JOURNAL OF PERSONALITY ASSESSMENT, *73*(2), 149–178

Writing a Good Cookbook: I. A Review of MMPI High-Point Code System Studies

Robert E. McGrath and Joel Ingersoll

School of Psychology
Fairleigh Dickinson University

A small set of studies that will be called the high-point code system studies has contributed greatly to the perception of the MMPI as an empirically grounded clinical instrument. Ten published studies have provided data concerning the interpretive significance of high-point codes in general. This article provides a methodological evaluation of these studies. We reviewed the features that distinguish these studies from other MMPI actuarial studies. The high-point code system studies were designed to maximize clinical usefulness, more so than even other actuarial studies, but sometimes did so at a potential cost in power. We address this issue further in a second article on effect sizes derived from the high-point code system studies (McGrath & Ingersoll, this issue). We found striking differences across studies in the code-definition strategies used. Which strategy represents the optimal high-point coding strategy for clinicians who use the MMPI remains an unresolved question because any strategy requires some tradeoff between the ideals of code group homogeneity and inclusiveness. We also address the issue of whether complete description or unique description of code group members is a more desirable goal for such studies and discuss implications of the findings for future research on this topic.

Although now more than 50 years old, the Minnesota Multiphasic Personality Inventory (MMPI) remains the most commonly administered self-report measure (e.g., Piotrowski, 1997).[1] The inventory's enduring popularity is due to two reasons. One is its comprehensive array of well-validated response-style indicators (Baer, Wetter, & Berry, 1992; Berry, Baer, & Harris, 1991; Rogers, Sewell, &

[1]For several reasons, the abbreviation *MMPI–1* is used for the original version (Hathaway & McKinley, 1983), *MMPI–2* for the revised version (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), and *MMPI* for the inventory in general.

Salekin, 1994). The other is an unparalleled body of research demonstrating the validity and clinical utility of the inventory.

Perhaps no set of studies has contributed to the image of the MMPI as an empirically grounded clinical instrument as much as the high-point code system studies.[2] This term refers to studies that examine the clinical implications of high-point codes in general with implications for a general MMPI interpretive system based on the most elevated combination of clinical scales. The studies were initially inspired by Meehl's (1956) seminal piece "Wanted—A Good Cookbook." The article extended his discussion of actuarial versus clinical prediction (Meehl, 1954) to the more global domain of personality description. The first part of the article introduced the concept of a cookbook, an empirically based and mechanical system for identifying clinical descriptors based on patterns in a respondent's test data. This was a general discussion with implications for any psychometric instrument and any set of test patterns.

The remainder of the article was an extended discussion of a doctoral dissertation by C. C. Halbower (1955), representing the first contribution to the development of such a cookbook. The methods used by Halbower in his study provided the methodological foundation for all the studies included in this review.

Halbower (1955) began by classifying Department of Veterans Affairs (VA) outpatients according to the 2 most elevated clinical scales from their MMPI–1 profiles, the vector of *T* scores for the traditional 3 validity scales, and 10 clinical scales. He identified four patterns or codes that represented more than 50% of all patients seen at the facility. To increase homogeneity within the code groups, profiles included in the study had to meet additional rules that were intended to exclude potentially invalid profiles or to ensure a minimum difference between code scales and other clinical scales.

Therapists completed a 154-item 11-step Q-sort of descriptive statements for nine patients from each of the four MMPI code groups. Statements showing little variability had previously been eliminated from the set. Correlations between Q-sort scores were computed for each pair of patients within a code group. The five patients with the highest mean correlations were chosen as the most representative cases to reduce the impact of sampling error and faulty clinical judgments on

[2]These are more commonly called *codetype studies* in the literature, but we prefer our label for several reasons. Dahlstrom (1992) thought references to *codetypes* were confusing because the term can also refer to categories within a classification scheme for profile codes such as Lachar's (1974) division of high-point codes into neurotic, character disordered, and psychotic codetypes. The term *codetype* is also imprecise because any actuarial study involves some sort of profile coding. One individual who commented on the article correctly pointed out that most of the studies on MMPI high-point codes are not included in this review. These are generally more focused studies limited to a single code, target variable, or type of patient. We added the word *system* to clarify that our domain was limited to those studies that treated high-point codes as the basis for a general MMPI interpretive system.

group means. The cookbook description for patients with that test pattern consisted of the average Q-sort scores for the five cases.

The next step involved applying the cookbook descriptions to a second sample of four cases from each code group. Each patient's therapist completed a Q-sort. Four or five clinicians who were unfamiliar with the patient also completed Q-sorts based on the MMPI profile. The most important finding was that therapist Q-sort scores consistently correlated more highly with the cookbook scores than they did with mean Q-sort ratings by clinicians. Where the mean validity coefficient for the clinicians was .48, the mean for the cookbook ratings was .78.

Although the results were striking, the study was limited to four patterns and was never independently published. Further research was needed before a general MMPI cookbook could be produced. In the subsequent years a small number of such studies were published with a dramatic impact on interpretive strategies for the inventory. Most texts describing the clinical use of the MMPI recommend the classification of profiles according to high-point code and base interpretive information for the codes either implicitly or explicitly on the high-point code system studies (e.g., Archer, 1997; Greene, 1991; Groth-Marnat, 1997; Lachar, 1974; Lewak, Marks, & Nelson, 1990).

This article provides a methodological review of the high-point code system studies and addresses implications for future research on interpretive systems for the MMPI. In addition, the information contained herein provides the background for a second article having to do with the size of effects found in these studies (McGrath & Ingersoll, this issue).

This article is divided into four sections. The first lists the common characteristics of the high-point code system studies, the second summarizes each of the studies, the third compares the studies and discusses the methodological issues that emerge from this comparison, and the fourth discusses implications for future MMPI interpretive research.

