Chapter One

HISTORY AND DEVELOPMENT

The road that led to the development of the Minnesota Multiphasic Personality Inventory (MMPI) was circuitous, its course passing by several notable psychometric failures and one surprising success in the area of occupational psychology. The achievement of Starke Rosecrans Hathaway, a physiological psychologist, and J. Charnley McKinley, a neuropsychiatrist, rested on 30 years of experience with self-report questionnaires created to discern personality features. The need to evaluate the fitness of military personnel to serve in the Armed Forces during World War I was a major stimulus to the development of personality questionnaires. The best-known example of these was the Woodworth Personal Data Sheet (1920), which incorporated its author's *Psychoneurotic Tendencies*, a scale developed three years earlier in an effort to assess the vulnerability of military personnel to shell shock, battle fatigue, or other emotional instability. The test was a qualified success, presumably, in part, because prospective soldiers who feared combat or otherwise considered themselves in need of evaluation were more than willing to acknowledge their attitudes on the test (Hathaway, 1965).

The *Psychoneurotic Tendencies* scale was adapted as one of the scales of the Bernreuter Personality Inventory (1933), which was to become hugely popular despite telling research failures that appeared almost immediately following its publication. Landis and Katz (1934) found, for example, that its scale for neuroticism was unable to discriminate neurotics from psychotics. Moreover, some of the scale's items were endorsed by members of a normal sample more frequently than by neurotics. The Achilles' heel of the Bernreuter lay in the strategy of its construction. The composition of the test items was guided by rational considerations applied to textbooks or case histories on the assumption that neurotics, for example, would respond to the items as expert opinion suggested they should. The work of Landis and Katz (1934) and others showed that neurotics, among others, could be somewhat perverse in this regard.

The Humm-Wadsworth Temperament Scale (1935) was the first personality questionnaire to use the actual responses of psychiatric patients to determine the direction in which items should be scored and their suitability for

scale development. It was also the first inventory to attempt to model a specific theory of personality, that of psychiatrist Aaron Rosanoff (1921). The Humm-Wadsworth contained 318 items and provided scores for seven scales: *Normal, Hysteroid, Manic, Depressive, Autistic, Paranoid*, and *Epileptoid*. The item assignments for each scale were based on the difference between the item's frequency of endorsement among a group of patients judged high on the trait and a comparison group of normals. For whatever reason, the test didn't catch on in clinical settings, but it became popular for use in industry (Goldberg, 1971). The method of contrasted groups was not discovered by Humm and Wadsworth. It had been known since at least the early 1920s and had been used to conspicuous advantage by E.K. Strong in the development of the Strong Vocational Interest Blank (SVIB; 1927). Strong had built his interest inventory by requiring that the items on his scales discriminate between occupational groups (e.g., lawyers) and "men-in-general."

As Hathaway began to develop the MMPI, he was aware of the criticisms of the Bernreuter and of the rational approach to inventory construction in general. He also knew both of the methods used to develop the SVIB and of the success of this inventory in practice. Finally, Hathaway knew of the Humm-Wadsworth, which figured in his first publication in the field of clinical psychology (1939), and from which he borrowed about one-quarter of his MMPI items.

It might be supposed that the method of contrasted groups appealed only to Hathaway's scientific scruples, but this is far from the case. Hathaway was a tinkerer, an inventor, and a builder of gadgetry and apparatus. He was by temperament a thoroughgoing pragmatist with a deep distrust of theory and an abiding belief in practical experience. The method of contrasted groups provided Hathaway with a practical means of avoiding theory and sidestepping rational or intuitive guidance in the selection of the items for the MMPI clinical scales. Hathaway did not pretend to know how different kinds of patients would respond to his items. The method of contrasted groups allowed him a satisfactory way of finding out: It allowed him to ask them.

There was another advantage to this method that was not wasted on Hathaway. He knew that the previous rationally developed inventories had failed at least in part because their developers' attention had wandered from the central question of *validity*. By relying on the method of contrasted groups, Hathaway required items to surmount a validity hurdle from the outset: Each item had to demonstrate construct relevant variance by concretely discriminating between groups. This was no guarantee that the item would survive subsequent challenges to its validity. For example, it could fail on cross-validation to discriminate the same group of criterion cases from a different group of normal controls; or a similar but separately constituted group of criterion cases from the same group of normal

controls; or in comparisons in which both the criterion cases and normal controls involved previously untested subjects. The advantage of the method was that the items exposed to such subsequent challenges had achieved at least some initial, concrete, discriminative advantage, a claim no rationally derived item could make.

Hathaway intended the MMPI to achieve a "sampling of behavior of significance to the psychiatrist," and this aim determined the range of clinical scales to be developed for the inventory. From an initial pool of more than 1,000 items drawn from psychiatric textbooks, guides for the mental status examination, and previously published tests, Hathaway and McKinley winnowed the pool to 504 items divided into 26 content areas. These included items related to general medical and neurological symptoms, political and social attitudes, affective and cognitive symptoms, and fears and obsessions; items implicating family, educational, and occupational experience; and a set of items to reveal an overly virtuous self-presentation on the inventory. An additional 55 items thought to be related to masculinity-femininity were later added, and 9 items were subsequently deleted to achieve the final pool of 550 items.

