body of literature has focused on the CPI. Each new piece of literature extends the utility of the test and adds to its construct validity. Therefore, the future of the CPI as a measure of normal personalities has good potential despite its limitations.

The Factor Analytic Strategy

Structured personality tests, as they exist today, share one common set of assumptions. These assumptions, simply stated, are that humans possess characteristics or traits that are stable, vary from individual to individual, and can be measured. Nowhere are these assumptions better illustrated than in the factor analytic strategy of test construction.

Recall that factor analysis is a statistical procedure for reducing the redundancy in a set of intercorrelated scores. For example, one major technique of factor analysis, the principal-components method (Bentler & de Leeuw, 2011), finds the minimum number of common factors that can account for an interrelated set of scores. As noted in the previous section, no more than two factors, perhaps only one, can account for most of the variance in both the CPI and the MMPI, which suggests that these tests are actually measuring no more than two unique components and that all scales are related to these one or two components.

The advantages of factor analysis are quite evident. However, before computers, even simple factor analyses required several weeks or even months of tedious arithmetic operations on a hand calculator. Therefore, the development of the factor analytic strategy awaited computer technology. R. B. Cattell has particularly distinguished himself in using the factor analytic strategy of structured personality assessment; this section focuses largely on his work.

Guilford's Pioneer Efforts

One usual strategy in validating a new test is to correlate the scores on the new test with the scores on other tests that purport to measure the same entity. J. R. Guilford's approach was related to this procedure. However, instead of comparing one test at a time to a series of other tests, Guilford and his associates determined the interrelationship (intercorrelation) of a wide variety of tests and then factor analyzed the results in an effort to find the main dimensions underlying all personality tests. If the results from existing personality tests could be reduced to a few factors, then items that correlated highly with these factors could be used in a new test, which would therefore capture the major dimensions of personality.

The result of the initial attempt to apply this strategy was a series of inventories that Guilford and his associates published in the 1940s and which were ultimately collapsed into a single scale—the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1956).

This survey reduces personality to 10 dimensions, each of which is measured by 30 different items. The 10 dimensions are general activity, restraint, ascendance (leadership), sociability, emotional stability, objectivity, friendliness, thoughtfulness, personal relations, and masculinity. The test presents a list of statements, most of which are self-statements as in the MMPI and MMPI-2.

The subject must indicate "Yes" or "No" for each statement. Three verification keys are included to detect falsification and to evaluate the validity of the profile. However, this first major factor analytic structured personality test failed to catch on, perhaps because it was overshadowed by the MMPI and because of its arbitrary, subjective way of naming factors. Today, the Guilford-Zimmerman Temperament Survey primarily serves as a research tool, with strong indications based on the latest and most powerful research tools that the 10 factors have an overarching, higher-order single factor, social efficacy, as discussed earlier in this chapter (Rushton & Irwing, 2009).

Cattell's Contribution

Rather than attempting to uncover the major dimensions of personality by intercorrelating personality tests, R. B. Cattell began with all the adjectives applicable to human beings so that he could empirically determine and measure the essence of personality. Beginning with a monumental catalog of all the adjectives (trait names) in an unabridged dictionary that apply to humans, Allport and Odbert (1936) reduced their list to 4504 "real" traits. Adding to the list traits found in the psychological and psychiatric literature, Cattell then reduced the list to 171 items that he believed accounted for the meaning of all items on the original list. College students then rated their friends on these 171 terms, and the results were intercorrelated and factor analyzed. The 171 terms were reduced to 36 dimensions, called *surface traits.* Subsequent investigation by factor analysis finally produced 16 distinct factors that accounted for all the variables. Thus, Cattell had reduced personality to 16 basic dimensions, which he called *source traits* (see Table 13.4).

The product of Cattell's marathon task was the Sixteen Personality Factor Questionnaire, better known as the 16PF (Cattell & Cattell, 1995; Schuerger, 1995), which was subsequently revised following continued factor analysis. Consistent with the factor analytic strategy, items that correlated highly with each of the 16 major factors, or source traits, were included, and those with relatively low correlations were excluded.

Developers took great care in standardizing the 16PF. Separate norms were provided for men alone, women alone, and men and women combined for each of three U.S. groups: adults, college students, and high-school seniors. Thus, nine sets of norms are available. To deal further with the covariation of structured personality test data and demographic variables that plagues the MMPI, the 16PF provides age corrections for those scales that change significantly with age. Six forms of the test are available: two parallel forms for each of the three levels of vocabulary proficiency, ranging from newspaper-literate adults through the educationally disadvantaged. For the latter, a tape-recorded (oral) form is also available. Norms for the various forms are based on more than 15,000 subjects representative of geographic area, population density, age, family income, and race according to figures provided by the U.S. census. Unlike the MMPI and CPI, the 16PF contains no item overlap, and keying is balanced among the various alternative responses.

