Chapter 1

Historical Trends

Chapter Questions

- Who were the major innovators in the field of personality assessment?
- How were these innovations extended to the assessment of children and adolescents?
- What is meant by the terms *personality* and *behavior*?
- What is meant by the terms *objective* and *projective* personality assessment?
- Who conducted the seminal research and coined the terms *internalizing* and *externalizing behavior* problems?

Personality assessment is a process that most individuals engage in throughout their lives (Martin, 1988). Mothers label their children as happy, cranky, or similarly shortly after birth, and often in utero (e.g., active). The musings of Alfred Binet about the personality of his two daughters are typical of observations made by parents. He described Madeleine as silent, cool, and concentrated, while Alice was gay, thoughtless, giddy, and turbulent (Wolf, 1966).

Adolescents are keenly aware of personality evaluation as they carefully consider feedback from their peers to perform their own self-assessments. Personality assessment is also prized by the business community, in which human resources personnel consult with managers and others to gauge the effects of their personality on coworkers and productivity.

Early personality assessment emphasized the assessment of enduring traits that were thought to underlie behavior or, in modern terminology, latent traits. Kleinmuntz (1967) described personality as a unique organization of factors (i.e., traits) that characterizes an individual and determines his or her pattern of interaction with the environment. Thus, personality structure is commonly thought to be a result of multiple individual traits interacting with one another, and with the person's environment.

P.J. Frick et al., *Clinical Assessment of Child and Adolescent Personality and Behavior*,
DOI 10.1007/978-1-4419-0641-0_1, © Springer Science+Business Media, LLC 2010

Definitions of Terms in Personality Assessment

Traits

A *trait* is often conceptualized as a relatively stable disposition to engage in particular acts or ways of thinking (Kamphaus, 2001 and in press). A child, for example, may be described by her parents as either shy or extroverted. The shy (introverted in psychological terms) child may tend to cope with stressful situations by withdrawing from social contact, whereas the extrovert readily approaches social situations. For parents and psychologists alike, these traits are often thought to have value for predicting human behavior, because of the presumption of trait stability across time and, in many cases, environments. In fact, because of trait stability, parents may take special precautions to ensure that the shy child adapts well to the social aspects of attending a new school by asking one of their child's friends who attends the same school to accompany the child on the first day. Similarly, a stable tendency to be shy or introverted should manifest itself in numerous social situations such as interactions in the neighborhood, at church, and in ballet class. Personality traits, then, are characterized by longitudinal and situational stability, not unlike other enduring characteristics of a person such as intelligence, height, and activity level.

The Big Five Personality Traits (Factors)

In 1961, Tupes and Christal discovered five factors of personality that appeared in the reanalysis of numerous data sets from scales of bipolar personality descriptors. These central personality traits have subsequently become the focus of an extensive research effort, including the development of tests designed to assess the constructs. One of the well known scales used to identify the "big five" in adults is the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985).

Although commonly referred to as "factors" because of their origins in factor analysis, they are prototypical examples of traits with the requisite characteristic of presumed stability. The big five factors are typically identified by bipolar comparisons that are summarized in Table 1.1. These factors are often assessed using forced-choice item formats in which adjectives are used as personality descriptors. This item format is in direct contrast to the more commonplace true/false item format that is typical of many psychological tests.

Table 1.1 Early Descriptions of the Big Five Personality Dimensions (Goldberg, 1992)

Factor I – Surgency (or introversion–extroversion)
Unenergetic vs. energetic
Silent vs. talkative
Timid vs. bold
Factor II – Agreeableness (or pleasantness)
Cold vs. warm
Unkind vs. kind
Uncooperative vs. cooperative
Factor III – Conscientiousness (or dependability)
Disorganized vs. organized
Irresponsible vs. responsible
Negligent vs. conscientious
Factor IV – Emotional stability (vs. neuroticism)
Tense vs. relaxed
Nervous vs. at ease
Factor V – Culture, intellect, openness, or sophistication
Unintelligent vs. intelligent
Unanalytical vs. analytical
Unreflective vs. reflective

Commercially available instruments such as the NEO-PI have provided new opportunities to study and refine these constructs. Given the amount of research and development in this area, the big five personality factors could eventually have a substantial impact on the field of child and adolescent personality assessment. With some noteworthy exceptions (e.g., Lynam et al., 2005) however, big five research has largely been focused on adult populations.

Temperament

A concept related to personality is *temperament*, which also emphasizes the measurement of specific traits that are hypothesized as underlying behavior across settings. In this regard, Goldsmith and Rieser-Danner (1990) observed, "most researchers consider temperament to be the behavioral manifestation of biologically influenced processes determining both the infant's characteristic response to the environment and his or her style of initiating behavior" (p. 250). Therefore, some researchers distinguish temperament from personality based on the presumed biological basis of temperament, whereas personality is thought to be formed by a dynamic interplay of biological and social factors over development (Frick, 2004; Martin, 1988). Predictably, much of the research on temperament is conducted with infants and young children. In this conceptualization, personality may be viewed as being superimposed on a person's temperamental foundation. This distinction between temperament and personality, however, is not universally agreed upon.

Behavior

In contrast to temperament and personality trait assessment, the assessment of behavior focuses on the measurement of observable behaviors, although recently the definition has been broadened to include cognitions as a type of behavior. For most purposes, Martin (1988) provides a useful definition of behavior.

> When applied psychologists speak of behavior, they are usually referring to that range of human responses that are observable with the naked eye during a relatively brief period of time in a particular environment. This conception of behavior rules out biochemical and neurological events, for example, because they are not observable by the unaided eye. Behavior is differentiated from traits or dispositions because the latter may only be seen if behavior is aggregated over relatively long periods of time and in a number of environmental contexts. Classical examples of observed behaviors of interest to child psychologists include tantrum behavior among young children, aggressive interactions with peers, attempts at conversation initiation, and so forth (p. 13).