The item format chosen was the first-person declarative sentence, written with simplified wording based on contemporary word-frequency tables. Brevity, clarity, and simplicity were occasionally given precedence over grammatical precision. Common English slang and idioms were used, but esoteric or specialized language was avoided. Responses were limited to *True, False*, and *Cannot Say (?)*.

The group of normals selected to contrast with the pathological criterion groups were 724 University of Minnesota hospital and outpatient clinic visitors between the ages of 16 and 65 (hereafter referred to as the "Minnesota Normals"). Each affirmed that he or she was not currently under a physician's care for the treatment of any illness. The demographics of this group of normals corresponded well to 1930 census values for Minnesota regarding age, sex, and marital status (Colligan, Osborne, Swenson, & Offord, 1983). "In 1940, such a Minnesota normal adult was about 35 years old, was married, lived in a small town or rural area, had eight years of general schooling, and worked at a skilled or semi-skilled trade (or was married to a man with such an occupational level)," (Dahlstrom, Welsh, and Dahlstrom, 1972, p. 8).

First published in 1942, the MMPI included scales for *Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychopathic Deviate (Pd), Masculinity-Femininity (Mf), Paranoia (Pa), Psychasthenia (Pt)*, and *Schizophrenia (Sc)*, and was an immediate success. By the end of April 1943, 230 copies had been sold for use in clinics, correctional facilities, corporations, and universities, generating enough revenue to more than cover the costs of initial publication. The Psychological Corporation, which had turned down the test when it was first submitted, became its licensed distributor

in 1943, by which time the ninth and last of the clinical scales, *Hypomania (Ma)*, was ready to be included in the revised *Manual* (Hathaway & McKinley, 1943) and test materials. A manual supplement introducing a booklet form of the test as an alternative to the original box form was published in 1946. This supplement also introduced the *K* scale—which thereafter took its place next to F on the standard profile form—as well as the practice of adding fractions of *K* to suppress some of the scale score variances related to response style. The MMPI arrived in its final form in the 1951 *Manual* with the addition of the *Social Introversion scale (Si)*, the last of the standard clinical scales. Rapid Reference 1.1 summarizes the MMPI-2 Standard Validity and Clinical scales.

≡ Rapid Reference 1.1

Summary of MMPI-2 Standard Validity and Clinical Scales

Validity Scales

L *Lie.* Assesses naive attempts to place oneself in a morally and culturally favorable light by denying moral imperfections.

F *Infrequency.* Assesses the tendency to claim highly unusual attitudes and behaviors as a function of severe psychopathology; a subject's seeking to place himself or herself in an unfavorable light; or a subject's difficulties completing the inventory (e.g., reading problems or random or careless responding).

K *Correction.* Assesses the tendency to control and limit the disclosure of distress, discomfort, and problems relating to others. Fractions of K are added as a correction to Scales 1, 4, 7, 8, and 9 to discourage false-negative/false-positive scores on these scales.

Clinical Scales

1. *Hypochondriasis (Hs).* Measures the tendency to manifest physical symptoms as an expression of emotional discomfort, to be preoccupied with one's health, and to reject nonmedical (i.e., psychological) explanations for such symptoms.

2. *Depression (D).* Measures aspects of symptomatic depression: dysphoria, distress, pessimism, low morale, inhibition, intropunitiveness, physical discomfort and vegetative symptoms, problems in thinking, and social vulnerability.

3. *Hysteria (Hy).* Measures the tendency to develop physical symptoms under stress, to experience pain, and to deny social friction or discord with others.

4. *Psychopathic Deviate (Pd).* Measures alienation, social disinhibition, and the tendency to come into conflict with family, authorities, and others through rebellion, exploitation, misconduct, poorly developed conscience, and the lack of internalized moral standards.

5. *Masculinity-Femininity (Mf)*. Measures broad patterns of interests, activities, attitudes, and sentiments that tend to follow gender stereotypes.

6. *Paranoia (Pa)*. Measures personal/moral rigidity, interpersonal sensitivity, resentment, and ideas of being misunderstood, mistreated, persecuted, or controlled by others, and the tendency to construe the actions, intentions, and motives of others as unfair, degrading, or hostile.

7. *Psychasthenia (Pt)*. Measures the tendency to express stresses through tension, anxiety, apprehensiveness, worry, phobias, obsessions, rumination, compulsions, and fears of losing control, with willful and inflexible efforts to control such symptoms.

8. *Schizophrenia (Sc)*. Measures severe alienation, self-contempt, apathy, cognitive disruption, inertia, feelings of unreality, alien impulses, and motor and sensory impairment.

9. *Hypomania (Ma)*. Measures a rapid and energetic personal tempo, hyperarousal, hyperactivity, stimulation-seeking, euphoria, imperviousness, undercontrol, and rebellious impulses (high scores), versus lethargy, slowness, submissiveness, vulnerability, scrupulousness and, occasionally depression (low scores).

10. *Social Introversion (Si)*. Measures introversion, shyness, social anxiety, social timidity and awkwardness, and social avoidance (high scores), versus extroversion, outgoingness, social comfort and skill, social intrepidity, and social stimulation-seeking (low scores).