DELIMITING THE HIGH-POINT CODE SYSTEM STUDIES

Of the various standard MMPI texts providing interpretive information for the high-point codes, we found two that listed the high-point code system studies on which their interpretive narratives were based (Greene, 1991; Lachar, 1974). Using these texts we identified seven of the studies included in this review. We used these studies to identify the key characteristics of a high-point code system study. We found four methodological elements in all seven studies that seemed fundamental to considering them high-point code system studies. First, profiles were classified according to the most elevated combination of clinical scales. Second, predictor variables represented the key variables in the study. Third, multiple high-point codes were investigated. Finally, the studies used diagnostically mixed psychiatric

samples. As the following discussion indicates, each is consistent with the goal of developing a comprehensive and clinically useful interpretive system.

The cardinal trait is of course the classification of MMPI profiles based primarily, although not exclusively, on the most elevated combination of clinical scales. Unlike Halbower (1955), subsequent researchers have then dichotomized cases according to whether they comply with the rule and examined the relation between the resulting binary variables and criterion variables.

There are several ways in which the dichotomization of profiles based on the standard clinical scales is probably less than optimal as a strategy for actuarial description. Compared to a linear combination of the clinical scales, both the configural approach to scale combination (e.g., Goldberg, 1965) and the dichotomization of cases (Cohen, 1983) can result in less powerful prediction of clinical variables. The strategy also excludes supplementary MMPI scales that may enhance prediction over the basic profile scales (Levitt, 1989). For example, the content scales are more homogeneous in item content than the basic scales, suggesting they could be purer predictors of clinically relevant variables. In some cases researchers have already demonstrated the incremental validity of the MMPI–2 content scales over the basic scales (Ben-Porath, Butcher, & Graham, 1991; Ben-Porath, McCully, & Almagor, 1993).

The enduring value of high-point coding lies in its accessibility. The dichotomization of cases is more consistent with the categorical nature of most clinical judgments. Configural coding based on a small set of scales is efficient, straightforward, and easy to grasp. Even at a time when computers are readily available to handle the task, whether clinicians would prefer a classification system that requires the linear combination of a larger set of scores unless research had clearly established its superiority to the traditional approach is questionable.

The second characteristic shared by the high-point code system studies is the relatively greater emphasis placed on predictor as opposed to outcome variables. Meehl's (1954) original discussion of actuarial approaches to data combination focused on the topic of prediction, and the actuarial research inspired by his book typically has been geared to the prediction of some fixed criterion. The cookbook studies conceived by Meehl (1956) were instead intended to contribute to the "detailed and ambitious enterprise roughly characterizable as 'describing the person'" (p. 264). This goal encourages a discovery-oriented approach to research in which the "meaning" of a patient's classification based on test data emerges out of empirical relations within a large array of clinically relevant variables.

The relative importance of predictor and criterion variables in the high-point code system studies is often obvious from the descriptions of the variables provided by the authors. Where the rules for high-point coding are often outlined and justified in great detail, the criterion variables are often listed in cursory fashion. In most studies the criteria consisted of whatever clinical variables were available to the researchers, regardless whether any a priori reason existed to expect those vari-

ables were related to the code groups under investigation. This exploratory approach to the identification of code correlates has been referred to as the study of free criteria, examining for unexpected covariations, or prediction of a taxonomic class (Gough, 1965; Sines, 1966).

When the goal is the prediction of a fixed target, such as suicidality or treatment outcome, researchers commonly sample retrospectively to ensure that approximately equal numbers of cases in the sample represent the presence and absence of the criterion condition (Fleiss, 1973). In contrast, most high-point code system researchers sample cross-sectionally without conditioning inclusion on either the predictor or the criterion.

The distinction has important implications for the statistical outcomes of the study. Dawes (1993) demonstrated that, in general, retrospective sampling results in larger effect sizes, and therefore more powerful significance tests, than cross-sectional sampling. He concluded "to the degree to which the program of psychology—and other social sciences—is to predict and influence behavior (prospectively), we are misled by retrospective analyses" (p. 7).

The potential impact of sampling method on outcomes is often ignored. For example, meta-analyses have consistently suggested a large mean effect size when using the MMPI to predict malingering (e.g., Berry et al., 1991; Rogers et al., 1994). However, many of the studies in this literature are analogue studies in which the base rate for malingering and accurate responding are similar; in some instances the same participants serve as both malingerers and controls. If the cross-sectional base rate of invalid responding is more asymmetrical than in the studies reviewed, as may be expected, the true validity of the MMPI scales reviewed could be much lower. The cross-sectional approach favored by the high-point code system researchers should produce smaller effect sizes when compared to studies that adopt a retrospective approach, but the outcomes are more reflective of what would be expected in clinical settings.

Third, in keeping with the goal of developing a general interpretive strategy, the high-point code system studies include a set of codes large enough to ensure that a substantial portion of the available MMPI profiles could be interpreted using the results of the study. In most cases, the researchers were able to generate interpretive statements for more than half of the MMPI profiles at their disposal. To do so, the researchers must identify enough cases representing each code to consider the results reliable. As Butcher, Graham, and Ben-Porath (1995) have noted, this implies the need for large sample sizes. In most of the studies reviewed, the samples used for code correlate analyses included more than 500 participants. In every case where the sample used for these analyses was less than 500, the target sample was drawn from a larger initial pool of profiles, although the authors did not always indicate the size of the pool.

The fourth common element we noted is the use of a diagnostically mixed sample of psychiatric patients. Although settings and age groups have varied, no limi-

tations were placed on the types of psychopathology included in the sample. The resulting interpretive statements are intended to be generally applicable to individuals who are most likely to complete the MMPI in psychiatric settings.

Experts have recognized for many years that clinical research requires a balance of clinical values with research values (Loevinger, 1963). This issue has emerged in many forms. Balancing internal and external validity has long been identified as a problem for the clinical researcher with the discussion in the psychotherapy literature of efficacy versus effectiveness (Seligman, 1995) representing one of its more recent instances. The high-point code system studies, more so than even other actuarial studies, reflect a strong bias in favor of the clinical value of the results over other research goals such as power. Although the clinical prejudice in the high-point code system studies is admirable, it can be expected to have consequences for the size of the relations observed when compared to those from other studies that manipulate base rates or that examine linear relations between predictors and criteria.