Short-term test–retest correlation coefficients for the 16 source traits are impressive, with a range of .65 to .93 and a median coefficient of .83. Long-term test–retest coefficients, however, are not so impressive (.21 to .64), and

TABLE 13.4 The Primary Source Traits Covered by the 16PF Test

Factor	Low Sten score description (1–3)	High Sten score description (8–10)
A	*Cool,* **reserved, impersonal, detached, formal, aloof** Sizothymia*	*Warm,* **outgoing, kindly, easygoing, participating, likes people** Affectothymia
B	*Concrete-thinking,* **less intelligent** Lower scholastic mental capacity	*Abstract-thinking,* **more intelligent, bright** Higher scholastic mental capacity
C	*Affected by feelings,* **emotionally less stable, easily annoyed** Lower ego strength	*Emotionally stable,* **mature, faces reality, calm** Higher ego strength
E	*Submissive,* **humble, mild, easily led, accommodating** Submissiveness	*Dominant,* **assertive, aggressive, stubborn, competitive, bossy** Dominance
F	*Sober,* **restrained, prudent, taciturn, serious** Desurgency	*Enthusiastic,* **spontaneous, heedless, expressive, cheerful** Surgency
G	*Expedient,* **disregards rules, self-indulgent** Weaker superego strength	*Conscientious,* **conforming, moralistic, staid, rule bound** Stronger superego strength
H	*Shy,* **threat-sensitive, timid, hesitant, intimidated** Threctia	*Bold,* **venturesome, uninhibited, can take stress** Parmia
I	*Tough-minded,* **self-reliant, no-nonsense, rough, realistic** Harria	*Tender-minded,* **sensitive, overprotected, intuitive, refined** Premsia
L	*Trusting,* **accepting conditions, easy to get on with** Alaxia	*Suspicious,* **hard to fool, distrustful, skeptical** Protension
M	*Practical,* **concerned with "down-to-earth" issues, steady** Praxernia	*Imaginative,* **absent-minded, absorbed in thought, impractical** Autia
N	*Forthright,* **unpretentious, open, genuine, artless** Artlessness	*Shrewd,* **polished, socially aware, diplomatic, calculating** Shrewdness
O	*Self-assured,* **secure, feels free of guilt, untroubled, self-satisfied** Untroubled adequacy	*Apprehensive,* **self-blaming, guilt prone, insecure, worrying** Guilt proneness
Q1	*Conservative,* **respecting traditional ideas** Conservatism of temperament	*Experimenting,* **liberal, critical, open to change** Radicalism
Q2	*Group-oriented,* **a "joiner" and sound follower, listens to others** Group adherence	*Self-sufficient,* **resourceful, prefers own decisions** Self-sufficiency
Q3	*Undisciplined self-conflict,* **lax, careless of social rules** Low integration	*Following self-image,* **socially precise, compulsive** High self-concept control
Q4	*Relaxed,* **tranquil, composed, has low drive, unfrustrated** Low ergic tension	*Tense,* **frustrated, overwrought, has high drive** High ergic tension

*Titles in roman type are the technical names for the factors and are explained more fully in the *Handbook.*

most such coefficients reported in the literature are lower than those reported for the MMPI and MMPI-2 (Schuerger, Tait, & Tavernelli, 1982). Also a bit disappointing are the correlations between the various forms, which range from a low of .16 to a high of .79, with median coefficients in the .50's and .60's, depending on which forms are correlated. Moreover, despite the method used for deriving the factors, the 16 source traits of the 16PF do intercorrelate, with some correlations as high as .75. To deal with this overlap, the 16 factors themselves were factor analyzed, resulting in four second-order factors, for which one can obtain scores.

Other important features of the test are its provision of a parallel inventory for ages 12 to 18, the Junior Senior High School Personality Questionnaire, and still another parallel extension for use with ages 8 to 12, the Children's Personality Questionnaire. Cross-cultural studies have been conducted in Western Europe, Eastern Europe, the Middle East, Australia, Canada (Schuerger, 1995), and Korea (Sohn, 2002). To extend the test to the assessment of clinical populations, items related to psychological disorders have been factor analyzed, resulting in 12 new factors in addition to the 16 needed to measure normal personalities. These new factors were then used to construct a clinical instrument, the Clinical Analysis Questionnaire (CAQ) (Delhees & Cattell, 1971).

Despite the care that has gone into the 16PF, its research base and use pale when compared with those of the MMPI and MMPI-2. The fact is, neither clinicians nor researchers have found the 16PF to be as useful as the MMPI. Moreover, the claims of the 16PF to have identified the basic source traits of the personality are simply not true. However, various research investigations have supported the validity of Cattell's personality test (see Meyer, 1993).

