5-point scale, how much they have been distressed by each of the problems during the past seven days. The SCL-90-R provides separate norms for male and female adult and adolescent nonpatients, as well as for psychiatric inpatients and outpatients. However, some of the norms are not sufficiently representative; for example, the psychiatric inpatient normative group was predominantly of lower socioeconomic status, and the adolescent group was mostly middle class and almost totally White.

The SCL-90-R items resemble those of earlier checklist-type instruments not only in that they were selected on the basis of content relevance and clinical utility, but also in that some of them can be traced back to the Woodworth Personal Data Sheet through intermediate scales like the Hopkins Symptom Checklist and the Cornell Medical Index (Derogatis & Lazarus, 1994). The items are organized into nine dimensions of psychopathology, namely, Somatization, Depression, Anxiety, Hostility, Psychoticism, Interpersonal Sensitivity, Phobic Anxiety, Paranoid Ideation, and Obsessive-Compulsive symptoms. Factor analytic studies of these scales suggest that they are intercorrelated and thus not very useful in differential diagnosis; nevertheless, the global indices derived from the Checklist have proved to be reliable indicators of the presence and severity of psychopathology (Payne, 1985). The SCL-90-R and related instruments, such as the Brief Symptom Inventory, can be used most appropriately as part of a battery, in the evaluation of change through a course of therapy, and in research on the outcome of various treatments.

The principal advantage of the content-related approach to personality inventory development lies in the simplicity and directness of the method. While these features make it possible to have relatively brief and economical instruments, their transparency also affords examinees a greater opportunity for conscious attempts at manipulation of the results than do other methods. Content-based instruments typically do not have the features designed to prevent or detect response biases that are discussed later in this chapter (Bornstein, Rossner, Hill, & Stepanian, 1994). For this reason, exclusive reliance on their results as a basis for any decision is not recommended.

**EMPIRICAL CRITERION KEYING**

**Basic Approach.** *Empirical criterion keying* refers to the development of a scoring key in terms of some external criterion. The procedure involves the selection of items to be retained and the assignment of scoring weights to each response. In the construction of the previously cited Woodworth Personal Data Sheet, some of the statistical checks applied in the final selection of items pointed the way for criterion keying. Thus, no item was retained in this inventory if 25% or more of a normal sample answered it in the unfavorable direction. The rationale underlying this procedure was that a behavior characteristic that occurs with such frequency in an essentially normal sample cannot be indicative of
abnormality. The method of contrasted groups was likewise employed in the selection of items. Only symptoms reported at least twice as often in a previously diagnosed psychoneurotic group than in a normal group were retained.

Despite some use of such empirical checks, however, content-related approaches rely essentially on a literal or veridical interpretation of questionnaire items. The response to each question is regarded as an index of the actual presence or absence of the specific problem, belief, or behavior described by the question. In empirical criterion keying, on the other hand, the responses are treated as diagnostic or symptomatic of the criterion behavior with which they were found to be associated. In an early description of this approach, Meehl (1945) wrote:

... the verbal type of personality inventory is not most fruitfully seen as a "self-rating" or self-description whose value requires the assumption of accuracy on the part of the testee in his observations of self. Rather is the response to a test item taken as an intrinsically interesting segment of verbal behavior, knowledge regarding which may be of more value than any knowledge of the "factual" material about which the item superficially purports to inquire. Thus if a hypochondriac says that he had "many headaches" the fact of interest is that he says this. (p. 9)

A self-report inventory is indubitably a series of standardized verbal stimuli. When criterion-keying procedures have been followed, the responses elicited by these stimuli are scored in terms of their empirically established behavior correlates. They are thus treated like any other psychological test responses. That questionnaire responses may correspond to the person's perception of reality does not alter this situation. It merely provides one hypothesis to account for the empirically established validity of certain items.

The Minnesota Multiphasic Personality Inventories

The outstanding example of empirical criterion keying in personality test construction is the Minnesota Multiphasic Personality Inventory (MMPI). In recent years, the MMPI has been revised and reconstituted into two separate versions, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the MMPI-Adolescent (MMPI-A—Butcher et al., 1992). In spite of the existence of these newer versions, no discussion of either can proceed without reference to the original MMPI and the role it played in the history of personality assessment. Although a thorough description of it is beyond the scope of this text, it must be noted that for a period of almost half a century, the MMPI was the most widely used and the most thoroughly researched personality test.²