STUDY SUMMARIES

Once we had identified a set of defining characteristics for high-point code system studies, the next step was to identify instances not included in the two texts. We conducted a search of PsycLIT from 1974 to the present. We examined all records that included the word *MMPI* or *multiphasic* and the root term *code*. This search produced no more studies published before 1980, and only two studies since then. Discussions with the authors of the available high-point code system studies and other MMPI researchers produced one more study that was recently published.

The decline in the number of studies in recent years most likely reflects the anticipation of the publication of the MMPI–2 during the 1980s and the time required to gather the large samples needed for these studies. Given that the MMPI–2 has now been available for 10 years, the number of high-point code studies will probably increase soon. At least two more high-point code system studies are known to be in progress, one involving the MMPI–2 (Arbisi, Ben-Porath, & McNulty, 1998) and another, the newer adolescent MMPI (Powis, McGrath, Pogge, Borgaro, & Stokes, 1998). The following describes the 10 projects completed so far.

Marks and Seeman

Although several investigations of limited scope appeared in the literature earlier (e.g., Gilberstadt & Duker, 1960), Marks and Seeman's (1963) book was the first widely available and complete report of a large-scale high-point code system study to be published. Given that Marks and Seeman were the first to address many of the

methodological issues later high-point code system researchers faced and that the study was complicated (it is, for example, the only system reviewed that has undergone major revision), their work merits detailed description.

The number of male patients who completed the MMPI–1 at the time of admission was too small to consider results reliable. Except for descriptive data concerning the admission MMPI–1 for certain codes, all analyses were restricted to female patients. The sample consisted of 350 adult female psychiatric inpatients from a university medical center who met criteria for one of 16 MMPI–1 codes. Only codes represented by at least 20 cases were included. Fifteen of the codes were high-point codes involving two or three of the clinical scales. The last code was called the *normal K+*. Patients in this group met several conditions, the most important of which were *T* scores lower than 70 on all clinical scales, and a *T* score on the *K* scale that was at least 5 points higher than the *T* score for *F*.

All the scales comprising the high-point code had to equal or exceed 70 *T*. Profiles had to meet up to nine additional rules before they could be included in the code group. The additional rules were meant to maximize group homogeneity, but the authors recognized they could be overly restrictive in other settings. If the profile did not meet the criteria for any of their 16 codes, Marks and Seeman indicated that the additional rules could be relaxed. However, they did not indicate how many additional rules could be ignored, or how severely they could be violated, before application of the associated descriptors was inappropriate. Because their first rule only required the high-point code scales exceed 70 *T*, if all but the first rule were ignored it would theoretically be acceptable for the clinician to include a profile in a certain code even when the scales defining the code were not the most elevated.

Although Marks and Seeman (1963) were able to classify 78% of available profiles, others found more than 50% of MMPI–1 profiles did not meet criteria for any of their codes (see Marks, Seeman, & Haller, 1974, and Wiggins, 1972, for reviews). The inclusion of hierarchical 2-point and 3-point codes, such as 3-1 and 2-3-1, also meant that some profiles met criteria for more than one code. In addition, during the development of the Missouri system described later, Gynther, Altman, Warbin, and Sletten (1972) concluded that relaxing the classification rules did not impact much on the strength of relationships. In response to these findings, Marks and Seeman revised the classification rules for the second edition of their book (Marks et al., 1974). For 12 of the codes, the only requirement was that code scales exceed all other clinical scales, although certain scales were excluded for certain codes. This represented a dramatic change in the system. It is surprising, then, that the authors did not modify their code descriptions for the second edition; the original data were presented with the assurance that the revised model had little effect on outcomes.

Marks and Seeman (1963) examined three classes of criterion variables. They rated patient charts on 225 variables. Psychometric data were available in the form

of IQ scores, a second MMPI–1 administered at the time of hospitalization requesting the patient estimate his or her results at the time of discharge, and actual discharge MMPI–1s. Therapists also sorted 108 Q-sort descriptors into a 9-step rectangular distribution from 1 (*least descriptive*) to 9 (*most descriptive*). Scores across descriptors were correlated between each pair of patients within a code group. Based on these correlations, the Q-sort data from the "most homogeneous" five or six patients for each code group were used for descriptor derivation.

For each criterion variable, they computed percents or means for the code groups and rank ordered the groups. If the mean or percent for a code exceeded the third quartile or fell below the second quartile for all codes, it was considered an interpretable difference. The use of a relative criterion rather than an absolute standard such as statistical significance ensured that every criterion variable in the study was related to some codes.

Gilberstadt and Duker

Gilberstadt and Duker (1965) studied 266 male VA psychiatric inpatients meeting criteria for one of 19 MMPI–1 codes the authors selected because they were considered particularly representative of certain clusters of traits. They did not require a minimum number of cases per code, and code groups included as few as six patients. The comparison group was a random sample of 100 VA psychiatric patients.

All 19 were high-point codes involving between one and four clinical scales. The procedure for classifying codes was similar to that described by Marks and Seeman (1963). The first rule usually required a minimum elevation, a higher elevation relative to the other clinical scales, or both for the code scales. Up to nine other conditions were required for each code. As one would expect given the discussion so far, when their system was applied to other samples, the percentage of cases that could be classified was low (Marks et al., 1974; Wiggins, 1972). Gilberstadt and Duker's (1965) study was also the first high-point code system study to use the validity scales for exclusionary criteria independent of their code classification rules.

Three judges completed a 131-item chart review for each patient. This was the first study in which the identification of descriptors was based on a significance test comparing code and comparison groups. A difference in frequency from the comparison group at $p < .05$ was used to identify descriptors for each code group. Although the authors did not indicate this to be the case, we may presume that some members of their comparison group met the criteria for the MMPI codes under investigation, which probably reduced the power of their comparisons. On the other hand, patients were excluded from the code groups if the authors thought their records were not typical of individuals with the code.

Conceptually, this is comparable to using the most representative cases for the Q-sort analyses in the Halbower (1955) and Marks and Seeman (1963) studies with the primary difference being a rational rather than empirical approach to case selection. However, the decision has statistical implications for this study that did not apply in the previous cases because of the use of significance tests. The reduction in group variability resulting from the elimination of dissimilar cases should increase standardized effect sizes, in turn increasing the number of significant outcomes.