There are, therefore, several distinguishing features of behavioral assessment methods that differentiate them from trait assessment measures. First, behavioral assessment methods have a different theoretical foundation and associated set of premises. Behavioral assessment methods draw heavily on the theory and research tradition of operant conditioning as exemplified by the work of B. F. Skinner (Skinner, 1963). This research tradition also emanates primarily from laboratory research, as opposed to clinical practice; thus, it is often considered to be more empirically based.

Second, behavioral assessment methods are distinguished from medical models of assessment more than are trait-based methods. The medical model assumes that symptoms are caused by underlying conditions, and it is the medical condition that must be measured, diagnosed, and treated to remove the symptoms (see Chap. 3 for a more extended discussion of the medical model). In direct contrast, behavioral assessment emphasizes the measurement and treatment of the symptoms

or behavior itself, and makes no assumptions regarding underlying cause.

Third, behavioral assessment places a premium on the assessment of discrete behaviors. For example, behavioral assessment may emphasize the measurement of finger tapping while completing seatwork in school, as opposed to aggregating several behaviors to form a test or scale with several items that measure "motor activity" in the classroom. This situation is changing, however, with the work of Thomas Achenbach, Keith Conners, Cecil Reynolds, and others (see Chap. 7), all of whom began combining behaviors into dimensions of behavior that may or may not differ from trait-based assessment methods.

As we suggested in the preface, we think that it is premature to reify any of these approaches as the ultimate method for assessing children's psychological adjustment. We merely seek progress in our methods and theories. This cautious approach seems warranted as the distinctions between the methods have become blurred as the science of assessment emerges. Furthermore, we think that each approach may be more or less helpful for answering particular assessment questions. Clearly, some of the questions directed at psychologists are trait-based, whereas others require the measurement of distinct behaviors. For example, a parent who asks, "Will my child ever become more outgoing like his sister?" is asking for trait assessment, but the parent who queries, "How can I get him to stop wetting the bed?" may require behavioral assessment expertise.

Early History of Personality Assessment

Formal personality measures emerged as a logical outgrowth of other efforts to measure individual differences, most notably the experimental methods of Wundt, Galton, and others (Chandler, 1990). Of the early assessment luminaries, Sir Francis Galton is one of the most notable. Although well known for his intelligence measurement contributions, he also studied the measurement of "character." In order to introduce the utility of personality measurement, Galton (1884) recounted the personality test invented by Benjamin Franklin as a crude form of personality measure. The scale was described in the tale of "The Handsome and Deformed Leg," in which Franklin recounts how his friend tested people so as to avoid those who, "being discontented themselves, sour the pleasures of society, offend many people, and make themselves everywhere disagreeable" (p. 9). This friend sought to diagnose such pessimistic individuals by showing them an attractive leg and a malformed one. If the stranger showed more interest in the ugly leg, the friend became suspicious and subsequently avoided this person. Franklin astutely identified this "test" as a grotesque but, nevertheless, an effective personality assessment device. Galton concluded: "The other chief point that I wish to impress is, that a practice of deliberately and methodically testing the character of others and of ourselves is not wholly fanciful, but de-serves consideration and experiment" (p. 10).

Intelligence tests, acknowledged as the first fruits of the psychometric movement, reached prominence early in the twentieth century with the introduction of the original Binet and Simon scale and numerous variants (Kamphaus, 2001 and in press). A lesser known fact is that Alfred Binet developed some intelligence test items that resembled stimuli used 30 years later in apperceptive techniques for assessing personality (DuBois, 1970). Test development activity also received a boost from the World War I effort, when ability testing became widespread (Kamphaus, 2001 and in press). Thus, it is no coincidence that the first formal and widely used measures of personality were developed about this same time.

Robert S. Woodworth

The Woodworth Personal Data Sheet was published in 1918 as a result of the surge of interest in testing potential soldiers. Woodworth developed a list of 116 questions about daydreaming, worry, and other problems. Some sample items from the Woodworth Personal Data Sheet (Woodworth, 1918) include:

> Do people find fault with you much?
> Are you happy most of the time?
> Do you suffer from headaches and dizziness?
> Do you sometimes wish that you had never been born?
> Is your speech free from stutter or stammer?
> Have you failed to get a square deal in life?

The examinee responded to each question with "yes" or "no" (French & Hale, 1990).

According to French and Hale (1990), the Woodworth Personal Data Sheet served as the foundation for the development of the Thurstone Personality Scale and the Allport Ascendance-Submission Test, among others. DuBois (1970) described the Personal Data Sheet as "the lineal ancestor of all subsequent personality inventories, schedules and questionnaires" (p. 94).

The Personal Data Sheet was an important practical innovation because, prior to this time, all military recruits suspected of having mental health disorders, stress disorders in particular, had to be identified by being interviewed by trained interviewers. The Personal Data Sheet allowed for the screening of large numbers of recruits without the time and expense of using huge cadres of interviewers (Kleinmuntz, 1967).

Thus, it was not basic research on personality or employee selection that led to the eventual popularity of personality testing. Instead, it was the need for diagnosis created by World War I which provided considerable evidence of the need for personality tests. The successful World War I and then later World War II applications of psychological testing proved that psychology could make practical contributions to society by identifying, accurately and time-efficiently, those in need of mental health services.

After World War II the mental health needs of citizens, veterans in particular, were the focus of greater attention. In the postwar years, the U.S. Veteran's Administration began to hire psychologists in large numbers to diagnose and treat veterans suffering from significant emotional disturbance. Psychologists brought their psychometric expertise to bear again by contributing new methods to the diagnostic process. The increased need for postwar mental health services, therefore, created the fertile ground in which personality testing flourished. As Kleinmuntz (1967) noted, "The most popular personality tests of the past 30 years grew out of the need to diagnose or detect individuals whose behavior patterns were psychopathological" (p. 10). The use of personality tests after the first and second world wars expanded beyond diagnosis into many areas including counseling, personnel selection, and personality research (Kleinmuntz, 1967).