²For a fairly concise description of the original MMPI, see earlier editions of this text (e.g., Anastasi, 1988b). More extensive treatments of this instrument can be found in the classic volumes by Dahlstrom, Welsh, and Dahlstrom (1972, 1975).
In many ways the MMPI, as an instrument, was a victim of its own success. Conceived in the 1930s by Starke R. Hathaway, a clinical psychologist, and J. Charnley McKinley, a neuropsychiatrist, it was initially published through a series of articles in the 1940s to serve as an aid in the process of psychiatric diagnosis. Thereafter, its effectiveness in detecting psychopathology and differentiating between the then rather crude nosological categories led to its being used for an ever-expanding set of purposes, beyond its original aim. By the 1960s, the MMPI was firmly entrenched as the leading personality test and used as frequently, or more, with normal subjects in counseling, employment, medical, military, and forensic settings, as with psychiatric patients. By the 1980s, the MMPI literature numbered several thousand references documenting, among many other things, the use of its 13 basic scales with a large variety of populations, the development of hundreds of special scales based on its items, and a vast array of empirical correlates of scale elevations and profile patterns. However, by that time, its well-documented conceptual and psychometric problems appeared more troublesome in light of advances in the fields of psychopathology and personality theory, as well as in test construction. Furthermore, by then it had also been clearly demonstrated that the narrowly based and outdated norms of the test were inappropriate for contemporary examinees, so that the very basis for determining abnormality rested on an uncertain foundation (Colligan, Osborne, Swenson, & Offord, 1983, 1989). In effect, the original standardization sample had become something akin to a nonnormative reference group, in terms of which the score scale was defined. The much more extensive data subsequently collected with reference to profile codes then provided the basis for normative interpretation.

Thus, the committee charged with restandardizing the MMPI faced the difficult task of modernizing the instrument while trying to save the wealth of interpretive material relevant to the assessment of personality and, especially, of psychopathology embedded within the MMPI's basic structure. For the sake of continuity, the committee chose to retain the vast majority of the items, all of the original clinical and validity scales, and many of the supplementary scales, along with their built-in weaknesses. The major changes were: a complete reformatting of the inventory; the development of uniform T scores for the eight original clinical scales and all of the content scales; the revision and deletion of outdated or otherwise objectionable items, as well as the addition of new ones; the creation of several new validity, supplementary, and content scales; and the separation of the inventory into two versions, suitable for different age groups.

The Minnesota Multiphasic Personality Inventory-2. The MMPI-2 items consist of 567 affirmative statements to which the test taker gives the responses "True" or "False." The first 370 items, which are virtually identical to those in the MMPI except for editorial changes and reordering, provide all the responses needed to score the original 10 "clinical" and three "validity" scales. The remaining 197 items (107 of which are new) are needed to score

---

3 The original articles have been reproduced in Dahlstrom and Dahlstrom (1980).
the full complement of 104 new, revised, and retained validity, content, and supplementary scales and subscales that make up the complete inventory. The items range widely in content, covering such areas as general health; affective, neurological, and motor symptoms; sexual, political, and social attitudes; educational, occupational, family, and marital questions; and many well-known neurotic or psychotic behavior manifestations, such as obsessive and compulsive states, delusions, hallucinations, ideas of reference, phobias, and sadistic and masochistic trends. Dahlstrom (1993a) has prepared a manual supplement that provides all the information necessary to compare the items of the MMPI-2 with the original ones. A few illustrative items, followed by the numbers they bear in the current form of the test, are shown below:

- My sleep is fitful and disturbed. (39)
- I believe I am being plotted against. (138)
- I am worried about sex. (166)
- When I get bored I like to stir up some excitement. (169)
- Most people inwardly dislike putting themselves out to help other people. (286)

The MMPI-2 provides scores on 10 basic “clinical scales,” which are the same as those in the original MMPI and are listed below:

1. Hs: Hypochondriasis
2. D: Depression
3. Hy: Hysteria
4. Pd: Psychopathic deviate
5. Mf: Masculinity-femininity
6. Pa: Paranoia
7. Pt: Psychasthenia
8. Sc: Schizophrenia
9. Ma: Mania
10. Si: Social introversion

Eight of these scales were developed empirically, in the 1940s, by criterion keying of items that differentiated between small clinical samples, mostly of about 50 persons each, representing the traditional psychiatric diagnosis in use at the time, and a normal control group of 724 visitors and relatives of patients in the University of Minnesota hospitals (Hathaway & McKinley, 1940, 1943). The Masculinity-femininity scale, originally intended to distinguish between homosexual and heterosexual men, was developed from the differences between male soldiers and female airline employees in item-endorsement frequency. Scores on this scale indicate the extent to which a person's interests and attitudes match the stereotype of her or his sex group. The Social introversion scale, added later, was derived from the responses of two contrasted groups of college students selected on the basis of extreme scores on a test of introversion-extraversion.