Lachar

Lachar (1968) chose 13 high-point codes based on their appearance in at least 75 MMPI–1s in a state hospital population over a 16-year period. Each code was represented in the sample by 30 cases, drawn at regular intervals from the pool of cases reflecting the code. In contrast to the restrictive classification strategies employed in the previous two studies, Lachar based his strategy on a more liberal model that Tellegen and Ben-Porath (1993) indicated was introduced by Hathaway and Meehl in 1951. First, Lachar restricted his attention to 2-point codes. Permutations of the two most elevated scales were combined so that, for example, 6-8 and 8-6 profiles were grouped together. No minimum elevation was required: even profiles within normal limits or with only one scale higher than 70 *T* were classified. Finally, in cases of ties lower-numbered scales took precedence. If Scales *2, 4,* and *6* were all equally elevated above the remaining clinical scales, the profile was included in the 2-4 code group.

Based on these classification rules, 100% of cases can be classified into a code group. The inclusiveness of the strategy has made it popular among subsequent high-point code researchers. However, additional conditions such as excluding potentially invalid profiles, requiring a minimum elevation of the code scales, or requiring a minimum number of cases before the code is studied has meant that in practice at least some profiles remain unclassified or some codes are not investigated.

Lachar's (1968) criterion variables were the most limited of the studies in this review. They included only demographic, IQ, and diagnostic data. The Lachar study is often not even considered one of the high-point code system studies. Even so, we included it because judgments about the diversity of criterion variables necessary to consider it a true high-point code system study would inevitably prove arbitrary, and the exploratory nature of the study met the conditions we had set for inclusion in the review. Lachar commented on differences in percentages across subgroups on diagnosis and demographic categories but did not compute inferential statistics.

Lewandowski and Graham

Lewandowski and Graham (1972) published the first study using a cross-sectional approach to sampling. They adopted the code definition strategy Lachar described (1968) and identified 19 MMPI–1 2-point codes represented by at least 5 cases in the first 292 they collected. The final sample included 588 psychiatric inpatients, 84% of whom met criteria for one of the 19 codes. As each MMPI–1 was completed it was alternately placed in one of two groups, producing two samples of 294 patients for purposes of cross-validation. Criterion variables included demographic data, case history information, ratings on the Brief Psychiatric Rating Scale, and nurse ratings. The relation between code and criterion was considered significant if the *p* value associated with *t* was less than .10 in both samples, resulting in an assumed joint probability of .01.

The Missouri System

Gynther and his colleagues presented the Missouri system in a series of 13 articles (Altman, Gynther, Warbin, & Sletten, 1972; Altman, Warbin, Sletten, & Gynther, 1973; Gynther, Altman, & Sletten, 1973a, 1973b; Gynther, Altman, & Warbin, 1972a, 1972b, 1973a, 1973b, 1973c, 1973d; Gynther, Altman, Warbin, & Sletten, 1972, 1973; Warbin, Altman, Gynther, & Sletten, 1972). It represented the largest actuarial study ever conducted, involving 3,758 psychiatric inpatients divided into four samples. The methodology bore several similarities to that of Lewandowski and Graham (1972). Gynther and colleagues used the same rules for determining the MMPI–1 profile code, although permutations of the code scales were treated separately. They selected twenty 2-point codes for further analysis on the basis of at least 30 profiles in an initial sample of 1,869 patients. They also included a high-*F* code group that included all patients with an *F* raw score above 25. Not all profiles in the sample fell into the predefined code groups.

As in Lewandowski and Graham's (1972) study, the criterion for significance was $p < .10$ in each of two samples. Four samples were used, and the second and fourth samples served as cross-validation samples for the first and third samples. Criterion variables included items and factor scores from a 111-item mental status interview in the first two samples and a 101-item list of problems and symptoms in the second two. Demographic data and diagnosis were also examined.

Marks, Seeman, and Haller

In addition to revising the criteria for high-point code classification in the Marks and Seeman (1963) study of adult psychiatric patients, Marks et al. (1974) pre-

sented a high-point code study of adolescent MMPI–1 profiles. The authors were concerned about the applicability of the adult norms, so they began by developing a new set of norms using a general sample of 1,806 adolescents divided by age (14–17) and gender. Based on previous evidence questioning the value of the *K* correction in adolescents, their raw scores were not *K*-corrected.

A national sample of 834 adolescents seen in a variety of mental health settings provided MMPI–1 profiles that were scored using the new adolescent norms. In contrast to the very restrictive classification strategy Marks and Seeman described (1963), Marks et al. (1974) used the most liberal code definition strategy found in this literature. They began with the model Lachar (1968) described. Codes that were represented by at least 10 cases in the first round of classification were selected for further investigation. If the profile did not fall into one of these groups, the code was redefined by replacing the less elevated of the two high-point code scales with the next most elevated clinical scale. Again, in cases of ties, they gave precedence to numerically lower scales. Based on these rules all but 12 profiles fell into one of the 29 code groups they examined.

Each adolescent completed a 72-item questionnaire that allowed up to 990 self-ratings and the 300-item Adjective Checklist (ACL). In addition, the therapist completed the 108-item Q-sort Marks and Seeman (1963) described, a 174-item case description questionnaire that allowed 904 ratings of the patient, and the 300-item ACL describing the patient. Descriptors that demonstrated gender differences or with base rates of greater than or equal to 90% or less than or equal to 10% were eliminated. This reduced the number of descriptors examined to 1,265. Because descriptor identification was based on comparisons with all other cases via significance test, the exclusion of criteria with extreme base rates was probably intended to avoid descriptors that—although associated with a less extreme base rate than the rest of the sample—were still unusual in the code group.

In cases where the group size was 20 or more, the group was randomly divided (matched on age and gender) into two subgroups. A descriptor had to be significantly related to the code group when compared to all other profiles at a joint alpha level of less than .06. In cases where the group size was less than 20, the descriptor had to be significantly related to code group at $p < .04$.