The success of the MMPI was measured not only by its sales and distribution but also by the amount of research interest it attracted. By the mid-1950s, the MMPI literature had accumulated 700 references, and by 1965, this number had expanded to more than 2,000. In the next decade the latter number would triple. Over time, the success of the MMPI in the United States drew the interest of psychologists and psychiatrists from countries around the world, eventually culminating in scores of translations for use in dozens of countries.

With the accumulation of empirical correlates made possible by research interest and as a means of facilitating classification and communication about MMPI data, a coding system was devised whereby the clinical scales could be identified by number in order to bypass the implications of the diagnostic constructs implicit in their names, as follows: Hypochondriasis (Hs) came to be represented as Scale 1, Depression (D) as Scale 2, and so on, up to Hypomania (Ma) as Scale 9. This convention enabled a convenient shorthand for describing MMPI profile patterns such that, for example, a profile having its primary elevations on D and

Pt could be coded as a 27/72. This coding system is described more fully in Chapter 3.

With the passage of time, several weaknesses of the MMPI came to be exposed. Some that were present from the beginning had been allowed to go uncorrected; others became evident only gradually, with the accumulation of research, the march of cultural and demographic changes within the United States, and the introduction of previously unforeseen applications. The chief chronic inadequacy of the MMPI stemmed from Hathaway's need to base his test norms on the original sample of 724 normals who had served as the primary nonpathological reference group for the development of the eight *basic* clinical scales (i.e., Scales *Hs-1, D-2, Hy-3, Pd-4, Pa-6, Pt-7, Sc-8*, and *Ma-9*); these basic clinical scales, plus Scales *Mf-5* and *Si-0*, comprise the *standard* clinical scales of the MMPI/MMPI-2. The repeated use of these subjects as contrasts for Hathaway's pathological criterion groups deprived those groups, in a statistical sense, of their "normal" levels of abnormality. Hathaway and McKinley wanted to collect data from a large additional sample of normal subjects on which a proper set of norms could be based, but they were prevented from doing so by a lack of funds. They therefore had to make do with their tainted sample to establish the MMPI norms.

The inadequacy of these original norms was later made strikingly clear (Colligan et al., 1983; Pancoast & Archer, 1989), as the results of research conducted around the country repeatedly found the scores of newly collected normal samples to hover around a *T*-score of 55. There were additional reasons for newly collected normals to score higher than Hathaway's normals. For the most part, the newer subjects were not excluded from samples because they were under a doctor's care at the time of testing, as Hathaway's subjects had been. Another factor was the tendency for later test administrators to discourage the use of the *Cannot Say (?)* response category, thereby increasing the number of responses that could contribute to scale elevations (the original normals had not been discouraged from leaving items unmarked). Perhaps anticipating this consequence of the hypernormal bias of the original normative sample, Hathaway set the optimal boundary for distinguishing nonpathological from pathological elevations at a *T*-score of 70, two standard deviations from the mean. The wisdom of this choice will become evident later when we discuss the MMPI-2.

In the postwar period, and with the passage of the GI Bill, the average educational attainment of the U.S. population began to rise sharply. The original Minnesota Normals had averaged an eighth-grade education; by 1970, the average years of education had soared to 12, and college enrollment levels were swelling. The population was also becoming more ethnically and culturally diverse, and more women were entering the labor force. With the passage of time, the colloquial

language of some of the items (e.g., references to "sleeping powders," "street-cars," "drop-the-handkerchief") had become dated and at risk of becoming obscure. Other items contained grammatical errors that were overdue for correction, and still others contained references to cultural activities that had become less familiar. With the expansion of the MMPI beyond the hospital and clinic to applications in employment screening and the forensic arena came increasing complaints about sexist wording and items dealing with religious matters, eliminatory functioning, and sexual adjustment that were deemed to be intrusive or offensive. Some areas of item content were thought to be underrepresented in an instrument that had already begun to be more frequently applied to the assessment of substance abuse, suicide risk, and treatment planning. By the 1970s, it had become clear that the time for restandardizing the MMPI had come. The copyright holder, the University of Minnesota Press, appointed a committee to undertake this work.

RESTANDARDIZATION LEADING TO THE MMPI-2

In preparation for the restandardization, the committee developed a new form of the MMPI, *MMPI-AX*, containing all of the original MMPI items, less 16 items that repeated earlier items on the original MMPI, plus 154 newly written items. Subjects between the ages of 18 and 84 were recruited by newspaper ads and solicited using directories and mailing lists from Minnesota (21.6% of total sample), North Carolina (18.8%), Ohio (17.3%), Pennsylvania (11.7%), Virginia (9.7%), California (9.4%), and Washington (8.3%). Subsamples of Native Americans from a federal reservation in Washington state (2.2%) and military personnel on active duty from several U.S. bases (0.92%) completed the restandardization sample. Unlike the Minnesota Normals, restandardization subjects were not disqualified for being under the care of a physician or mental health professional. Subjects were paid and were required to provide basic demographic data on sex, age, ethnicity, attained education, marital status, and income, and to complete a Recent Life Events Survey (*LES*; Holmes & Rahe, 1967), in addition to the MMPI-AX. A small proportion of the sample (111 women and 82 men) was retested an average of eight and a half days later to provide preliminary data on temporal stability. Marital couples or unmarried partners (832 women, 823 men) sharing a household for at least one year completed the Spanier Dyadic Adjustment Scale (Spanier, 1976) and provided ratings on their spouses or partners using a modified form of the Katz Adjustment Scales (Katz, 1968). The final restandardization sample of 2,600 (1,462 women, 1,138 men) remained from a larger sample of about 2,900 subjects, with the data from about 300 subjects removed because of omitted demographic or *LES* information, excessive (> 39)

item omissions on the MMPI-AX, or excessively (> 19) deviant scores on the F scale or on an experimental scale that was devised to detect infrequent responses to items on the second half of the test.