*From Minnesota Multiphasic Personality Inventory-2. Copyright © by The Regents of the University of Minnesota. All rights reserved. Reproduced by permission.*
In retaining the basic scales of the MMPI intact (except for nine deleted items, a number of editorial changes, and a reordering of the items), the MMPI-2 developers sought to preserve the wealth of clinically useful information associated with the interpretation of profile codes based on patterns of scores on those scales (Graham, 1993; Greene, 1991). However, also retained were the obsolete notions of psychopathology implicit in those scales and the consequences of the naive and flawed application of the empirical method of contrasted criterion group keying. Other problems, such as the multidimensionality and overlap of the basic scales, were kept as well (Helmes & Reddon, 1993).

An outstanding feature of the original MMPI was its use of three so-called validity scales, which have also been retained in the MMPI-2. These scales are not concerned with validity in the technical sense. In effect, they represent checks on carelessness, misunderstanding, malingering, and the operation of special response sets and test-taking attitudes. The validity scores include:

- **Lie Score (L):** based on a group of items that make the respondent appear in a favorable light but are unlikely to be truthfully answered in the favorable direction (e.g., I do not like everyone I know).
- **Infrequency Score (F):** determined from a set of 60 (out of the original 64) items answered in the scored direction by no more than 10% of the MMPI standardization group. Although representing undesirable behavior, these items do not fit any particular pattern of abnormality. Hence, it is unlikely that any one person will actually show all or most of these symptoms. A high F score may indicate scoring errors, carelessness in responding, gross eccentricity, psychotic processes, or deliberate malingering.
- **Correction Score (K):** using still another combination of specially chosen items, this score provides a measure of test-taking attitude believed to be more subtle. A high K score may indicate defensiveness or an attempt to "fake good." A low K score may represent excessive frankness and self-criticism or a deliberate attempt to "fake bad."

The first two scores (L, F) are ordinarily used for an overall evaluation of the test record. If either of these scores exceeds a specified value, the record is considered invalid. The K score, on the other hand, was designed to function as a suppressor variable. It is employed to compute a correction factor that is added to the scores on some clinical scales to obtain adjusted totals. The effectiveness of this use of the K score is questionable; therefore, scores on the affected scales can be reported with and without this correction. Although an unusually high K score would in itself make a record suspect and call for further scrutiny, moderate elevations of the K score may in fact reflect ego strength and a positive adjustment. It is particularly important to evaluate elevations in light of the individual's history and life circumstances.

---

5 There is also a Cannot Say (?) score representing the number of items that were double-marked or omitted. If this count exceeds 30 items, the test record is considered highly suspect and probably invalid.
Among the 21 supplementary scales of the MMPI-2 are three new "validity" indicators that can help to assess the care and veracity with which the test takers respond to the inventory. They are the Back F (Fb) scale, the Variable Response Inconsistency Scale (VRIN), and the True Response Inconsistency Scale (TRIN). Whereas the Fb scale is basically an extension of the original F scale for items that appear in the second half of the inventory, VRIN and TRIN are new scales that consist of pairs of items with similar or opposite meanings and are aimed at detecting inconsistent or contradictory responses.

The basic profile form for the MMPI-2 (Fig. 13-1) includes the 13 validity and clinical scales carried over from the original version. There are also separate profile forms for 15 content scales, 27 content component scales, 21 supplementary scales, and 28 Harris-Lingoes subscales. Some of these scales and subscales are new, and some have been retained from the original; all of them, however, are scored by using the MMPI-2 normative sample of 2,600 adults, aged 16 to 84. This sample is far more representative of the current population of the United States than the original Minnesota normative group, having been collected in seven different states in an attempt to reflect the U.S. population in terms of significant demographic variables, including gender, age, and ethnicity (Dahlstrom & Tellegen, 1993). Nevertheless, the representativeness of the sample has been questioned, primarily because of its high levels of occupational and educational attainment and underrepresentation of Hispanics and Asian Americans, compared with the 1980 census figures (Duckworth, 1991).

The main result of the renorming of the MMPI was a lowering in the score elevation of clinical profiles. This change, which had been widely anticipated, is probably due to generational differences, as well as to special factors unique to the original Minnesota sample and to the way in which it was used in the development of the MMPI (D. S. Nichols, 1992). At any rate, the cutoff T score necessary for considering a scale elevation to be of clinical interest is now 65, or approximately 1.5 SDs above the mean, as opposed to 70. Another innovation introduced in the MMPI-2 is the use of uniform—as opposed to linearly derived or normalized—T scores in 8 of the 10 clinical and all of the content scales. This involved equating the scores on all these scales to an average composite distribution; the uniform T scores allow comparisons across scales in terms of percentile equivalents, without significantly distorting the shape of the raw score distributions which are all positively skewed, albeit to different degrees (Tellegen & Ben-Porath, 1992).