Projective Techniques

The central assumption underlying projective testing is that the use of less well-defined stimuli that are prone to a variety of interpretations will encourage clients to reveal information that they otherwise would not share in response to direct questioning (Chandler, 1990). Given that test stimuli (e.g., ink blots) or questions were not clearly linked to known personality traits, projective testing depended heavily on the explicit or implicit personality theory that was favored by the test developer. Theory was necessary to determine the underlying nature or cause of the projected thoughts, emotions, or behaviors.

The most popular theory of the post-World War I era was psychodynamic personality theory as espoused by Sigmund Freud (1936) and others. Psychodynamic personality theory provided a useful theoretical framework for the development of projective assessment measures due to the concepts of repression, projection, and other constructs that are entirely consistent with the use of atypical test stimuli for identifying personality traits (see Chap. 10 for a more extended discussion of the basic assumptions of the projective technique).

Association Techniques

The use of association techniques, such as word association methods, for assessment purposes can be traced as far back as the work of Aristotle (DuBois, 1970). In relatively recent history, Sir Francis Galton began studying association techniques as early as 1879. Galton's contribution to the study of association was his introduction of scientific rigor to the enterprise. He used experimental methods to study association methods, including quantitative scaling of the results (DuBois, 1970).

Subsequently, Kraepelin, Wundt, Cattell, Kent, and Rosanoff studied the associations of patients and research participants to word lists, recording such variables as response time and type of association. The latter names, Kent and Rosanoff, may be least familiar to many readers because the other names are linked with the illustrious history of intellectual assessment. Kent and Rosanoff made their contribution solely to the study of associations by developing a list of 100 stimulus words and systematically recording the associations of 1,000 normal subjects (DuBois, 1970). This effort represents an important initial attempt at developing norms to which researchers and clinicians could compare the responses of clinical subjects.

The renowned psychoanalyst Carl Jung made extensive use of association techniques for the study of personality. In an address at Clark University in 1909, he described his research efforts in detail and provided some insight into the types of interpretations commonly made of these measures. Jung described his association word list as *a formulary.* His list consisted of 54 words including *head*, *to dance*, *ink*, *new*, *foolish*, and *white*. According to Jung (1910), normality could be distinguished from psychopathology with this formulary using variables such as reaction time and response content. In his speech, he provided a transcript of the responses of a normal individual and of a "hysteric." A sampling of their associations to the formulary follows:

Stimulus	Normal	Hysterical
To sin	Much	This is totally unfamiliar to me, I do not recognize it
To pay	Bills	Money
Bread	Good	To eat
Window	Room	Big
Rich	Nice	Good, convenient
Friendly	Children	A man

As noted previously for other measures, reaction time to the stimulus words was also interpreted by Jung. He gave a glimpse of one such interpretation in the following quote:

> The test person waives any reaction; for the moment he totally fails to obey the original instructions, and shows himself incapable of adapting himself to the experimenter. If this phenomenon occurs frequently in an experiment, it signifies a higher degree of disturbance in adjustment (p. 27).

These early word association methods set the stage for the development of other association (projective) techniques, such as the Thematic Apperception Technique and Rorschach's test, both of which used pictures in lieu of word lists to elicit associations.

Thematic Apperception Test

The Thematic Apperception Test (TAT) of Henry A. Murray constitutes a prototypical example of a projective device. Charles E. Thompson summarized the central tenet of the projective approach in the following quote taken from his 1949 adaptation of Murray's TAT.[1]

> If the pictures are presented as a test of imagination, the subject's interest, together with his need for approval, can be so involved in the task that he forgets his sensitive self and the necessity of defending it against the probing of the examiner, and, before he knows it, he has said things about an invented character that apply to himself, things which he would have been reluctant to confess in response to a direct question (p. 5).

The TAT is unique among projective measures in that it has traditionally been interpreted qualitatively, even though quantitative scoring methods are available (Kleinmuntz, 1967). Murray's original approach to TAT scoring was entirely qualitative and psychoanalytically based. He proposed the following categories for analyzing the characteristics of the stories given by the subject (Murray, 1943).

1. *The Hero.* This is the person with whom the subject seems to identify. The hero may share characteristics such as age, gender, occupation, or other features with the subject that aid identification. The hero's traits should be evaluated to determine the self-perceptions of the subject including superiority, intelligence, leadership, belongingness, solitariness, and quarrelsomeness.
2. *Needs of the Hero.* Needs may include those for order, achievement, and nurturance.
3. *Environmental Forces.* Factors that affect the hero and these are also referred to as *press.* An example would involve scoring aggression if the hero's property or possessions were destroyed in a story.
4. *Outcomes.* The success of the hero and the hero's competencies are assessed by evaluating the outcomes of stories.
5. *Themas.* Themas assess the interplay of needs and presses, and they reveal the primary concerns of the hero.
6. *Interests, Sentiments, and Relationships.* For this aspect of scoring the examiner records the hero's preferences for topics.

Murray's qualitative scoring system for the TAT is a classic example of systems that dominated the early interpretation of projective devices. Numerous quantitative scoring systems followed Murray's original work as exemplified by scoring systems eventually developed for Rorschach's test.

The Rorschach

Hermann Rorschach (1884–1922) was a major figure in Swiss psychiatric research who began his work as a physician in 1910. He married a Russian colleague who became his comrade and collaborator (Morgenthaler, 1954). He served as a physician in a hospital in Herisau until his death from complications of appendicitis in 1922. His death was

[1]Thompson's modification is identical to the original TAT with the exception that African American figures are used as characters on the stimulus cards. Thompson found that African Americans did not respond optimally to the original TAT pictures. In fact, one of his patients asked if he could imagine that the people in the pictures were "colored," and if he could make up some stories about "colored people."

described as a critical blow to Swiss psychiatry. In a 1954 eulogy to Dr. Rorschach, published in the English translation of his original work, Morgenthaler attempted to describe Rorschach for future generations.