The restandardization sample was made to conform as closely as possible to 1980 census data, excluding geographic distribution. In terms of marital status, income distribution, and ethnic diversity, this goal was largely met, but with some underrepresentation of Hispanics and Asian Americans. There was also some underrepresentation of subjects at the extremes of the age distribution, particularly for younger (< 20) men and older (> 70) women. However, the sample grossly exceeded census estimates for educational level and occupational status. Schinka and LaLone (1997) drew a subsample of 1,000 subjects from the restandardization sample stratified in accordance with census projections and 1995 educational statistics for age, gender, ethnicity, and educational attainment. They found that differences on the standard validity and clinical scales, the content scales, and the supplementary scales between the full restandardization sample and their census-matched subsample were virtually nonexistent. Thus, there do not appear to be large sources of systematic bias in the restandardization sample that would lead to significant errors of measurement or interpretation in most situations. This does not mean that the MMPI-2 can be assumed to be free of biasing factors in all situations. Mere correspondence to census values cannot guarantee that the MMPI-2 will provide a reliable normative standard against which the profiles of atypical samples can be interpreted. For example, the profiles of poor, inner-city African Americans, Native Americans in most parts of the United States, itinerant Hispanic laborers, or even homeless or geographically isolated Caucasians may deviate significantly from the restandardization norms in ways that may lead to overpathologizing or underpathologizing the interpretations derived from them.

The restandardization project culminated in the publication of the MMPI-2 in 1989 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Rapid Reference 1.2 provides publication information.

CAUTION

Norms for the MMPI-2 (or any other such instrument) can never be assumed to be equally applicable for all examinees. For example, the normative reference group may be a poor fit for recent immigrants, residents of poverty-impacted inner-city neighborhoods, members of marginal religious organizations/cults or fringe sociopolitical organizations like white supremacy groups, patients awaiting medical treatment (e.g., transplant surgery) for life-threatening conditions, criminal suspects under prosecution, and so forth. Consideration of subcultural membership and differences can be important for test interpretation.

≡ *Rapid Reference 1.2*

Minnesota Multiphasic Personality Inventory–2 (MMPI-2)

Authors: Starke R. Hathaway, PhD, and J. Charnley McKinley, MD. Restandardized by James N. Butcher, PhD, W. Grant Dahlstrom, PhD, John R. Graham, PhD, Auke Tellegen, PhD, and Beverly Kaemmer

Publication date: 1989

Copyright holder: University of Minnesota Press

What the test measures: Psychopathology and normal/abnormal personality functioning

Age range: 18 years and above

Administration time: 1–2 hours

Norms: Norms for the copyright holder-approved MMPI-2 scales are available in the two editions of the *MMPI-2 Manual* of 1989 (includes the Obvious & Subtle subscales and PS not included in the 2001 *Manual*) and its 2001 revision which contains norms for the *PSY-5* and content component scales, not included in the 1989 *Manual*. More comprehensive collections of approved and as yet unapproved scale norms, such as the *Mf* and *S* subscales, the Wiggins content scales, and many others are contained in Greene, 2000, and/or Friedman, Lewak, Nichols, and Webb, 2001.

Qualifications of examiners: Graduate-level training in psychodiagnostic assessment. Purchase of MMPI materials requires an "A Level" qualification (licensed mental health professionals).

Publisher and distributor: Pearson Assessments
19500 Bulverde Road
San Antonio, TX 78259-3701
Phone: 800-232-1223

Products and services: Manual, test materials, and scoring and/or interpretation services, including on-site scoring, are available from Pearson Assessments. Alternate computer scoring, with or without interpretation, is available from Caldwell Report, 5839 Green Valley Circle, Suite 203, Culver City, CA 90230, phone 877-667-4248, and from Behaviordata, 20833 Stevens Creek Boulevard, Suite 100, Cupertino, CA 95014, phone 800-627-2673.

SIMILARITIES AND DIFFERENCES BETWEEN THE MMPI AND MMPI-2

Apart from the deletion of a few items that previous test-takers had identified as objectionable (mostly items with religious, sexual, bowel, or bladder content),

from scales *F* (3 items), *Hs* (1), *D* (3), *Mf* (4), and *Si* (1), the standard validity and clinical scales of the MMPI are unchanged in the MMPI-2, assuring substantial continuity with the research and clinical literature built up over the previous half-century. The revised instrument contains 567 items, of which none are repeated (vs. the MMPI's 566 items, of which 16 were repeated). Ninety items, 15 with religious content, were dropped from the original MMPI item pool; 107 were new to the MMPI-2. With very few exceptions, any scale developed for the MMPI can be adapted for use with the MMPI-2. Sixty-eight of the items retained from the MMPI were rewritten to correct grammar, eliminate sexist language, or reduce ambiguity. None of the changes materially affect the performance of these items (Ben-Porath & Butcher, 1989). The scale level factor structure of the MMPI-2 is essentially identical to that of the MMPI. Although short forms of the MMPI/MMPI-2 render a considerable amount of test information unavailable to the interpreter (e.g., content scale scores), the completion of the first 370 items enables the full scoring of L, F, K, all of the standard clinical scales and Harris-Lingoes subscales, and PK.