Kelley and King

Kelley and King presented their research in a series of seven articles (Kelley & King, 1978, 1979a, 1979b, 1979c, 1980; King & Kelley, 1977a, 1977b). Their sample consisted of 550 outpatients who completed the MMPI–1 at a university student mental health center. The study was unique in that methods were somewhat different across publications. For example, only correlates of the 2-7-8 code were cross-validated. Their classification strategy was similar to that described by Lachar

(1968), except that code scales had to meet or exceed 70 *T*. In addition to fourteen 2-point codes, they studied five 1-point (spike) codes and one 3-point code. They also examined three variants of the within-normal-limits code: profiles with $K < 65$ *T* (which they referred to as *Within Normal Limits*), with *K* between 65 and 70 (High *K*), and with *K* above 70. This last group was called *K*+ although the criteria differed from the normal *K*+ described by Marks and Seeman (1963). Codes were represented by as few as five profiles.

Criterion variables were drawn from a 179-item record review. The standard for a significant relation was generally set at $p < .05$.

Williams and Butcher

Two articles present the methodology for the Williams and Butcher (1989a, 1989b) study. Data were collected as part of the MMPI revision process. MMPI Form TX was an experimental version containing items from the MMPI–1, MMPI–2, and the more recent adolescent version (Butcher et al., 1992). It was administered to 844 adolescents in a variety of mental-health-related settings. MMPI–1 profiles were generated using the Marks et al. (1974) normative data. They eliminated potentially invalid profiles, reducing the sample to 725. They used the code definition strategy Lachar (1968) described, although they required a minimum *T* score of 65 for scales included in the code. They identified five 2-point codes that were represented by at least 20 profiles. Criterion variables included five factor scores from the Devereux Adolescent Behavior Rating Scale, which was completed by treatment staff; nine scales from the Child Behavior Checklist, completed by parents; and a 27-item record review (after elimination of 7 items with extreme base rates). A significant relation was based on $p < .01$ when adolescents matching the code were compared with the rest of the sample. The use of a more conservative alpha level rather than cross-validation was based on a recommendation by Green (1982), who noted that although both procedures maintain the same alpha level, the former is associated with more powerful statistical tests than the latter. This was the first high-point code system study in which the authors identified criterion variables they hypothesized would be related to each code.

Archer, Griffin, and Aiduk

The Archer, Griffin, and Aiduk (1995) study represents the first one published with the MMPI–2. Their sample consisted of 597 psychiatric inpatients. The procedures used for code classification, selecting codes to study, and statistical analyses were identical to those Williams and Butcher (1989b) described, except for some differences in the criteria used to identify potentially invalid profiles. In addition to nine

2-point codes represented by at least 20 cases, they included a Within Normal Limits group. Criterion variables included a demographic data form, the 18-item version of the Brief Psychiatric Rating Scale, and the Global Assessment Scale, all completed by a staff psychologist; the Symptom Checklist 90 (SCL–90); and an 88-item measure of ward behavior completed by nursing staff.

Graham, Ben-Porath, and McNulty

Graham, Ben-Porath, and McNulty (1999) conducted a study of high-point code correlates as part of a larger investigation providing clinical correlates for a broad array of MMPI–2 patterns. The high-point code study involved a sample of 429 outpatients seen at an urban community mental health center. The authors adopted a relatively restrictive approach to code classification. Potentially invalid profiles were excluded, and code scales had to exceed 64 *T*. In addition, only profiles that met the criterion for a well-defined code were included in the code groups. A well-defined code is one in which the code scales exceed all other clinical scales by at least 5 *T* points. This requirement was based on evidence that high-point codes are unreliable when less than a 5-point difference exists between code scales and other scales (e.g., Graham, Smith, & Schwartz, 1986; Graham, Timbrook, Ben-Porath, & Butcher, 1991). Graham et al. (1991) questioned the appropriateness of applying high-point code interpretive data to a profile when the code is not well-defined. McNulty, Ben-Porath, and Graham (1998) have also provided evidence that effects in their sample were larger and more consistent with expectations when the well-defined standard was used.

The study focused on twelve 2-point codes and five 3-point codes represented by at least 10 profiles. In some cases 2-point and 3-point codes were hierarchical, and 27 patients met criteria for more than one code group. Graham and his associates resolved this issue by including those patients in both code groups. Although this strategy is unique, it is not inconsistent with common MMPI practice. When a patient meets criteria for two codes (usually because clinical scales are tied), anecdotal reports suggest that many clinicians review interpretive material for both codes to generate a set of descriptive statements.

The choice of a comparison group for the analyses was also distinctive. Graham et al. (1999) noted that the usual practice of comparing the code group members to all other members of the sample will not identify descriptors that are accurate but not unique. For example, if individuals with a 1-2 code tend to be depressed, but are no more likely (or are only slightly more likely) to be depressed than the typical psychiatric patient, "depressed" probably would not emerge as a correlate of the code.

The authors thought a better alternative would involve comparison to individuals who are not experiencing serious psychopathology. Given that gathering in-

take, self-report, and therapist data for a nonclinical sample was infeasible, their comparison group consisted of those individuals who generated within normal limits profiles (all clinical scales lower than 65 *T*) in their sample but who did not seem to be underreporting according to validity scales. Unfortunately, this strategy is still not optimal as they indicated themselves. The use of within normal limits cases as the comparison group means that descriptors common to persons who seek treatment will not be identified, and to what extent their comparison group consisted of successful underreporters is unclear.

Criterion variables in this study included intake data (psychiatric history, substance abuse history, diagnosis, and mental status variables), and an SCL–90 completed shortly after intake. After three sessions, therapists completed an analogue version of the SCL–90 and a Patient Description Form. The latter included 188 items, each of which was completed on a 5-point scale. The criterion for significance was a correlation of at least .15, associated with a *p* value of .001 or less.

COMPARISON OF THE STUDIES

The 10 high-point code system studies involved a total of 8,614 psychiatric patients, including adult and adolescent outpatients and inpatients. They examined relations between 172 codes and more than 3,900 criterion variables. Tables 1 through 5 summarize a number of methodological elements from the studies.