> Flexibility of character, rapid adaptability, fine acumen, and a sense for the practical were combined in Hermann Rorschach with a talent for introspection and synthesis. It was this combination which made him outstanding. In addition to this rare nature, which tempered personal emotional experience with practical knowledge, he possessed sound traits of character most valuable in a psychiatrist. Most important of these were an unerring tendency to search for the truth, a strict critical faculty which he did not hesitate to apply to himself, and a warmth of feeling and kindness. (p. 9)

Rorschach's approach to personality assessment was novel in many respects. The test stimuli used were inkblots placed on paper that was then folded in half. Rorschach was not, however, interested in the content of the subject's response to the inkblots. Rather, he was interested in the form of the response (or its *function*). Some functions of interest included the number of responses, perception of color or movement, and perception of the whole vs. the parts. These and other characteristics of Rorschach responses continue as part of modern scoring systems (see Chap. 10).

Rorschach first offered his method as an experiment. His original sample is described in Table 1.2. He expressed a desire for larger sample sizes but noted that the number of experiments was limited because the stimulus figures were damaged by passing through hundreds of hands.

Rorschach's legacy, his original inkblots, and many of the associated scoring criteria remain influential as the test continues to enjoy popularity. Several scoring systems have been offered for the Rorschach, with the Exner Comprehensive System (Exner & Weiner, 1982) contributing most to the continuing usage of the instrument.

Table 1.2 Rorschach's Original Research Sample

Sample	N
Normal, educated	55
Normal, uneducated	62
Psychopathic personality	20
Alcoholic cases	8
Morons, imbeciles	12
Schizophrenics	188
Manic-depressives	14
Epileptics	20
Paretics	8
Senile dements	10
Arteriosclerotic dements	5
Korsakoff and similar states	3

Source: Adapted from Rorschach (1951).

Sentence Completion Techniques

Sentence completion techniques are venerable personality assessment methods of the association tradition that can trace their roots to Payne (1928). The sentence completion method, however, obtained a substantial boost in popularity because of its use by the U.S. Office of Strategic Services (OSS), the forerunner of the Central Intelligence Agency (CIA). Henry Murray was the coordinator of a sophisticated OSS assessment effort. About 60 military assessment stations, staffed by American psychologists, were situated in the USA and abroad to screen recruits for sensitive and dangerous assignments. Some of the methods used in this ambitious program are described in the following quote from DuBois (1970):

> In one of the stations near Washington, recruits in fatigue uniforms assumed a false identity and developed a cover story, which the staff members during the three-day stay endeavored to break. The procedures described in a comprehensive report (OSS, Assessment Staff, 1948) were varied: casual conversations,

> searching interviews, the sentence completion test, questionnaires about health and working conditions and personal history, conventional aptitude tests such as map memory and mechanical comprehension, and a number of situational tests (p. 111).

After World War II, the sentence completion technique continued to enjoy some favor among psychologists. The well-known Rotter Incomplete Sentences Blank was published in 1950 (Rotter & Rafferty, 1950). Sentence completion methods have a lengthy history of use with children and adolescents as well. They enjoyed worldwide use in countries such as Finland, Germany, Denmark, India, Japan, and Taiwan (Haak, 1990).

Projective Techniques for Children

The use of projective techniques with children dates back to the early part of the twentieth century when Florence Goodenough began to study children's human figure drawings (DuBois, 1970). Goodenough noted, as did others, that children's drawings were affected by their emotionality. The typical paradigm for drawing techniques has been to have the child draw a picture of a person. Traditionally, the content of the drawings has been interpreted as a measure of child adjustment and personality. Some aspects of content that were extensively studied for adults included (Swensen, 1968):

> Size of the person depicted
> Placement on the page (bottom, top, corner, etc.)
> Stance (vertical, horizontal, balanced, etc.)
> Line quality (heavy, light, etc.)
> Shading
> Erasures
> Omissions (missing body parts)
> Distortion (poor proportion of body parts)

Various interpretations have been associated with these and other content variables over the years. Heavy lines, for example, have been associated with assertive and aggressive individuals and light lines have been viewed as being indicative of passive individuals (Koppitz, 1968). Swensen (1968) found such interpretations to be highly unreliable. The most reliable and valid interpretive approach involved making general judgments about the mental health status of the individual based on the overall quality of the drawing, rather than specific content interpretations (Cummings, 1986).

The TAT, among other projective methods, has also been adapted by many for use with children and adolescents. One of the most well-known TAT adaptations is the Children's Apperception Test (CAT) (Bellak & Bellak, 1949b), designed for ages 3–10. The CAT consists of ten pictures with animals as stimuli in contrast to the TAT's depictions primarily of people. The Rorschach has also been widely used with children, and several compendiums of child responses have been published to aid interpretation (e.g., Ames, Metraux, Rodell, & Walker, 1974).

The proper interpretation of children's projective responses remains a topic of debate. Indeed, the degree to which children obey the *projective hypothesis* has been questioned. Chandler (1990) elucidates the nature of the projective hypothesis as follows:

> *Projection*, in common usage, means to cast forward. In this sense, *projection* implies a direct extension of psychological characteristics onto the outer world. But *projection* also has a specific meaning within psychoanalytic theory. Freud (1936) used the term to refer to the process that occurs when the ego, faced with unacceptable wishes or ideas, thrusts them out onto the external world as a means of defense. In projection the individual attributes his or her own thoughts and actions to someone else. Thus, if one's own faults or feelings are unacceptable to

> the ego, they may be seen as belonging to someone else; in the process, the material may become distorted or remain partially repressed. From such a perspective, projective material would not be seen as direct representation of aspects of the personality, certainly not with the sort of one-to-one correspondence that the first meaning of projection implies (p. 57).