The Harris-Lingoes subscales for six of the eight basic scales were for the most part unchanged for the MMPI-2, although the subscales for Scale *4 (Pd)*, which had originally included 14 items not on *Pd* itself, 6 of these on Pd3, no longer include these items. The 13 Wiggins (1966) content scales for the MMPI were replaced by 15 MMPI-2 content scales (Butcher, Graham, Williams, & Ben-Porath, 1990). The Koss-Butcher (Koss & Butcher, 1973) and Lachar-Wrobel (Lachar & Wrobel, 1979) critical items were retained, with two of the Koss-Butcher item sets (those dealing with depression/suicide and alcohol abuse) augmented by some of the new MMPI-2 items. Several new scales were developed for the MMPI-2, including five validity or response-style indicators, three subscales for Scale *0 (Si)* to replace the earlier six Serkownek (1975) subscales, two gender-role scales (GM & GF), two posttraumatic stress disorder (PTSD) scales (PK & PS), two alcohol/substance abuse scales (AAS & APS, and a revision of MAC-R), eight Martin-Finn subscales for Scale *5 (Mf)* to replace the six Serkownek (1975) *Mf* subscales, the Marital Distress Scale (MDS), and the Personality Psychopathology–Five (PSY-5) scales. All of these scales are discussed in detail in Chapters 6 and 7.

Scores for all of the scales of the MMPI were transformed onto the T distribution to enable comparison between scales with different numbers of items and with different means and variances. Because each scale had its own characteristic skewness and kurtosis, the standard linear *T*-score formula did not permit a direct comparison of percentile ranks. For the MMPI-2, the distributional characteristics of the basic clinical and content

scales are adjusted to enable MMPI-2 *T*-scores to be represented as percentile equivalents.

The new uniform *T*-score distributions were designed by first creating a composite distribution for each scale set and then mapping each scale in each set (basic and content) onto its own composite distribution (Tellegen & Ben-Porath, 1992). In this way, percentile-rank uniformity among scales could be achieved without distorting the characteristic positive skew of these scales. Following publication of the MMPI-2, the uniform *T*-distribution was extended to include the content component (Ben-Porath & Sherwood, 1993; Green, Handel, & Archer, 2006) and *PSY-5* (Harkness, McNulty, & Ben-Porath, 1995; Bagby, Ryder, Ben-Dat, Bacchiochi, & Parker, 2002) scales.

Finally, unlike the original normals, the restandardization subjects were actively discouraged from leaving items unmarked, so that their average *Cannot Say (?)* score was lower than that of the Minnesota Normals, thereby reducing the distorting influence of omitted items on MMPI-2 mean scale scores. Hathaway had set a fifth-grade reading level as the minimum competency for taking the MMPI; subsequent research on reading skills showed that the difficulty of many of the items was well beyond this level of reading competency. As a result of studies on reading difficulty of items carried out as a part of the MMPI-2 restandardization, the authors set an eighth-grade level of reading proficiency as the new minimum.

ESSENTIAL REFERENCES FOR THE MMPI-2

The MMPI-2: Manual for Administration and Scoring (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) is the most basic reference, containing the publisher's guidelines for use and extensive psychometric information about the test. Many important references on the test, including the original articles on scale development, are collected in *Basic Sources on the MMPI-2* (Butcher, 2000). Among several guides and manuals for the MMPI-2, the most comprehensive are *The MMPI-2/MMPI-2-RF: An Interpretive Manual* (Greene, 2011a) and *Psychological Assessment with the MMPI-2* (Friedman et al, 2001).

MORE RECENT DEVELOPMENTS

Students and newcomers to the MMPI-2 will find the following sections recondite if not inaccessible. For now, they may be safely skipped pending greater familiarity/mastery of the material presented in Chapters 2 through 10.

MMPI/MMPI-2 Personality Disorder Scales

The first set of scales devised for the assessment of personality disorders was developed for the original MMPI by Morey, Waugh, and Blashfield (1985). These were then followed by two additional sets of personality disorder scales by Levitt and Gotts (1995; see also Gotts & Knudsen, 2005), then by Somwaru and Ben-Porath (1995) for the MMPI-2. Most of these scales have demonstrated acceptable psychometric characteristics (e.g., temporal stability, convergent and discriminant validity), but none are available through commercial scoring services at present and will not be further described or discussed here. Interested readers are referred to the convenient review by Widiger and Boyd (2009), the references included therein, and several additional reports included among the references here (Castlebury, Hilsenroth, Handler, & Durham, 1997; Guthrie & Mobley, 1994; Hurt, Clarkin, & Morey, 1990; Lenzenweger & Korfine, 1992; McCann, 1992; Miller, Streiner, & Parkinson, 1992; Pincus & Wiggins, 1990; Scheidt & Windle, 1994; Sinha & Watson, 1999; Trull, 1991; Wagner, Riley, Schmidt, McCormick, & Butler, 1999; Wiggins & Pincus, 1989; Wise, 1994, 1995, 2002; and Woolley, 2004).

