clinical interview	electrodermal responding	neurologist	standardization
comorbidity	electroencephalogram (EEG)	neuropsychological tests	stress
concurrent validity	functional magnetic resonance	neuropsychologist	structured interview
construct validity	imaging (fMRI)	personality inventory	test–retest reliability
content validity	intelligence test	PET scan	Thematic Apperception Test
criterion validity	internal consistency reliability	predictive validity	(TAT)
CT or CAT scan	interrater reliability	projective hypothesis	validity
<i>Diagnostic and Statistical Manual</i>	magnetic resonance imaging	projective test	
<i>of Mental Disorders</i>	(MRI)	psychological tests	
diagnosis	metabolite	psychophysiology	