Factor analysis is one of many ways of constructing tests. It will identify only those traits about which questions are asked, however, and it has no more claim to uniqueness than any other method. Even so, the 16PF remains an exemplary illustration of the factor analytic approach to structured personality testing.

Problems With the Factor Analytic Strategy

One major criticism of factor analytic approaches centers on the subjective nature of naming factors. To understand this problem, one must understand that each score on any given set of tests or variables can be broken down into three components: common variance, unique variance, and error variance. *Common variance* is the amount of variance a particular variable holds in common with other variables. It results from the overlap of what two or more variables are measuring. *Unique variance* refers to factors uniquely measured by the variable. In other words, it refers to some construct measured *only* by the variable in question. *Error variance* is variance attributable to error.

Factor analytic procedures generally identify sources of common variance at the expense of unique variance. Thus, important factors may be overlooked when the data are categorized solely on the basis of blind groupings by computers. Furthermore, all the computer can do is identify the groupings. The factor analyst must determine which factors these groupings measure, but no definite

criteria or rules exist for naming factors. If five items such as *daring*, *outgoing*, *determined*, *excitable*, and *fearless* load high on a factor, then what should one call this factor? In factor analysis, one name for this factor has about as much validity as any other.

round and contains no
of the stimulus does not conform
perceptions in general may not be accurate. You
are unwilling to provide the obvious, conventional response.
you feel out of place, like a square peg in a round hole.

Of course, examiners can never draw absolute, definite conclusions from any single response to an ambiguous stimulus. They can only hypothesize what a test response means. Even the same response to the same stimulus may have several possible meanings, depending on the characteristics of the people who make the response. A problem with all projective tests is that many factors can influence one's response to them. For example, a response may reflect a recent experience or an early experience one has forgotten. It may reflect something one has witnessed (a bloody murder) or something one imagines (flunking out of college) rather than something one has actually experienced directly. It may reflect day-to-day problems, such as an argument with a boyfriend or girlfriend. With all of these possible factors influencing a response, it is no wonder that the validity of projective tests has been questioned. Arguably, the interpretation of projective tests requires highly trained, experienced practitioners, but as in any field, even an expert can draw the wrong conclusions. Further, even the most experienced experts often disagree among themselves (Exner, 1995; Nezworski & Wood, 1995). As in the example at the beginning of the chapter, Rorschach users believe that they can use projective tests to draw valid conclusions. Many researchers, however, remain firmly unconvinced (Erickson, Lilienfeld, & Vitacco, 2007; Sechrest, Stickle, & Stewart, 1998; Wood et al., 1996, 2003).

The Rorschach Inkblot Test

As an example of a psychological test based on the projective hypothesis, the Rorschach has few peers. Indeed, no general discussion of psychological tests is complete without reference to the Rorschach, despite heated scientific controversies. The Rorschach has been called everything from a psychological X-ray (Piotrowski, 1980) and "perhaps the most powerful psychometric instrument ever envisioned" (Board of Professional Affairs, 1998, p. 392) to an instrument that "bears a charming resemblance to a party game" (Wood et al., 2003, p. 1) and should be "banned in clinical and forensic settings" (Garb, 1999, p. 316). Strangely, the Rorschach is both revered and reviled (compare, e.g., Erickson et al., 2007 and Mattlar, 2004).

Historical Antecedents

Like most concepts, the notion of using inkblots to study human functioning did not simply appear out of thin air. More than 25 years before the birth of Herman Rorschach, the originator of the test that bears his name, J. Kerner (1857) noted that individuals frequently report idiosyncratic or unique personal meanings when viewing inkblot stimuli. The wide variety of possible responses to inkblots does provide a rationale for using them to study individuals. Indeed, Binet proposed the idea of using inkblots to assess personality functioning (Binet & Henri, 1896)

when Rorschach was only 10 years old. Several historic investigators then supported Binet's position concerning the potential value of inkblots for investigating human personality (Dearborn, 1897; Kirkpatrick, 1900). Their support led to the publication of the first set of standardized inkblots by Whipple (1910). Rorschach, however, receives credit for finding an original and important use for inkblots: identifying psychological disorders. His investigation of inkblots began in 1911 and culminated in 1921 with the publication of his famous book *Psychodiagnostik.* A year later, he suddenly and unexpectedly died of a serious illness at age 37.

Rorschach's work was viewed with suspicion and even disdain right from the outset. Not even the sole psychiatric journal of Switzerland, Rorschach's homeland, reviewed *Psychodiagnostik* (Allison, Blatt, & Zimet, 1968). In fact, only a few foreign reviews of the book appeared, and these tended to be critical. When David Levy first brought Rorschach's test to the United States from Europe, he found a cold, unenthusiastic response. U.S. psychologists judged the test to be scientifically unsound, and psychiatrists found little use for it. Nevertheless, the use of the test gradually increased, and eventually it became quite popular.