PSY-5 Facet Subscales

Seeking to replicate the clinical utility of sub- or component scales for the MMPI-2 clinical and content scales, respectively, for the *PSY-5* scales, Arnau, Handel, & Archer (2005) used principal components analysis to devise 13 facet subscales for them. Although mixed, subsequent evaluations of the reliability and validity of these facet scales have been, on balance, discouraging (Jones, 2008; Quilty & Bagby, 2007; Wang, Zhang, Shi, Zhou, & Li, 2010), but they are nevertheless helpful for understanding the varieties of item content comprising the *PSY-5* scales.

The Restructured Clinical (RC) Scales

One of the problems unanticipated by Hathaway and McKinley in the construction of the MMPI that became evident over the course of ensuing decades of research was the extensive covariation among the clinical scales. In clinical samples the average intercorrelation among the basic clinical scales is roughly in the .55–.60 range. In order to address this problem, Auke Tellegen began a program of investigation in the 1990s that culminated in a new set of nine Restructured Clinical (RC) Scales (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003). These new scales represent a substantial departure from the empirical traditions

of the MMPI and from the familiar clinical scales that have formed its foundation. The first of these scales, *Demoralization (RCd)*, was conceived as a measure of the broad dimension of general maladjustment or subjective distress, the "First Factor" of the MMPI-2 item pool, analogous to Welsh's (1956) *A* scale (described in Chapters 4 and 7). The remaining eight scales, RC1–RC4 and RC6–RC9, were intended to capture "the distinctive substantive core" of each of the clinical scales, 1–4 and 6–9.

It is beyond the scope of this book to review the methodology described in Tellegen et al. (2003) for developing the RC scales, but readers should note that these matters have met with extensive and detailed criticism in Nichols (2006a, 2006b), and in Ranson, Nichols, Rouse, and Harrington (2009). There is reason to doubt, however, that seeking any supposed core for the clinical scales is desirable, even if possible. And Greene's (2011b) finding that *each* of the RC scales reliably decomposes into two or three factors indicates that the procedures followed by Tellegen et al. failed to achieve their goal of unidimensional constructs. However, their effort to create a set of scales that demonstrate reduced covariation as compared with the clinical scales was at least nominally successful. Across the five clinical samples reported in Tellegen et al., the average intercorrelation among the clinical scales is .54, whereas that among their RC counterparts is .41, a 24% increase in scale independence. Unfortunately, the average intercorrelation among the RC scales and their clinical scale counterparts is only .70, a 29% decrease in the fidelity of the RC version to its parent clinical scale. Such a drop is easily understood when it is realized that fewer than half of the items on the RC scales derive from their clinical scale precursors.

Nevertheless, with some notable exceptions (Binford & Liljequist, 2008; Wolf, Miller, Orazem, Weierich, Castillo, Milford, & Keane, 2008). subsequent research has shown that, at least in general, the RC scales have performed reasonably well in predicting external criteria when such criteria and comparison scales have not been unduly selective (see, e.g., Sellbom, Graham, & Schenk, 2006, and the comment thereupon by Greene, Rouse, Butcher, Nichols, & Williams, 2009), and have often performed better than their clinical scale parents against usually unidimensional external criteria. However, the latter finding needs to be understood in the context of the differences in the structure of the RC scales as compared with that of the clinical scales. Whereas most of the clinical scales are heterogeneous in item content, Scales 1 and 7 being the exceptions, most of the RC scales are, like the MMPI-2 content scales, more homogeneous, with *RC4* and *RC9* the exceptions (Cheng, 2008; Depaoli & Meyers, 2007; Nichols, 2009b). Save for only *RC9*, the RC scales are more highly correlated with—indeed are redundant with—various MMPI-2 content-based scales (Greene et al., 2009;

Rouse, Greene, Butcher, Nichols, & Williams, 2008; Tellegen & Ben-Porath, 2008, pp. 339–365; Tellegen, Ben-Porath, & Sellbom, 2009) than they are with their corresponding clinical scales.

On the average, the RC scales are more highly correlated with the content scales as a group among the restandardization men (.392) and women (.414) than they are with the clinical scales for the same men (.351) and women (.355; Tellegen & Ben-Porath, 2008, pp. 315–328). By comparison, the clinical scales are less highly correlated with the content scales among these men (.358) and women (.385; Butcher, Graham, Williams, & Ben-Porath, 1990, p. 61) than are the RC scales. Although these differences are small in absolute terms, they suggest that, if anything, the RC scales were "restructured" away from the clinical scales and toward content-based scales.