The changes just described were certainly justified on psychometric grounds; however, because these changes result in differences between the profile patterns and codes obtained from the MMPI and the MMPI-2, a great debate has arisen concerning the viability of generalizing the findings from the vast MMPI profile interpretation literature, and from clinical lore, to the MMPI-2 (Chojnacki & Walsh, 1992; Morrison, Edwards, & Weissman, 1994; Tellegen & Ben-Porath, 1993). The data suggest that, for well-defined profiles with clear separation between scale scores, there seems to be about as much congruence between MMPI

---

6As of 1996, the 10 Wiers-Hammer Suicide-Obvious subscales are no longer available from the MMPI-2 publisher.
Figure 13-1. MMPI-2
Basic Scales Profile of a 41-Year-Old-Man. The code summarizes the relationships among the MMPI-2 scale scores and provides a basis for profile interpretation and comparisons.

(The scores in the profile illustration are from the Annotated Sample Report for The Minnesota Repert Adult Clinical System—Revised [University of Minnesota, 1993, pp. 2–7]).
and MMPI-2 code types as between those obtained from repeated administrations of either one of the versions (Archer, 1992b; Graham, 1993). At any rate, the MMPI-2 manual has information that permits users to compare scores generated from the two versions based on the responses to either one. Although this suggestion is not without its problems (see, e.g., Ben-Porath & Tellegen, 1995), it has been endorsed by some as an empirically defensible method for users to negotiate the period of transition between the two versions (see, e.g., Humphrey & Dahlstrom, 1995).

The Minnesota Multiphasic Personality Inventory-Adolescent. The MMPI-A is the new form of the MMPI developed specifically for use with adolescents. It incorporates most features of the MMPI and MMPI-2, including all of the 13 basic scales, but accommodates younger test takers through the reduction of the overall length of the inventory to only 478 items, the inclusion of new items and scales covering areas specifically relevant to them such as school and family problems, and, above all, the provision of age-appropriate norms. The MMPI-A used a normative sample of 1,620 contemporary adolescents between the ages of 14 and 18; a clinical sample of 713 adolescents in the same age range was collected concurrently for use in comparisons and validity studies.

In addition to the basic clinical and validity scales it shares with the MMPI-2, the MMPI-A has its own validity scales (F1 and F2), as well as some content and supplementary scales and subscales that are unique to it and some that are common to both instruments. Although a great deal of research, including norms and conversion tables published by Dahlstrom et al., (1972) and Marks, Seeman, and Haller (1974), supported the use of the MMPI with adolescents, that research does not necessarily apply to the MMPI-A, which is more of a brand new instrument than a revision. As such, its usefulness will have to be determined through the accumulation of research and interpretive materials that started concurrently with its publication (Archer, 1992a; Butcher & Williams, 1992; Williams, Butcher, Ben-Porath, & Graham, 1992).

Concluding Comments on the Minnesota Multiphasic Personality Inventories. In spite of its origin as a prototypical product of a naively applied empiricism, and of repeated rumors about its imminent demise, the MMPI has managed to survive. In fact, although its original purpose was to aid in the process of psychiatric classification, and the procedures followed in its development made it unsuitable for personality assessment in normal individuals, the MMPI has been widely used with normals as well as psychiatric patients. It and its revised versions have incorporated a number of additional procedures and interpretive strategies into the original, empirically derived framework of the inventories. Such features, which have already been mentioned in passing but deserve further notice, include the many scales developed by grouping items on the basis of their content (Butcher, Graham, Williams, & Ben-Porath, 1990) as well as the use of factor analysis in the development of some of the supplementary scales (Welsh, 1956).
New ways of approaching the complex task of MMPI interpretation continue to evolve. One of the most recent is the use of structural summaries to bring some coherence and ease to the use of the multiplicity of intercorrelated scales produced by the Minnesota inventories. The approach, based on analyses of scales, subscales, and items, aims at reducing the number of dimensions necessary to interpret the results of the inventories by cutting across arbitrary classifications—such as “Supplementary,” “Content,” and “Harris and Lingoes” scales. The most salient dimensions are used to organize the categories into a “Structural Summary” format similar to the one developed by Exner for use with the Rorschach (see chap. 15). This approach to MMPI interpretation is at the incipient stages of development and in need of further investigation and validation. Nevertheless, it is already being applied both to the MMPI-2, with structural dimensions derived primarily on the basis of content analysis, and to the MMPI-A, with dimensions derived through factor analysis, and it seems promising (Archer & Krishnamurthy, 1994; Archer, Krishnamurthy, & Jacobson, 1994; Nichols & Greene, 1995).