Five individuals have played dominant roles in the use and investigation of the Rorschach. One of these, Samuel J. Beck, was a student of Levy's. Beck was especially interested in studying certain patterns or, as he called them, "configurational tendencies" in Rorschach responses (Beck, 1933). Beck, who died in 1980, eventually wrote several books on the Rorschach and influenced generations of Rorschach practitioners (Beck, 1944, 1945, 1952). Like Beck, Marguerite Hertz stimulated considerable research on the Rorschach during the years when the test first established its foothold in the United States (Hertz, 1937, 1938). Bruno Klopfer, who immigrated to the United States from Germany, published several key Rorschach books and articles and played an important role in the early development of the test (Klopfer & Davidson, 1944; Klopfer & Kelley, 1942). Zygmunt Piotrowski (1947, 1964) and David Rapaport (Rapaport, Gill, & Schafer, 1945–1946) came somewhat later than Beck Hertz, and Klopfer, but like them continues to exert an influence on clinical practitioners who use the Rorschach. The development of the Rorschach can be attributed primarily to the efforts of these five individuals. Like most experts, however, the five often disagreed. Their disagreements are the source of many of the current problems with the Rorschach. Each expert developed a unique system of administration, scoring, and interpretation; they all found disciples who were willing to accept their biases and use their systems.

Stimuli, Administration, and Interpretation

Rorschach constructed each stimulus card by dropping ink onto a piece of paper and folding it. The result was a unique, bilaterally symmetrical form on a white background. After experimenting with thousands of such blots, Rorschach selected 20. However, the test publisher would only pay for 10. Of the 10 finally selected, five were black and gray; two contained black, gray, and red; and three contained pastel colors of various shades. An example of a Rorschach card is shown in Figure 14.1.

The Rorschach is an individual test. In the administration procedure, each of the 10 cards is presented to the subject with minimum structure. After preliminary remarks concerning the purpose of testing, the examiner hands the first card to the

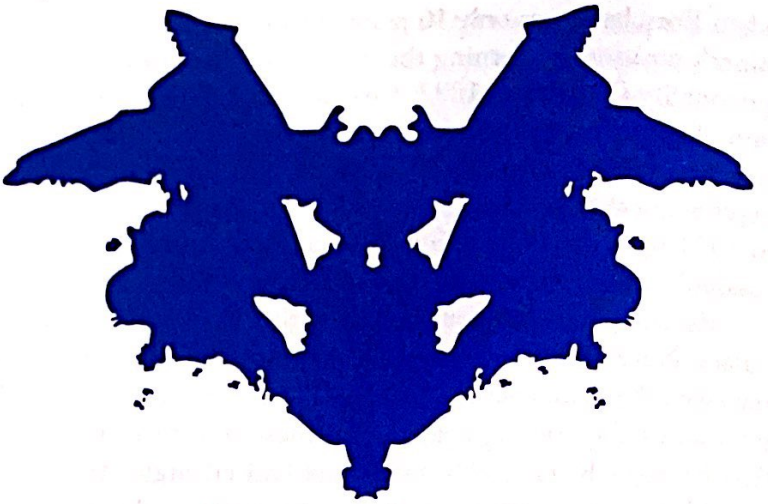

FIGURE 14.1
A Rorschach-type image is created by dropping ink onto a piece of paper and folding it. This is a reproduction of an actual card from the Rorschach.

subject and asks something like, "What might this be?" No restriction is placed on the type of response permitted, and no clues are given concerning what is expected. If the subject asks for guidance or clarification, the examiner gives little information. If, for example, the subject asks, "Do I use the whole thing or just part of it?" the examiner replies, "As you like" or "Whatever you choose." Anxious subjects or individuals who are made uncomfortable by unstructured situations frequently ask questions, attempting to find out as much as possible before committing themselves. The examiner, however, must not give any cues that might reveal the nature of the expected response. Furthermore, in view of the finding that the examiner may inadvertently reveal information or reinforce certain types of responses through facial expressions and other forms of nonverbal communication (Lord, 1950; Wood, Lilienfeld, Garb, & Nezworski, 2000a), Exner (1993) advocated an administration procedure in which the examiner sits next to the subject rather than face-to-face as in Rapaport's system (Blais, Norman, Quintar, & Herzog, 1995).

Notice that the examiner is nonspecific and largely vague. This lack of clear structure or direction with regard to demands and expectations is a primary feature of all projective tests. The idea is to provide as much ambiguity as possible so that the subject's response reflects only the subject. If the examiner inadvertently provides too many guidelines, the response may simply reflect the subject's tendency to perform as expected or to provide a socially desirable response, as discussed in Chapter 13. Therefore, an administration that provides too much structure is antithetical to the main idea behind projective tests.