The RC scales have been controversial almost from their original publication and have stimulated several research studies and critical evaluations questioning their basis and the methods used in their development (Butcher, Hamilton, Rouse, & Cumella, 2006; Butcher & Williams, 2009; Caldwell, 2006; Gordon, 2006; Nichols, 2006a, 2006b; Ranson, Nichols, Rouse, & Harrington, 2009; Rogers & Sewell, 2006; Rogers, Sewell, Harrison, & Jordan, 2006). The vast majority of the supportive research has come from the authors of the RC scales and their students or colleagues.

One area requiring more investigation involves the comparison of the levels of elevation between the clinical and RC scales among known groups of participants. Although the trends in the research literature are not yet uniform, there are indications that in many contexts the RC scales elevate less readily, that is to say are less sensitive, or more prone to false negatives, than are the clinical scales (Binford & Liljequist, 2008; Cumella, Kally, & Butcher, 2009; Gucker, Kreuch, & Butcher, 2009; Henry, Heilbronner, Mittenberg, Enders, & Domboski, 2009; Megargee, 2006; Pizitz & McCullaugh, 2011; Rogers, Sewell, Harrison, & Jordan, 2006; Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006; Wallace & Liljequist, 2005; Weiss, Bell, & Weiss, 2010; but see also Osberg, Haseley, & Kamas, 2008). The data reported by Megargee and by Pizitz and McCullaugh are particularly concerning in this respect. In a large (> 2,000) sample of incarcerated felons, Megargee found that their mean scores on the RC scales were, on the average, lower than the mean RC scores of the MMPI-2 restandardization sample, and all were below a *T*-score of 56, including *RC4*, a scale one would expect to be significantly elevated among prisoners. Pizitz and McCullaugh, in a sample of convicted male stalkers, found that five of the RC scales (*RC2, RC3, RC7, RC8*, and *RC9*) showed a mean *T*-score below 50, and that the mean *T*-score for *RC4*, a scale that one would expect to be elevated in such a sample, was only 51.7, more

than a standard deviation below that for Scale 4. Converting the MMPI-2 to the MMPI-2-RF form (see following section), these investigators found that of the 42 substantive (i.e., nonvalidity) scales on this form, only eight achieved mean scores greater than *T*-50, the highest of these being on Mechanical-Physical Interests (MEC), at a *T*-score of 57, for this all-male criminal sample.

The problem of false negatives is familiar to the student of the MMPI (see the discussion under *K* in Chapter 5). Just as some of the clinical scales failed to elevate when it seemed they should have (i.e., when the examinee was manifestly disordered) before the adoption of the *K*-correction, so it may be that some, perhaps all, of the RC scales are similarly afflicted, and possibly more so. Homogeneous scales with obvious item content, of which the RC (and RF; see following section) and content scales are examples, tend to be somewhat more vulnerable to both under- and over-reporting than are the more complex, multivariate clinical scales. Although the extent and regularity of such vulnerability for the RC scales is as yet not fully understood, the trend of the evidence accumulated so far suggests that they are likely to underperform the clinical scales in assessment contexts in which sensitivity to psychopathology is an especially desirable characteristic, such as in employment screening.

To be sure, the authors of the RC scales have defended them vigorously (see Tellegen, Ben-Porath, Sellbom, Arbisi, & Graham, 2006), though not always responsively. To take only one example, they extensively criticized the sample of 26,118 male and 26,425 female protocols upon which this author's (Nichols, 2006a) analyses were based, but replicated virtually none of these on their own, presumably more adequate samples, in their response. Thanks to the publication of Tellegen and Ben-Porath (2008), 56 of the correlations between previously extant scales and the RC scales are reported for five of their clinical samples. The median difference in value between these correlations and those reported by Nichols (2006a; Table 4) is negligible and, for 39 of them, their own values are actually identical with, or higher than, those of Nichols for at least one of their samples. This example suggests that had these authors chosen to replicate Nichols analyses on their own samples, their findings would have been in substantial accord with his (see also Greene, 2011a, pp. 405–406).

THE MMPI-2-RF (Restructured Form)

The most recent MMPI development was the release of an abbreviated and restructured form, the MMPI-2-RF, in 2008. This new form consists of 338 MMPI-2 items divided among 50 scales (a categorical listing of the 229 items not included in the MMPI-2-RF is available in Butcher, 2011). The clinical scales have been abandoned in the RF form and replaced with the RC scales.

DON'T FORGET

Upon the presently available weight of the evidence, the RC scales, like the MMPI-2 content scales, the substantive MMPI-2-RF scales, and other content-based measures lack sensitivity and therefore may underestimate the presence, significance, and magnitude of clinical problems, a liability that may render them less than suitable for screening purposes (e.g., employment screening) for which maximal sensitivity is desirable.

Abbreviated versions of the *PSY-5* scales are retained. Among the remaining scales are (1) abbreviated or otherwise modified versions of *VRIN, TRIN, F, Fp*, the Symptom Validity Scale (FBS; formerly the Fake Bad Scale), *L* and *K*, plus a new validity measure, Infrequent Somatic Responses (Fs); (2) three Higher-Orderscales:Emotional/Internalizing Dysfunction (EID), Thought Dysfunction (THD), and Behavior/Externalizing Dysfunction (essentially sets of items to model the 2-7/7-2 codetypes or *RCd, RC2*, and *RC7*, the 6-8/8-6 codetypes or *RC6* and *RC8*, and the 4-9/9-4 codetypes or *RC4* and *RC9*, respectively); (3) five Somatic/Cognitive scales; (4) nine Internalizing scales; (5) four Externalizing scales; (6) five Interpersonal scales; and (7) two Interest scales. Consistent with its abbreviated design, more than half of the 50 RF scales have ten or fewer items: 5 scales with ten items, 5 with nine, 2 with eight, 5 with seven, 3 with six, 4 with five, and 1 with four. The norms for all of the RF scales are based on the MMPI-2 restandardization sample.