The tables indicate striking diversity across studies in the rules used for the definition of high-point codes. Researchers varied almost every aspect of the coding strategy, including whether the order of code scales was considered, how profiles that met criteria for more than one code group were handled, and whether a minimum elevation was required. It is perhaps true that no two operationalizations in psychology are ever quite the same. Researcher A's cognitive–behavioral therapy for depression can be different from Researcher B's cognitive–behavioral therapy for depression, and even from Researcher A's definition of the term in previous studies, without necessarily invalidating the aggregation of findings across studies to draw common conclusions. A difference in degree seems to exist, however, between the high-point code system studies and, for example, the research on cognitive–behavioral therapy for depression. Except for basing the rules on which scales are most elevated, not one element of the code definition strategy has been constant across all these studies.

The differences in code definition strategies can be seen as alternative approaches to balancing what Tellegen and Ben-Porath (1993) referred to as homogeneity versus inclusiveness. This in turn represents a special case of the tension between clinical usefulness and methodological values discussed earlier. More restrictive coding rules have traditionally been justified as a means of en-

TABLE 1
Summary of the MMPI High-Point Code System Studies: Part 1

Study	*MMPI Version*	*Setting*	*Dates of Data Collection*	*Cross-Validated?*	*Samples Used for Analyses*
Marks and Seeman (1963)	MMPI–1	University medical center department of psychiatry	1960–1962	No	350 adult female psychiatric inpatients demonstrating codes
Gilberstadt and Duker (1965)	MMPI–1	VA hospital psychology service	1952–1960 for code groups, 1949–1951 for comparison group	No	Male VA patients: 266 patients demonstrating codes and 100 patients in a random sample
Lachar (1968)	MMPI–1	State hospital	Collected during a 16-year period, presumably starting in the early 1950s	No	390 adult psychiatric inpatients demonstrating codes
Lewandowski and Graham (1972)	MMPI–1	Public psychiatric hospital	Not indicated	Yes	588 consecutive protocols alternately placed in one of two samples
The Missouri System	MMPI–1	Five state hospitals and two mental health centers	1968–1972	Yes	Four samples: 1,259 psychiatric inpatients; 1,610 subsequent psychiatric inpatients; and 1,807 psychiatric inpatients rated on a new checklist (918 from Sample 2 and 889 subsequent patients), alternately assigned based on code to two samples
Marks, Seeman, and Haller (1974)	MMPI–1, adolescent norms	Various mental health settings nationwide	1965–1970	In cases where $n \geq 20$	834 White teenagers in therapy, ages 12–18

(Continued)

TABLE 1 (Continued)

Study	*MMPI Version*	*Setting*	*Dates of Data Collection*	*Cross-Validated?*	*Samples Used for Analyses*
Kelley and King	MMPI–1	University counseling center	Not indicated	Only with 2-7-8	550 student outpatients
Williams and Butcher (1989a, 1989b)	MMPI Form TX, Marks et al. (1974) adolescent norms	Inpatient and day treatment programs and special schools	1985–1988	No	725 adolescent patients
Archer, Griffin, and Aiduk (1995)	MMPI–2	Two state hospitals and one private medical hospital	Not indicated	No	597 adult psychiatric inpatients
Graham, Ben-Porath, and McNulty (1999)	MMPI–2	Urban community mental health center	1991–1992	No	Adult psychiatric outpatients: 339 with codes, 90 WNL cases

Note. MMPI = Minnesota Multiphasic Personality Inventory; VA = Department of Veterans Affairs; WNL = within normal limits. Information for Marks and Seeman (1963) does not reflect changes in the 1974 edition because data were not reanalyzed in light of the revisions.

TABLE 2
Summary of the MMPI High-Point Code System Studies: Part 2

Study	*Demographic Characteristics Described*	*Types of Codes Examined*	*Permutations Combined?*	*Exclusionary and Inclusionary Criteria*
Marks and Seeman (1963)	Gender	2- and 3-point codes, *K*+; up to 10 rules required for classification	In some cases	None indicated
Gilberstadt and Duker (1965)	None	1-, 2-, 3-, and 4-point codes; up to 10 rules required for classification	In some cases	Inclusionary: MMPI administered within 21 days of admission; ages 20 to 60; primary diagnosis not brain damage; $L \leq 60$, $F \leq 85$, $K \leq 70$ *T*; SHIQ ≥ 105
Lachar (1968)	Age, gender, years of education, and marital status	2-point codes	Yes	Inclusionary: Profile obtained within 1 month of hospitalization, patient discharged
Lewandowski and Graham (1972)	None provided	2-point codes	Yes	None indicated
The Missouri System	Age, gender, ethnicity, and SES	2-point codes, $F > 25$ raw	No	None indicated
Marks, Seeman, and Haller (1974)	Age, gender, and ethnicity	2-point codes	Yes	Exclusionary: mental deficiency, received less than 10 hr of therapy within 20 weeks
Kelley and King	Age, gender, ethnicity, and marital status	1- and 2-point codes, 2-7-8, WNL	Yes	None indicated
Williams and Butcher (1989a, 1989b)	Gender, ethnicity, and geographic distribution	2-point codes	Yes	Inclusionary: fifth-grade reading level, CNS ≤ 10 raw, *L* and $K \leq 70$ *T*, $F \leq 90$ *T*

(Continued)

TABLE 2 (Continued)

Study	*Demographic Characteristics Described*	*Types of Codes Examined*	*Permutations Combined?*	*Exclusionary and Inclusionary Criteria*
Archer, Griffin, and Aiduk (1995)	Age, gender, educational level, ethnicity, and marital status	2-point codes	Yes	Exclusionary: CNS > 29 raw, VRIN or TRIN > 79 *T*
Graham, Ben-Porath, and McNulty (1999)	Age, gender, educational level, ethnicity, marital status, and employment status	2- and 3-point codes	Yes	Inclusionary: CNS ≤ 30 raw, VRIN ≤ 80 *T*, TRIN 6–12 raw, *F* < 27 (men) or 29 (women) raw, *Fb* < 23 (men) or 24 (women)

Note. MMPI = Minnesota Multiphasic Personality Inventory; SHIQ = Shipley–Hartford Institute of Living Scale; SES = socioeconomic status; WNL = within normal limits; CNS = Cannot Say; VRIN = Variable Response Inconsistency Scale; TRIN = True Response Inconsistency Scale. Information for Marks and Seeman (1963) does not reflect changes in the 1974 edition because data were not reanalyzed in light of the revisions.