Each card is administered twice. During *the free-association* phase of the test, the examiner presents the cards one at a time. If the subject gives only one response to the first card, then the examiner may say, "Some people see more than one thing here." The examiner usually makes this remark only once. If the subject rejects the card—that is, states that he or she sees nothing—then the examiner may reply, "Most people do see something here, just take your time." The examiner records every word and even every sound made by the subject verbatim. In addition, the examiner records how long it takes a subject to respond to a card (reaction time) and the position of the card when the response is made (upside down, sideways).

In the second phase, the *inquiry*, the examiner shows the cards again and scores the subject's responses. Responses are scored according to at least five dimensions, including location (where the perception was seen), determinant (what determined the response), form quality (to what extent the response matched the stimulus properties of the inkblot), content (what the perception was), and frequency of occurrence (to what extent the response was popular or original; popular responses occur once in every three protocols on average). A complete discussion of these special scoring categories is beyond the scope of this text. For more information on scoring and interpretation, see Exner's (1993; Exner & Erdberg, 2005) Rorschach textbooks.

In scoring for location, the examiner must determine where the subject's perception is located on the inkblot. To facilitate determining this location, a small picture of each card, known as the *location chart*, is provided. If necessary, on rare occasions, an examiner may give a subject a pencil and ask the subject to outline the perception on the location chart. In scoring for location, the examiner notes whether the subject used the whole blot (W), a common detail (D), or an unusual detail (Dd). Location may be scored for other factors as well, such as the *confabulatory response* (DW). In this response, the subject overgeneralizes from a part to the whole.

According to such Rorschach proponents as Exner, a summary of a subject's location choices can be extremely valuable. The examiner may, for example, determine the number and percentage of W, D, and Dd responses. This type of information, in which scoring categories are summarized as a frequency or percentage, is known as the quantitative, structural, or statistical aspect of the Rorschach as opposed to the qualitative aspects, which pertain to the content and sequence of responses. Normal subjects typically produce a balance of W, D, and Dd responses. When a subject's pattern deviates from the typical balance, the examiner begins to suspect problems (Tarafder, Mukhopadhyay, & Basu, 2004). However, no one has been able to demonstrate that a particular deviation is linked to a specific problem (e.g., see Acklin, 1995; Bartell & Solanto, 1995; Frank, 1995). A substantial deviation from what is typical or average may suggest several possibilities. The protocol may be invalid. The subject may be original or unconventional and thus fail to respond according to the typical pattern. Or the subject may have a perceptual problem associated with certain types of brain damage or severe emotional problems. The relative proportion of W, D, and Dd location choices varies with maturational development. Ames, Metraux, and Walker (1971), for example, noted that W responses occur most frequently in the 3- to 4-year-old group. As the child grows older, the frequency of W responses gradually decreases until young adulthood. Theoretically, adult protocols with a preponderance of W responses suggest immaturity or low mental age.

Like other quantitative aspects of the Rorschach, location patterns and frequencies have been studied in experimental investigations. Presumably, these investigations provide information about the meaning of various response patterns and thus contribute to the construct validity of the Rorschach. Unfortunately, many of the results of the studies conflict with the opinions of experts. Furthermore, many studies that support the validity of the Rorschach have been denounced as unreplicated, methodologically unsound, and inconsistent (Gacono & Evans, 2007; Wood et al., 2003). There appears to be no room for compromise in this area of psychological testing.

Having ascertained the location of a response, the examiner must then determine what it was about the inkblot that led the subject to see that particular percept. This factor is known as the *determinant*. One or more of at least four properties of an inkblot may determine or lead to a response: its form or shape, its perceived movement, its color, and its shading. If the subject uses only the form of the blot to determine a response, then the response is scored F and is called a *pure form response*. Responses are scored for form when the subject justifies or elaborates a response by statements such as "It looks like one," "It is shaped like one," or "Here are the head, legs, feet, ears, and wings." In all of these examples, the response is determined exclusively on the basis of shape. In addition to form, a perception may be based on movement, color, shading, or some combination of these factors. These other determinants can be further subdivided. Movement may be human (M), such as two people hugging; animal (FM), such as two elephants playing; or inanimate (m), such as sparks flying. As you can see, the scoring can become quite complex.

As with location, several attempts have been made to link the presence (or absence) of each determinant as well as the relative proportion of the various determinants to various hypotheses and empirical findings (Daini & Bernardini, 2007; Exner, 1999; Perry, Sprock, Schaible, & McDougall, 1995). Consider the movement response. Most Rorschach practitioners agree that whether and how a subject uses movement can be revealing. Like most Rorschach indicators, however, the meaning of movement is unclear because of disagreements among experts and contradictory or unclear experimental findings. Many experts believe that the movement response is related to motor activity and impulses. Numerous movement responses, for example, may suggest high motor activity or strong impulses. The ratio of M (human movement) to FM (animal movement) responses has been linked by some experts to a person's control and expression of internal impulses.