Because the RF form includes no new items, all of the scales thereon may be scored within the MMPI-2 and, for the same reason and with rare exceptions, all of the scales on the MMPI-2-RF are redundant with scales already available on the MMPI-2 (see Tellegen & Ben-Porath, 2008, pp. 33–37).

Several problems and criticisms have emerged following the release of the MMPI-2-RF. Although space does not allow their adequate discussion here, some of the problems can be briefly highlighted, as follows:

1. The justification and adequacy of the theory recruited for the approach taken to construct the central scales of the RF form, the RC scales (Ranson et al., 2009)
2. The limited sensitivity of the scales on the new form for the assessment of clinical problems (Binford & Liljequist, 2008; McCullaugh et al., 2009; Megargee, 2006; Rogers et al., 2006; Sellbom et al., 2006; Wallace & Liljequist, 2005), as described in the previous section
3. Low internal consistency reliabilities for some of the RF scales (Butcher, 2011)

4. The absence of or confusing and/or misleading empirical correlates for many of the RC scales (Butcher, 2011; Greene, 2011a), raising questions about their construct validity

5. Combining the norms of men and women by the adoption of nongendered norms, thereby affecting scales' sensitivity to some clinical problems, e.g., depression (Hathaway & McKinley, 1942/2000), and contributing to gender bias (Dean, et al., 2008; Nichols, Greene, & Williams, 2009).

6. The loss of information from the deleted items, particularly in areas that may be related to work adjustment, family problems, antisocial attitudes, mood, and interests—topics that may be important in employment screening, child custody, and forensic evaluations

The MMPI-2-RF has been heavily promoted by its authors, publisher, and licensed distributor through advertising, Continuing Education workshops, and related means. Thanks to these efforts, it has established a foothold in personality assessment practice in a relatively short time. Moreover, there is no reason to suppose, given its pedigree, that generally favorable research findings as regards its validity and performance will not be forthcoming, at least within some contexts of use. The MMPI-2-RF is almost entirely a content-based instrument, and in most contexts content-based scales have a history of performing well. However, users will do well to bear in mind the vulnerability of content-based measures to impression management and to the false-negative problem described previously in connection with the RC scales. An additional concern relates to the length of many of the RF scales, just as it does to some of the shorter scales of the MMPI-2. As Emons, Sijtsma, and Meijer (2007) have shown, when used for classification purposes, scales of 15 or 12 or fewer items are not only less reliable than 20- or 40-item versions of the same scale, but they are also significantly more vulnerable to misclassification. And, of course, there is nothing to prevent these two vulnerabilities—content obviousness and short scale length—from operating synergistically. Hence, an added element of caution is indicated when interpreting many, perhaps the majority, of MMPI-2-RF scores.

To be sure, there are certainly cases/situations for which an abbreviated form of the MMPI-2 can be a convenience. Examinees with attentional or concentration deficits, dementia, traumatic brain injury, and similar disabilities may find the full MMPI-2 overly taxing. Even in these cases, however, the MMPI-2-RF may possess few if any advantages over the standard MMPI-2 370-item short form considering the extensive literature on numerous clinical scale pattern types from multiple psychiatric and medical settings, including numerous neurological and neuropsychological samples (see Arbisi & Seime, 2006; Gass, 2006, 2009, for recent reviews).

✍ TEST YOURSELF ✍

1. **The primary developer of the MMPI, Starke Hathaway, was a**
   (a) clinical psychologist.
   (b) physiological psychologist.
   (c) social psychologist.
   (d) psychoanalyst.

2. **The test whose purpose and development were most similar to those of the MMPI was**
   (a) the Woodworth Personal Data Sheet.
   (b) the Bernreuter Personality Inventory.
   (c) the Humm-Wadsworth Temperament Scale.
   (d) the Strong Vocational Interest Blank.

3. **The method of contrasted groups used in the development of the MMPI was the invention of Hathaway and McKinley.** True or False?

4. **The success of the MMPI following its initial publication was virtually immediate.** True or False?

5. **The last of the original 13 validity and clinical scales to be added to the MMPI was**
   (a) Scale *K*.
   (b) Scale *5* (*Mf*).
   (c) Scale *9* (*Ma*).
   (d) Scale *0* (*Si*).

6. **The MMPI-2 restandardization sample tended to match 1980 census values for**
   (a) marital status but not educational level.
   (b) educational level but not income distribution.
   (c) income distribution but not marital status.
   (d) educational level but not ethnic diversity.
   (e) b and d.

7. **Uniform *T*-scores were designed to correct for the characteristic positive skew of the basic clinical scales.** True or False?

Answers: 1. b; 2. c; 3. False; 4. True; 5. d; 6. a; 7. False