For adults as well as children, the process of projection still rests primarily on a theoretical rather than empirical foundation. In the absence of data to support the projective hypothesis, psychologists have focused on the use of psychometric methods to assess the reliability of obtained scores and the validity of score inferences. This shift to the accumulation of psychometric evidence for measures is reflected best in the work of Exner. In the 1960s, John Exner began a research program designed to take the best of the Rorschach scoring systems and incorporate their features into a comprehensive system (Exner & Weiner, 1982). Further, a standard method for scoring responses on the test has led to scores that have proven to be reliable and, as a result, has set the stage for direct tests of the validity of various interpretations that can be made from them. The application of psychometric standards to projective measures is a clear departure from a long history of qualitative analysis and interpretation. The efforts of Exner and others have set a new course for projective measures in that they are increasingly held to the same standard as tests of intelligence, adaptive behavior, and "objective" personality assessment methods.

Objective Tests

Although we acknowledge that the distinction between projective and objective testing is an oversimplification, it is nevertheless useful for pedagogical purposes. Objective methods can be differentiated from projective tests in several ways. First, objective methods are often considered to be atheoretical and/or empirical. As opposed to requiring the examiner to use theory to interpret results, the results often derive their meaning from empirical procedures, such as matching a person's results to those of a clinical sample. Second, objective methods are not likely to be based on psychodynamic theory. Hence, the results of objective measures are often considered to be less useful for providing insight into the dynamics of an individual's interactions with the world. Third, objective methods take greater advantage of measurement science for the development of tests. Issues of item selection, reliability, and validity are often emphasized in the test manuals.

Minnesota Multiphasic Personality Inventory (MMPI)

Until the advent of the MMPI, projective techniques reigned supreme. In a 1961 survey of tests used by psychologists in the USA, the MMPI was the only nonprojective measure mentioned among the top ten most used tests. Of the top ten tests, five were intelligence tests and four were projective measures (Sundberg, 1961). A confluence of circumstances, including the expansion of clinical psychology practice during and after World War II, and the emergence of an extensive research base led to almost immediate acceptance of this self-report personality inventory (Kleinmuntz, 1967). Further, the MMPI was one of the first tests to gain popularity with others outside of the mental health professions (see Box 1.1). However, this popularity led to significant friction and disagreements over the relative merits of the MMPI, and its objective methods, compared with the popular projective techniques. This tension is reflected in the comments of Paul Meehl that are summarized in Box 1.2.

The MMPI (Hathaway & McKinley, 1942) differed from its predecessors (such

Box 1.1

Sample Items from 1960s MMPI Spoofs

Personality testing eventually became popular enough to warrant derision by members of Congress, well-known humorists such as Art Buchwald, and others. Some of these alternate MMPI items were published in a 1965 issue of *American Psychologist* (p. 990) to poke fun at this method of personality assessment.

> When I was younger I used to tease vegetables.
>
> I think beavers work too hard.
>
> I use shoe polish to excess.
>
> When I was a child I was an imaginary playmate.

Box 1.2

Meehl on Science and Technics

Paul Meehl is considered one of the founders of modern personality assessment and diagnostic practice. His 1973 collection of selected papers published by the University of Minnesota Press provides a unique glimpse of the genius of an astute clinician. In the following quote, Dr. Meehl discusses the tension between science and practice in psychology and takes a stance against theoretical dogmatism:

> Doubtless every applied science (or would-be science) presents aspects of this problem to those working at the interface between science and technics, as is apparent when one listens to practicing attorneys talking about law professors, practitioners of medicine complaining about medical school teaching, real engineers in industry poking ambivalent fun at academic physicists, and the like. So I do not suggest that the existential predicament of the clinical psychologist is unique in this respect, which it certainly is not.
>
> But I strongly suspect that there are few if any fields of applied semiscientific knowledge in which the practitioner with scientific interests and training is presented *daily* with this problem in one guise or another, or in which its poignancy, urgency, and cognitive tensions are so acute. I am aware that there are *some* clinical psychologists who do not experience this conflict, but I have met, read, or listened to very few such during the thirty years since I first began working with patients as a clinical psychology trainee. Further, these rare exceptions have seemed to me in every case to be either lacking in perceptiveness and imagination or, more often, deficient in scientific training and critical habits of mind.
>
> When I encounter a hard-nosed behaviorist clinician who knows (for sure) that Freud's theory of dreams is 100 percent hogwash and is not worth five hours of his serious attention; or, toward the other end of the continuum, when I converse with a devoted Rorschacher who knows (for sure) that the magic inkblots are highly valid no matter what the published research data indicate-I find both of these attitudes hard to understand or sympathize with (p. viii).

as the Personal Data Sheet) in at least one fundamental way. It was one of the first tests to use an empirical approach to objective personality test development. Most tests of the day used a priori or rational–theoretical approaches (Martin, 1988). Rational approaches, as the name implies, depend heavily on the test author's theory of personality for many aspects of test construction, including item development and scoring methods. On the other hand, empirical approaches make greater use of empirical data to make such decisions (see Chap. 2 for a more detailed discussion of this distinction).

The MMPI used an item selection method called *empirical criterion keying* (Anastasi &

Urbina, 1998). Simply stated, this method involved selecting items that meet an empirical criterion. In the case of the MMPI, items were selected if they were able to routinely differentiate clinical groups from samples of "normal" subjects, and distinguish clinical groups from one another. For example, items for the Psychasthenia scale (a scale designed to assess anxiety-related problems such as obsessions and fears) were selected based on a clinical group of 20 cases, the results of which were compared with "normals" and other clinical groups to identify items that best differentiated the target clinical group from the others.