A special type of movement response is called *cooperative movement*. Such responses involve positive interaction between two or more humans or animals (Exner, 1999). Exner and colleagues believe that such responses provide information about a subject's attitude concerning how people interact. One study, for example, reported that individuals who give more than two such responses tended to be rated by others as fun to be with, easy to be around, and trustworthy (Exner & Farber, 1983). The conclusion seemed to be that such responses were positive. Subsequent research, however, could not confirm the initial findings (Shaffer & Erdberg, 1996). In a study of 20 individuals who had committed sexual homicide, 14 gave cooperative movement responses. Clearly, there is no simple or clear-cut approach to Rorschach interpretation (Gacono & Evans, 2007; Gacano & Meloy, 1994).

As you think about the inferences that can be drawn from the Rorschach, keep in mind that they are at best hypotheses. Competent psychologists never blindly accept one interpretation of a particular quantitative aspect of the Rorschach. Certainly, one who blindly accepts a particular interpretation of a Rorschach pattern is ignoring the available literature. Focused Example 14.1 explains some of the ways that highly trained experts use the Rorschach to make clinically useful inferences.

Identifying the determinant is the most difficult aspect of Rorschach administration. Because of the difficulties of conducting an adequate inquiry and the current lack of standardized administration procedures, examiners vary widely in the conduct of their inquiries (Blais et al., 1995). It has been known for years

14.1 FOCUSED EXAMPLE

Expert Interpretation of the Rorschach

Rorschach experts resolutely maintain that, if properly used, the Rorschach can be an invaluable tool. Modern scientists are completely unconvinced. At best, Rorschach interpretations should be viewed only as tentative hypotheses. Hypotheses that are confirmed by other sources of data usually have more validity than do those that cannot be confirmed. When the Rorschach is rigidly or blindly interpreted, scientific skepticism is justified. When the Rorschach is interpreted cautiously and in conjunction with other sources of data, however, a highly trained expert may surprise even the most critical scientist.

When Dennis Saccuzzo had a predoctoral internship at a Veterans Administration hospital, Marguerite Hertz, one of the five original Rorschach experts, was a consultant there. Every second Thursday of the month, Hertz would interpret an actual Rorschach protocol presented by interns or staff members. Her interpretations were so detailed and exact that Saccuzzo, who was inexperienced with the Rorschach, doubted their validity. When other interns or staff agreed with everything Hertz said, he became even more skeptical. He thought they were merely awed by Hertz's reputation and were afraid to challenge this spirited woman.

When Saccuzzo's turn came to present a Rorschach, he used the protocol of a patient he had been seeing in psychotherapy for several months. He knew this patient well and fully expected Hertz to make errors in her interpretation. He was surprised, however, when Hertz was able to describe this patient after reading only the first four or five responses and examining the quantitative summary of the various scoring categories and ratios. Within 25 minutes, Hertz told him not only what he already knew but also things he had not seen but were obviously true once pointed out. This experience was most unsettling. Having started with a strong bias against the Rorschach, and still doubting its scientific underpinnings, he could not dismiss what Hertz had done.

Later, he came to believe that Hertz's secret was her experience. She had given or studied so many Rorschachs that she had great insight into the meaning of each pattern. After having seen the Rorschach patterns of dozens, if not hundreds, of disturbed individuals, she could identify a problem. Indeed, her knowledge and experience were so broad that she could even distinguish specific types of disturbances based on the Rorschach.

However, until the experts can specify the exact processes underlying correct interpretations from the Rorschach, the criticism from scientists will continue, as Hertz herself (1986), who has repeatedly called for innovation and rigorous research, has acknowledged.

that examiner differences influence the subject's response (Gibby, Miller, & Walker, 1953; Hartman, 2001; Lis, Parolin, Calvo, Zennaro, & Meyer, 2007). As a result of this problem, much of the Rorschach literature is confounded by differences in administration and scoring alone, let alone interpretation. This is one reason why reliable experimental investigations of the Rorschach are rare (Lewandowski & Saccuzzo, 1976).

On the other hand, scoring content is relatively simple. Most authorities list content categories such as human (H), animal (A), and nature (N). An inquiry is generally not necessary to determine content.

Similarly, most experts generally agree on the so-called populars, those responses frequently given for each card. Exner's (1993, 2003) Comprehensive System, which includes as populars only those responses that occur once in three protocols on the average, provides a standardized method for scoring populars.

Form quality is the extent to which the percept (what the subject says the inkblot is) matches the stimulus properties of the inkblot. Scoring form quality is difficult. Some experts argue that if the examiner can also see the percept, then the response has adequate form quality, but if the examiner cannot see it, then the response has poor form quality and is scored F–. Obviously, such a subjective system is grossly inadequate because scoring depends on the intelligence, imagination, skill, and psychological state of the examiner. Exner's (1993) Comprehensive System, which uses the usual frequency of the occurrence of various responses in evaluating form quality, is more objective and thus more scientifically acceptable than the subjective method.