TABLE 3
Summary of the MMPI High-Point Code System Studies: Part 3

Study	*Minimum T Score Required*	*How Ties Were Handled*	*Criterion for Choosing Codes*	*Codes Examined*
Marks and Seeman (1963)	≥ 70	Additional rules meant to avoid ties	≥ 20 of 826 profiles from a 2-year period (2.4%)	2-3-1, 2-7, 2-7-4, 2-7-8, 2-8, 3-1, 3-2-1, 4-6, 4-6-2, 4-8-2, 4-9, 8-3, 8-6, 8-9, 9-6, *K*+
Gilberstadt and Duker (1965)	≥ 70	Additional rules meant to avoid ties	Codes that "had come to attention because of their apparent power as cardinal types in representing trait clusters" (p. 12)	1-2-3, 1-2-3-4, 1-2-3-7, 1-3-2, 1-3-7, 1-3-8, 1-3-9, 2-7, 2-7-4, 2-7-8, 4, 4-3, 4-9, 7-8, 8-1-2-3, 8-2-4, 8-6, 8-9, 9
Lachar (1968)	None	Lower numbered scale given precedence	≥ 75 first-administration profiles over 16-year period	1-2, 1-3, 2-3, 2-4, 2-7, 2-8, 3-4, 4-6, 4-8, 4-9, 6-8, 7-8, 8-9
Lewandowski and Graham (1972)	None	Lower numbered scale given precedence	≥ 5 profiles in an initial sample of 292 (1.7%)	1-2, 1-3, 1-4, 1-8, 2-3, 2-4, 2-7, 2-8, 3-4, 4-5, 4-6, 4-7, 4-8, 4-9, 5-9, 6-8, 6-9, 7-8, 8-9
The Missouri System	None	Lower numbered scale given precedence	≥ 30 profiles in an initial sample of 1869 (1.6%)	2-1, 2-3, 2-4, 2-7, 2-8, 4-2, 4-3, 4-5, 4-6, 4-8, 4-9, 6-4, 6-8, 8-2, 8-4, 8-6, 8-7, 8-9, 9-4, 9-8, High *F*
Marks, Seeman, and Haller (1974)	None	Lower numbered scale given precedence	≥ 10 profiles in the sample (1.2%); if *N* for a code was < 10, the code was redefined substituting the third elevation for the second	1-2, 1-3, 1-4, 1-5, 1-6, 1-8, 2-3, 2-4, 2-5, 2-7, 2-8, 2-0, 3-4, 3-5, 3-6, 4-5, 4-6, 4-7, 4-8, 4-9, 4-0, 5-6, 5-9, 5-0, 6-8, 7-8, 7-9, 7-0, 8-9
Kelley and King	≥ 70	Lower numbered scale given precedence	≥ 5 profiles in the sample (0.9%)	2, 2-0, 2-4, 2-5, 2-6, 2-7-8, 2-8, 3-4, 3-9, 4, 4-5, 4-7, 4-9, 5, 5-7, 5-8, 6-7, 7-9, 9, 0, WNL (*K* < 65), WNL (High *K*), *K*+
Williams and Butcher (1989a, 1989b)	≥ 65	Lower numbered scale given precedence	≥ 20 of 725 valid profiles (2.8%)	2-4, 4-6, 4-7, 4-8, 4-9
Archer, Griffin, and Aiduk (1995)	≥ 65	Lower numbered scale given precedence	≥ 20 of 597 valid profiles (3.4%)	2-3, 2-4, 2-7, 2-8, 4-6, 4-8, 4-9, 6-8, 7-8, WNL

(Continued)

TABLE 3 (Continued)

Study	*Minimum T Score Required*	*How Ties Were Handled*	*Criterion for Choosing Codes*	*Codes Examined*
Graham, Ben-Porath, and McNulty (1999)	≥ 65	Codes were well defined (code scales ≥ 5 *T* points higher than other scales); profiles could be included in both a 2-point code and 3-point code group	≥ 10 of 1,020 valid profiles (1.0%)	1-2, 1-3, 2-3, 2-4, 2-6, 2-7, 3-4, 4-6, 4-8, 4-9, 6-8, 7-8, 1-2-3, 2-4-7, 2-7-8, 4-6-8, 4-7-8

Note. MMPI = Minnesota Multiphasic Personality Inventory; WNL = within normal limits. Information for Marks and Seeman (1963) does not reflect changes in the 1974 edition because data were not reanalyzed in light of the revisions.

TABLE 4
Summary of the MMPI High-Point Code System Studies: Part 4

Study	*Moderator Variables Examined*	*Amount of Sample in Target Codes*	*Amount of Available Profiles in Target Codes*	*Dependent Variables*	*Comparison Group*
Marks and Seeman (1963)	None	100%	644 of 826 (78%)	Clinician ratings (N = 81 women), record review (N = 80 women), intellectual ability, self-report	Rest of sample
Gilberstadt and Duker (1965)	None	266 of 366 (73%)	Not indicated	Record review; diagnostic and "cardinal features" data also collected and described impressionistically	100 patients admitted 1949 to 1951; L < 61, F < 86, K < 71 T
Lachar (1968)	None	100%	Not indicated	Demographics, IQ, diagnosis	Rest of sample
Lewandowski and Graham (1972)	None	84% in both samples[a]	84%[a]	Demographics, biographical data, clinician ratings	Rest of sample
The Missouri System	Gender, ethnicity	76% for Samples 1 and 2[a]	76%[a]	Demographics, clinician ratings	Rest of sample
Marks, Seeman, and Haller (1974)	None	822 of 834 (99%)	822 of 834 (99%)	Clinician ratings, self-report	Rest of sample
Kelley and King	Gender	410 of 550 (75%)	410 of 550 (75%)	Record review	Rest of sample
Williams and Butcher (1989a, 1989b)	None	202 of 725 (28%)	202 of 844 (24%)	Clinician ratings (N = 527), parent ratings (N = 476), record review (N = 709)	Rest of sample
Archer, Griffin, and Aiduk (1995)	None	381 of 597 (64%)	381 of 704 (54%)	Demographics, clinician ratings, self-report	Rest of sample
Graham, Ben-Porath, and McNulty (1999)	None	339 of 429 (79%)	339 of 1,219 (28%)	Clinician ratings, self-report	90 WNL cases

Note. MMPI = Minnesota Multiphasic Personality Inventory; WNL = within normal limits. Information for Marks and Seeman (1963) does not reflect changes in the 1974 edition because data were not reanalyzed in light of the revisions.