The original version of the MMPI consisted of 550 statements printed on separate cards. The cards were separated by the patient into three categories: true, false, and cannot say. The first MMPI clinical scales were linked to the major diagnostic nosology of the day (Kleinmuntz, 1967), which is another factor that contributed to its popularity. The ten clinical scales of the original version included are provided in Table 1.3. The MMPI has undergone many changes since its inception, with the most recent edition entitled the MMPI-2. In fact, some of the scale names (e.g., Psychasthenia) had fallen into disuse at about the time of original publication (Kleinmuntz, 1967). A chronology of MMPI developments is listed next and a thorough discussion of this important measure is provided in Chapter 6.

Table 1.3 The Original Scales from the MMPI

Clinical scales
Hypochondriasis
Depression
Hysteria
Psychopathic deviate
Masculinity–Femininity
Paranoia
Psychasthenia
Schizophrenia
Hypomania
Social Introversion
Validity scales
Question scale
Lie scale
F scale
Correction scale

Source: Kleinmuntz, 1967.

MMPI Version	Publication Date
MMPI	1942
MMPI-2	1989
MMPI-A (Adolescent)	1992

The "Children's MMPI"

Not surprisingly, the MMPI profoundly influenced child assessment practice including the development of the Personality Inventory for Children (PIC) in the 1950s. The PIC was based on a pool of 600 items; hence, it was comparable in length to the MMPI. A central difference between the MMPI and the PIC was the informant. The PIC was not a self-report measure. Instead, a parent rated the child's behavior. Lachar (1990) gave the following rationale for this decision:

> Selection of the parents as the source of PIC test responses helps overcome two of the major obstacles posed by requesting the referred child or adolescent to respond to numerous self-report descriptions in order to obtain a multiple-scale objective evaluation. The majority of children seen by mental health professionals in a variety of settings appear for such an evaluation because of their noncompliant behaviors and/or documented problems in academic achievement, most notably in the development of reading skills. Therefore, it seems unlikely that a technique requiring such children to read and respond to a large set of self-descriptions will find broad acceptance in routine clinical practice (p. 299).

The scales of the PIC were derived using factor-analytic methods. Thus, the PIC, like the MMPI, was developed with a heavy emphasis on empirical methods (see Chap. 6). In the 1960s, empirical methods of test development were also applied to the development of other types of child assessment devices.

Rating Scales

Parent and teacher ratings of children's behavior and emotions trace their roots to the assessment of adult psychopathology in hospital settings. Conceptualized as one type of observational method, rating scales were developed in the 1950s for use by nurses and other caretakers who worked closely with patients for extended periods of time. One of the first such measures was the Wittenborn Psychiatric Rating Scales (1955). According to Lorr (1965), the scales were designed for recording currently observable behavior and symptoms in hospitalized mental patients. The Wittenborn could be completed by a social worker, psychologist, psychiatrist, nurse, attendant, or other individual familiar with the patient's day-to-day behavior. The original scale consisted of 52 symptoms that were combined to yield 9 scores for acute anxiety, conversion hysteria, manic state, depressed state, schizophrenic excitement, paranoid condition, paranoid schizophrenic, hebephrenic schizophrenic, and phobic compulsive. An item assessing withdrawal included the following options:

> No evidence of social withdrawal
> Does not appear to seek out the company of other people
> Definitely avoids people

The Wittenborn was used for diagnostic purposes as well as for the design and evaluation of treatment (Kleinmuntz, 1967). Reviewers of the day found many reasons to recommend the Wittenborn, including a thorough research base (Eysenck, 1965) and easy administration and scoring (Lorr, 1965). There was considerable concern, however, about overlapping scales. The hebephrenic schizophrenic and schizophrenic excitement scales correlated at .88 and the paranoid condition and paranoid schizophrenic scales correlated at .79. On the basis of these data, Eysenck (1965) and Lorr (1965) recommended that these scales be combined to reflect this overlap.

Other rating scales of adult psychopathology for use in inpatient settings included the Hospital Adjustment Scale (McReynolds, Ballachey, & Ferguson, 1952) and the Inpatient Multidimensional Psychiatric Scale (Lorr, 1965), a rating of symptomatology completed by the clinician after a diagnostic interview. Such measures probably fell into decline for many reasons, one of the most prominent being the deinstitutionalization movement of the 1970s. These instruments did, however, clearly demonstrate the utility of ratings of behavior as practical and useful assessment tools. These scales set the stage for the development of parent and teacher rating scales of child behavior.

Internalizing and Externalizing Dimensions

Research into the diagnosis of child psychopathology led to increased attention to the use of rating scales for child diagnosis. In a 1978 article in *Psychological Bulletin*, Thomas Achenbach and Craig Edelbrock introduced many clinicians to the terms *internalizing* and *externalizing* psychological disorders of childhood. These dimensions, or types of child psychopathology, were based on an extensive empirical analysis (typically using factor analysis) of parent and teacher behavior problem rating scales. Children experiencing adjustment difficulties of the internalizing variety have also been described as over-controlled, with problems of inhibition, anxiety, and, perhaps, shyness (Edelbrock, 1979). On

the other hand, children with externalizing problems have been described as undercontrolled with difficulties such as aggression, conduct problems, and acting-out behavior (Edelbrock, 1979).

These two dimensions of child psychopathology trace their roots to the work of Peterson (1961), who labeled the syndromes as *conduct problem* (externalizing) and *personality problem* (internalizing). The veracity of the broad internalizing and externalizing categorizations of child psychopathology is supported by many factor-analytic investigations of both parent and teacher rating scales alike (Edelbrock, 1979). The utility of these behavioral distinctions was also demonstrated in an early study of 163 consecutive referrals to a child psychiatry outpatient department (Cohen, Gotlieb, Kershner, & Wehrspann, 1985). Children were classified as externalizers and internalizers based on the Achenbach Child Behavior Checklist (CBCL; a parent report form) and the Teacher Report Form (TRF) (see Chap. 7). The resulting analyses uncovered distinct differences between the two groups, particularly on the Teacher Form. Internalizers were found to be more intelligent, better readers, less egocentric, and they used more adaptive means of coping with stressful situations. Internalizers were also generally rated as being less disruptive than externalizers.