Table 14.1 summarizes our discussion of Rorschach scoring. Though the discussion has been incomplete, we hope it has shown how a projective test can be scored to yield quantitative data. These quantitative data, in turn, permit the accumulation of norms for particular groups. If subjects deviate from the typical or expected performance, then the examiner must determine the reason underlying the deviation. Proponents argue that this process can lead to valuable information about individuals (Acklin, 1995; Groth-Marnat, 1999; Hilsenroth, Fowler, & Padawer, 1998).

Rorschach scoring is obviously difficult and complex. Use of the Rorschach requires advanced graduate training. You should not attempt to score or use a Rorschach without formal and didactic graduate instruction and supervised experience. Without this detailed training, you might make serious errors because the procedure is so complex.

Rorschach protocols may be evaluated not only for its quantitative data but also for qualitative features, including specific content (Moreland, Reznikoff, & Aronow, 1995) and sequence of responses (Exner, 1999). One important aspect of a qualitative interpretation is an evaluation of content reported frequently by emotionally disturbed, mentally retarded, or brain-damaged individuals but infrequently by the normal population. Such responses have been used to discriminate normal from disordered conditions (Moreland et al., 1995).

Confabulatory responses also illustrate the idea behind qualitative interpretations. In this type of response, the subject overgeneralizes from a part to a whole: "It looked like my mother because of the eyes. My mother has large piercing eyes just like these." Here the subject sees a detail—"large piercing eyes"—and overgeneralizes so that the entire inkblot looks like his or her mother. Although one such response has no clear or specific meaning, experts believe that the more confabulatory responses a subject makes, the more likely that she or he is in a disordered state.

Psychometric Properties

Clinical Validation

The mystique and popularity of the Rorschach became widespread in the 1940s and 1950s. This popularity was widely based on clinical evidence gathered from a select group of Rorschach virtuosos who had the ability to dazzle with blind analysis, a process by which a clinician conducts a Rorschach analysis of a patient with no former knowledge of the patient's history or diagnosis and then validates the results of the Rorschach evaluation by checking other sources (Klopfer & Davidson,

Irrational Beliefs Test

According to the cognitive viewpoint, human behavior is often determined by beliefs and expectations rather than reality. If, for example, your instructor announces that there will be an exam in the third week of classes, you will do most of your studying for it the day or two before it if you are like most students. Suppose, however, you miss the class

just before the announced exam. And suppose that a "friend" of yours plays a trick on you and tells you the exam has been canceled. If you believe your friend and therefore expect that there will be no exam, will you study as hard as you would have (if at all) had you known there would be a test? It's unlikely. The exam will still be given (reality), but your behavior will have changed because of your belief that the exam has been canceled. In view of the influence of beliefs and expectations, several cognitive-behavioral tests have been developed to measure them. In an early study, R. A. Jones (1968), for example, developed a 100-item Irrational Beliefs Test (IBT) to measure irrational beliefs (e.g., the belief that you must always succeed to be worthwhile).

The IBT requires subjects to indicate their level of agreement or disagreement with each of the 100 items on a 5-point scale (e.g., "I frequently worry about things over which I have no control"). Half of the items indicate the presence of a particular irrational belief; the other half, its absence.

The IBT found widespread use in clinical as well as research settings and has received considerable attention (Bridges & Sanderman, 2002; Taghavi, Goodarzi, Kazemi, & Ghorbani, 2006; Warren, 2011). Although the IBT was initially one of the most popular measures of irrational beliefs, its use has gradually diminished due to criticisms that these beliefs were not measured independently of the emotional consequences they were hypothesized to cause (Bridges & Harnish, 2010, p. 871). The reliability of the IBT appears to be similar to that of structured personality tests, with test–retest coefficients for short intervals (2 weeks or less) ranging from .48 to .90 for individual scales and .88 for the full scale. The validity documentation of the IBT is weak (Smith & Zurawski, 1983), although the IBT does appear to be related to both anxiety and depression (Cook & Peterson, 1986; Deffenbacher et al., 1986).

Irrational Beliefs Inventory (IBI)

The IBI was developed to improve upon some of the weaknesses of the IBT. This 50-item scale, developed in the Netherlands by Koopmans, Sanderman, Timmerman, and Emmelkamp, is based in part on the item pool of the IBT. Because it measures cognitions rather than negative affect, the IBI avoids the criticisms leveled at the IBT—that beliefs were not measured separately from the negative consequences of those beliefs. The IBI uses a 5-point scale and consists of five subscales (worrying, rigidity, need for approval, problem avoidance, and emotional irresponsibility) plus a total score.