Two other sets of developments that proceed space with the MMPI-2 and MMPI-A, as with most other tests, are the computerization of procedures for administration, scoring, and interpretation of the inventories and the development of translations of the instruments into many languages. Conventional computer administration and scoring of the inventories, which were available for the MMPI, are also available for the new inventories, as are several automated interpretation services. In addition, a computerized adaptive version of the MMPI-2 has been developed and tried with promising results (Roper, Ben-Porath, & Butcher, 1991, 1995).

Whereas it took almost a decade for the original MMPI to be translated for the first time, work on cross-cultural adaptations of the MMPI-2 began even prior to its publication. In the first three years of its existence, there were 15 translation projects of the MMPI-2 completed or in progress. Six Spanish translations or adaptations, including two versions for use with Hispanics in the United States, have been prepared. A handbook on translations and international adaptations of the MMPI-2 is also available (Butcher, 1996). One noteworthy finding of those who have collected data with the use of such translations and adaptations is that current cross-cultural normal samples score closer to the MMPI-2 standardization norms than earlier such samples did compared to the MMPI norms.

By trying to improve on the classic example of a personality inventory without altering it in any fundamental way, the MMPI Restandardization Committee set for itself two difficult and, to a large extent, contradictory goals. Time will tell whether the decisions the committee made will extend the supremacy of the Minnesota inventories into the next century or whether they will be overtaken by a new generation of similar instruments, such as Jackson's Basic Personality Inventory—discussed in a later section of this chapter—or the Personality Assessment Inventory (PAI) developed by Leslie Morey (1991) using a sophisticated sequential strategy that combined logical and empirical methods to ensure the psychometric soundness of its scales. In the meantime, the rate and range of pub-
lications of books and articles on the MMPI-2 and MMPI-A seems to be continuing unabated (Butcher, 1990; Butcher, Graham, & Ben-Porath, 1995; Keller & Butcher, 1991; Pope, Butcher, & Seelen, 1993).

California Psychological Inventory

Over the years of its existence, the MMPI has served as a basis for the development of other widely used inventories. An outstanding example is the California Psychological Inventory (CPI). While drawing nearly half of its items from the MMPI, the CPI was developed specifically for use with normal adult populations. In its latest revision, the CPI—Third Edition—consists of 434 items to be answered “true” or “false” and yields scores on 20 scales (Gough & Bradley, 1996). Three are “validity” scales designed to assess test-taking attitudes. These scales are designated as: Well-being (Wb), based on responses by normals asked to “fake bad”; Good impression (Gj), based on responses by normals asked to “fake good”; and Communality (Cm), based on a frequency count of highly popular responses. The remaining 17 scales provide scores in such personality dimensions as Dominance, Sociability, Self-acceptance, Responsibility, Socialization, Self-control, Achievement-via-conformance, Achievement-via-independence, Empathy, and Independence. The last two scales were added in the 1987 revision.

For 13 of these 17 scales, items were selected on the basis of contrasted group responses, against such criteria as course grades, social class membership, participation in extracurricular activities, and ratings. The ratings were obtained through peer nominations, which have been found to constitute an effective assessment technique for many interpersonal traits (see chap. 16). For the remaining 4 scales, items were originally grouped subjectively and then checked for internal consistency. Cross-validation of all scales on sizable samples has yielded significant group differences, although the overlapping of contrasted criterion groups is considerable and criterion correlations are often low.

As in the MMPI-2, all CPI scores are reported in terms of a standard score scale with a mean of 50 and an SD of 10; at present, this scale is based on a normative sample of 3,000 females and 3,000 males drawn from the CPI archives so as to represent the general population of the United States in terms of age, socioeconomic level, and geographic area. Norms are provided separately for females, males, and both genders combined. In addition, means and SDs of scores on each scale are given for many special groups.

The CPI, originally published in 1956, was conceptualized as an “open system” from which elements could be removed and to which elements could be added as the need arose (Gough, 1987, p. 1). Accordingly, through the revisions it has undergone, the inventory has been shortened from its original length of 480 items—down to 462 in the 1987 revision and most recently to 434 items—to delete those that might be objectionable to some respondents or legally questionable in light of the Americans with Disabilities Act of 1990 (PL. 101-336), especially in personnel selection settings. Using the extensive archival data available for more than 13,000 subjects on all of the CPI items, Gough and Bradley have striven to