Numerous independent research studies, many of which have been conducted internationally (Ivanova et al., 2007), have demonstrated strong factor-analytic support for these two types of child behavioral adjustment. This preponderance of evidence, supported by other types of validity evidence, has resulted in these factors serving as the foundation for the development of many teacher and parent rating scales, most notably the Achenbach Child Behavior Checklist (Achenbach, 1991b) and the Behavior Assessment System for Children (BASC-2; Reynolds & Kamphaus, 2004). Furthermore, the terms *internalizing* and *externalizing* are now a part of psychologists' everyday parlance when discussing child behavior problems.

The Diagnostic and Statistical Manual of Mental Disorders Diagnostic Systems

Diagnostic systems have had a profound impact on child assessment by defining symptoms and other diagnostic indices that have subsequently been incorporated into various assessment methods. The most obvious link exists between the various editions of the *Diagnostic and Statistical Manual of Mental Disorders* (currently *DSM-IV-TR;* American Psychiatric Association, 2000) and structured interview methods designed to assess symptomatology associated with various *DSM* diagnostic categories (see Chap. 11). Given this interdependence, a thorough knowledge of the nature of the *DSM* and its variants is prerequisite to the study of child assessment.

As mental disorders became recognized as conditions worthy of medical treatment, the need for diagnostic systems became more pressing. Consistent diagnosis was necessary for communication among clinicians and for the conduct of epidemiological research and other scientific investigations. The American Medico-Psychological Association (now the American Psychiatric Association) began efforts to standardize diagnostic procedures in 1917. The first diagnostic manual, a classification of mental disease, was produced by the American Psychiatric Association in conjunction with the U.S. Census Bureau (Widiger, Frances, Pincus, Davis, & First, 1991).

The first edition of the *DSM* (*Diagnostic and Statistical Manual*) appeared in 1952. Part of the impetus for the creation and

frequent updating of the *DSM* has been provided by the *International Statistical Classification of Diseases, Injuries, and Causes of Death* (*ICD*). The *ICD*, currently *ICD-11*, is published by the World Health Organization. The *DSM* has been revised both to coordinate with the *ICD* and to add criteria for conditions that are of concern to US clinicians, and delete conditions that are not apparent in the USA (Widiger et al., 1991). The *DSM* has also been revised because of a desire to make the diagnostic categories more evidence-based. Prior to the development of the *DSM-III*, the system was based primarily on the expert judgment of a relatively small number of clinicians. The *DSM-II*, for example, was finalized after review by 120 psychiatrists in February of 1967 (Widiger et al., 1991).

The *DSM-IV* (APA, 1994) was based on a more comprehensive research base than any of its predecessors. According to Widiger et al. (1991), three research methods have formed the empirical cornerstone for the development of *DSM-IV.*

1. *Literature reviews*: Comprehensive reviews of the research were completed to advise the various committees charged with proposing diagnostic criteria for conditions. These reviews were seen as a way to mitigate against biases on the part of some committees (Widiger et al., 1991).
2. *Data reanalyses*: Existing data sets were made available to the *DSM-IV* committees supported by funding from the John D. and Catherine T. MacArthur Foundation. According to Widiger et al. (1991), these data set reanalyses allowed the committees to evaluate the validity of current diagnostic algorithms and pilot-test new proposals for making diagnoses.
3. *Field trials*: These studies were particularly useful for testing the reliability and validity of diagnostic categories (Widiger et al., 1991).

Table 1.4 Chronology of Diagnostic Systems Developed Under the Auspices of the American Psychiatric Association

1917	Classification of Mental Disease
1933	Standard Classified Nomenclature of Disease
1952	Diagnostic and Statistical Manual of Mental Disorders I (DSM-I)
1968	Diagnostic and Statistical Manual of Mental Disorders II (DSM-II)
1980	Diagnostic and Statistical Manual of Mental Disorders III (DSM-III)
1987	Diagnostic and Statistical Manual of Mental Disorders III-Revised (DSM-III-R)
1994	Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV)
2000	Diagnostic and Statistical Manual of Mental Disorders IV-Text Revision (DSM-IV-TR)

The *DSM-IV-TR*, because of its greater reliance on empirical methods, has had an even more substantial impact on the personality assessment process (see Chap. 3). The chronology of the *DSM* is provided in Table 1.4.

IDEA and Special Education

The 1974 Education of Handicapped Children's Act, better known as Public Law 94–142 (IDEA), and its reauthorization, the Individuals with Disabilities Education Improvement Act (IDEIA), mandated special education and related services for children classified as having an emotional disturbance. As a result, some method had to be developed to define child problems and determine children's eligibility for special education services. Under IDEA, and the subsequent IDEIA with few substantive

changes, the classification of severe emotional disturbance was defined as follows:

> The term means a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree which adversely affects school performance: (a) an inability to learn which cannot be explained by intellectual, sensory, or health factors; (b) an inability to build or maintain satisfactory relationships with peers and teachers; (c) inappropriate types of behavior or feelings under normal circumstances; (d) a general pervasive mood of unhappiness or depression; or (e) a tendency to develop physical symptoms or fears associated with personal or school problems (Federal Register, 1999).

With its passage in the 1970s, this law effectively mandated US public schools to identify and serve children with behavioral or emotional problems, many of which had previously been educated in a variety of settings, including residential treatment programs or state mental hospitals. Consequently, these laws expanded school-based diagnostic practices to include evaluation for the presence of "emotional disturbance," just as had been more commonly done for developmental and learning disorders. These federal mandates also enhanced the popularity of rating scales (particularly teacher ratings) as assessment methods of choice in many school systems.