The IBI has consistent psychometric properties across several cultures (Bridges & Sanderman, 2002), including Arabic (Al-Heeti, Hamid, & Alghorani, 2012) as well as Spanish-speaking populations (Amutio & Smith, 2007). Its internal consistency is acceptable, and the five subscales were found to be independent of each other. It has been useful in clinical and research settings when exploring the role of irrational beliefs in obsessive-compulsive disorder, social phobias, and therapy for depression (Bridges & Sanderman, 2002).

Cognitive Functional Analysis

What people say to themselves also influences behavior. If you tell yourself that you can't learn statistics, then you will likely avoid statistics. Furthermore, when confronted with a difficult statistics problem, you will tend to give up quickly. If you tell

yourself you like statistics, you will probably confront difficult statistics problems by taking your time and systematically figuring out the answers. Self-statements have been shown to influence behaviors as diverse as coping behavior in patients with fear of public speaking (Gallego, Emmelkamp, van der Kooij, & Mees, 2011), pain during pregnancy (Chang, Yang, Jensen, Lee, & Lai, 2011), treatment of schizophrenia (Bennouna-Greene, Berna, Conway, Rathbone et al., 2011), and athletic performance (Perkos, Theodorakis, & Chroni, 2002). Interestingly, positive and negative self-statements do not function in the same way. Apparently, negative self-statements do far more harm than positive self-statements do good. Thus, treatment generally involves identifying and then eliminating negative self-statements rather than increasing positive self-statements. Try to become aware of your own self-statements for a moment. What do you say to yourself as you go about your daily activities? Odds are, if you make a lot of negative self-statements, you are hindering your personal efficiency and ability to cope.

One of the most important examples of cognitive-behavioral assessment is called *cognitive-functional analysis* (Meichenbaum, 1976, 2003). The premise underlying a cognitive-functional analysis is that what a person says to him- or herself plays a critical role in behavior. The cognitive-functional analyst is thus interested in internal dialogue such as self-appraisals and expectations. Again, what do you say to yourself about yourself as you go about your daily activities? Do you constantly criticize or belittle yourself? Or do you always reassure yourself of your capabilities? Research clearly indicates these self-statements influence your behavior and even your feelings (Longe, Maratos, & Gilbert, 2010; Martin & Swinson, 2000; Masuda, Twohig, Stormo et al., 2010).

Cognitive-functional analysis is concerned with ascertaining the environmental factors that precede behavior (environmental antecedents) as well as those that maintain behavior (environmental consequences). In addition, however, a cognitive-functional analysis attempts to ascertain the internal or cognitive antecedents and consequences for the behavioral sequence (the internal dialogue). What does the person say to him- or herself before, during, and following the behavior? What is said before the behavior may influence what is done. What is said during the behavior may influence the way the behavior manifests itself. What is said following the behavior may influence its probability of recurrence.

If thoughts influence overt behavior, then modifying one's thoughts can lead to modifications in one's actions. In other words, to the extent that thoughts play a role in eliciting or maintaining one's actions, modification of the thoughts underlying the actions should lead to behavioral changes. For example, if the thought "I must have a cigarette" is consistently associated with the behavioral sequence involved in smoking, then changing that thought to "My lungs are clean, I feel healthy, and I have no desire to smoke" could help to modify the person's pattern of smoking behavior.

Parallel to Meichenbaum's technique of cognitive-functional analysis are procedures and devices that allow a person to test him- or herself, or *self-monitoring devices*. Because cognitive-behavioral practitioners value the role and responsibility of the individual in the therapeutic process, they have developed a wide variety of these devices. In the simplest case, an individual must record the frequency of a particular behavior—that is, to monitor it so that he or she becomes aware of the

behavior. To monitor your smoking behavior, simply count the number of cigarettes you smoke each day. To monitor your weight, weigh yourself each morning and record the number of pounds.

Some self-monitoring procedures are quite sophisticated. For example, a mechanical counter, marketed to the general public, can be attached to the jaw to count the number of bites a person takes when eating. The idea is to take fewer bites each day, even if only one less than the day before. Presumably, this procedure will ultimately result in a lower intake of food and eventually weight loss. Similarly, timing devices and procedures allow people to assess how long they engage in an activity. In one method, the subject plugs in a clock every time she studies, thus recording total study time. The goal is to increase this length of time, either by increasing the length of individual study sessions or by increasing the total study time within a specific period. These self-monitoring assessment tools are limited only by the imagination of the practitioner. One of us (DPS) recently bought a smartphone that was equipped with a health meter that counted the number of stets, number of flight of stairs, and distance traveled per day. Later in this chapter in our section about computerized psychological tests and measurement, we discuss the latest computer equipment used to analyze thoughts and behaviors regarding everything from social regulation of emotion (Perrez, Wilhelm, Schoebi, & Horner, 2001) to fear responses (South, Larson, White, Dana, & Crowley, 2011).

Psychophysiological Procedures