The IDEA nosology of emotional disturbance (the word *severe* has now been removed) has long been the target of criticisms that it is invalid, restrictive, or otherwise flawed (Forness & Kritzer, 1992). Bower (1982), the recognized developer of the conceptual basis of the IDEA diagnostic categories, raised similar questions about the system. He noted that:

> Section ii [which excludes the socially maladjusted from the IDEA act] is, one would guess, a codicil to reassure traditional psychopathologists and budget personnel that schizophrenia and autism are indeed serious emotional disturbances on the one hand, and that just plain bad boys and girls, predelinquents, and sociopaths will not skyrocket the costs on the other hand. It is clear what these modifications and additions were intended to do. It is perhaps not clear what such public policy and fiscal modifications do to the conceptual integrity of the definition and the nature and design of its goals (p. 56).

Despite such controversy, the IDEA classification system remains as the "gold standard" nosology when determining child eligibility for oft-costly special education and related services.

Constructs (Dimensions) vs. Categories

The measurement problems associated with categorical diagnostic systems such as the DSM are well known (Kamphaus & Campbell, 2006), and the advantages of using dimensional methods are recognized as well. Achenbach and McConaughy (1996), for example, noted that the yes/no nature of categorical methods does not necessarily account for children whose problems vary in degree or severity. As a result, the nexus between normality and psychopathology cannot be well understood with categorical methods, since most high prevalence problem behaviors of childhood, such as inattention and hyperactivity, are not classifiable when below diagnostic threshold levels. Substantial evidence is emerging to suggest that child behavior problems such as inattention, hyperactivity, depression, and conduct problems, in fact, fall along continua in the population (Hudziak et al., 1998; Scahill et al., 1999).

As a result, dimensional classification methods have demonstrated their usefulness

in the study of psychopathology. For example, dimensional approaches have demonstrated more predictive validity than categorical approaches (Fergusson & Horwood, 1995), as well as statistical reliability (Cantwell, 1996). Such methods also minimize the need for clinical judgment and inference (Haynes & O'Brien, 1988), provide greater sensitivity to the presence of comorbid conditions (Caron & Rutter, 1991), and have the ability to depict multiple symptom patterns in a given individual simultaneously (Cantwell, 1996). Furthermore, the use of dimensional, person-oriented approaches to identify subtypes or clusters of individuals may lead to more efficient, streamlined subtype-specific intervention and prevention services (Achenbach, 1995; Bergman & Magnusson, 1997). The overlap and tensions between categorical and dimensional classification methods will be elucidated throughout this text. It is important to do so because a merger of these methodologies is likely in the DSM-V and other future diagnostic systems (Rounsaville et al., 2002).

Future Trends

The pace of change in personality assessment is ever hastening. There is increasing interest in the development of new child assessment methods, providing clinicians with a wide array of assessment options. In 1990, Tuma and Elbert (1990) identified test development and research trends that remain true to the present day.

> It is apparent that personality assessment is undergoing rapid development in all areas: projective, objective, and behavioral assessment; clinical interviewing and informal assessment; and environmental assessment. The developments outlined above encompass observable behavior, structured and unstructured use of tests and interviews, and assessment of broad- and narrow-band aspects of personality, all within the context of a person's situation/environment. Thus, in spite of various criticisms and some apparent decrease in the use of personality assessment instruments (*they were referring to projective devices primarily*), all indications point to vigorous activity in the area that promises to continue." (p. 23; italics added).

In the past, the technology of personality assessment has been viewed as lagging behind other areas of assessment, such as intelligence and achievement testing (Martin, 1998). This conclusion is no longer true. New measurement science rigor is being applied to the development of behavioral rating scales, interview methods, and diagnostic systems. Two trends of the past few decades are continuing; relatively less emphasis on training in projective methods (Belter & Piotrowski, 1999) and increased use of rating scales (Archer & Newsom, 2000). In fact, a veritable explosion in the creation and publication of behavior rating scales alone necessitated creation of this, our third edition of this text.

Chapter Summary

1. Personality is typically considered to be composed of traits, a more enduring set of characteristics of the individual.
2. Formal personality measures emerged as a logical outgrowth of other efforts to measure individual differences, most notably the experimental methods of Wundt, Galton, and others.
3. The Woodworth Personal Data Sheet was published in 1918 as a result of the

surge of interest in testing potential soldiers.

4. The needs for diagnosis created by World War I and World War II provided consider-able impetus for the development of personality tests.
5. A major assumption underlying projective testing is that the use of stimuli that are prone to a variety of interpretations will encourage clients to reveal information that they otherwise would not share in response to direct questioning.
6. The Rorschach test stimuli were originally inkblots placed on paper that was then folded in half.
7. The use of projective techniques with children dates back to the early part of this century, when Florence Goodenough began to study children's human figure drawings.
8. The MMPI was one of the first tests to use an empirical approach for personality test development and used an item selection method called *empirical criterion keying*.
9. The use of informant rating scales for the assessment of child psychopathology traces its roots to the assessment of adult psychopathology in hospital settings.
10. In a 1978 article in *Psychological Bulletin*, Achenbach and Edelbrock introduced the terms *internalizing* and *externalizing* when referring to psychological disorders of childhood.
11. The first edition of the *DSM* (*Diagnostic and Statistical Manual*) appeared in 1952.
12. The most recent edition of the manual, the *DSM-IV-TR*, is based on a more comprehensive research base than any of its predecessors.
13. The 1974 Education of all Handicapped Children's Act, better known as Public Law 94–142, and its reauthorization, the Individuals with Disabilities Education Improvement Act (IDEIA), have mandated special education and related services for children classified as emotionally disturbed.