[a]These values were provided in a source article.

TABLE 5
Summary of the MMPI High-Point Code System Studies: Part 5

Study	*Raters Blind to MMPI?*	*Interrater Reliability or Agreement Examined?*	*Criteria for Significance*	*Statistics Provided*
Marks and Seeman (1963)	Yes	Record review: two of three raters had to agree; clinical ratings: mean reliability .50	Above third quartile or below second quartile for all codes	Percentages, means, some ranges
Gilberstadt and Duker (1965)	Yes	Two out of three raters had to agree	$p < .05$	Percentages
Lachar (1968)	Not indicated	No	None	Percentages
Lewandowski and Graham (1972)	Not indicated	No	*t* tests, $p < .10$ in both samples	None
The Missouri System	Not indicated	No	$p < .10$ in each of two samples	In some cases, χ^2 and percentages
Marks, Seeman, and Haller (1974)	In 90% of cases	No	Joint $p < .06$ if cross-validated; otherwise, $p < .04$	None
Kelley and King	Yes	96.4% agreement on a random subset	Generally $p < .05$	None
Williams and Butcher (1989a, 1989b)	Yes	Record review: 14% rated twice, 26 of 27 kappas significant at .05	*t* and χ^2 tests, $p < .01$	For significant analyses, means, standard deviations, and *t* or χ^2
Archer, Griffin, and Aiduk (1995)	Yes	No	$p < .01$	For significant analyses, means, standard deviations, and *t* or χ^2
Graham, Ben-Porath, and McNulty (1999)	Yes	No	$r \geq .15$ and $p < .001$	Percentages, means, standard deviations; for significant analyses, correlations

Note. MMPI = Minnesota Multiphasic Personality Inventory. Information for Marks and Seeman (1963) does not reflect changes in the 1974 edition since data were not reanalyzed in light of the revisions.

hancing predictive validity, although the well-defined code was developed primarily to enhance code reliability. However, greater restrictiveness also means less inclusiveness.

The diversity of coding strategies creates a dilemma for the clinician who wants to use actuarial research for actuarial description. Little research exists comparing the impact of alternative coding strategies on statistical relations and inclusion rates. In fact, the study by McNulty et al. (1998) represents the first evaluation of the relation between coding strategy and statistical outcomes in a number of years.

Not a lot is known about how clinicians code MMPI profiles in the absence of clear guidelines. That clinicians will adopt different strategies seems likely, and they may even modify their approach across profiles. This would be ironic given that the MMPI has been used previously to demonstrate the value of a consistent approach to scale interpretation (Goldberg, 1970).

Variations in Inclusiveness

The impact of coding restrictions on inclusion may be found in column 4 of Table 4. This column indicates the percentage of all MMPIs available to the researchers that were included in the code groups studied. In determining this value, restrictions based on criteria other than the MMPI, such as the exclusion of "mentally deficient" candidates in the Marks et al. (1974) study with adolescents, were not considered (the authors usually did not provide data on the number excluded by these additional criteria anyway). For example, 78% of the MMPI profiles available to Marks and Seeman (1963) were included in one of their code groups, and Williams and Butcher (1989b) were able to classify only 24% of the 844 profiles in their initial sample. We have already noted that the rate for the Marks and Seeman study should be considered spuriously high for such a restrictive coding strategy because it was based on sample-derived criteria.

The first studies in the series (Gilberstadt & Duker, 1965; Marks & Seeman, 1963) demonstrated a clear bias in favor of homogeneity, although Marks and Seeman's suggestions for relaxing their inclusion rules indicate they were sensitive to the clinician's need for an inclusive interpretive system. Beginning with Lachar's (1968) study, the bias shifted toward inclusiveness. In response to evidence that restrictive coding strategies resulted in low levels of inclusiveness in other settings, and that they did not necessarily improve statistical outcomes, even Marks et al. (1974) joined the trend.

In more recent studies the pendulum has swung back. The consistent use of validity data and minimum elevations in the three most recent studies has resulted in a decline in inclusiveness. Graham et al.'s (1999) use of well-defined codes, in addition to eliminating potentially invalid profiles and requiring a minimum elevation for the code scales, represents the most restrictive coding strategy to appear in quite some time.

Factors Affecting Inclusiveness

The range in inclusiveness is large, from 24 to 99%. Several methodological factors clearly influenced inclusion rates, such as the recoding of profiles with infrequent codes in the Marks et al. (1974) adolescent study. Given the importance of inclusion as a criterion for a high-point coding system, we thought a more detailed examination of factors that influence inclusiveness worthwhile. After exploring several options, we found that a review of the three most recent studies proved to be the most informative analysis. These studies used a broader set of exclusionary criteria than earlier studies, which resulted in lower inclusion rates. They are also more informative than earlier studies about the impact of various exclusionary criteria on inclusion rate. Table 6 shows the results of this analysis.

Across all three studies, the exclusion of potentially invalid protocols eliminated about 15% of the available profiles, although studies used somewhat different criteria for potential invalidity. In the two studies that required at least 20 cases before a code group was targeted, this condition excluded about 25% of the sample. The proportion was consistent even though 20 cases translated into a larger proportion of the Archer et al. (1995) sample (3.4%) than of the Williams and Butcher (1989b) sample (2.8%).

Graham et al. (1999) examined codes represented by as few as 10 cases, or 1% of their initial sample. Even so, the criterion excluded more than 30% of their pro-

TABLE 6
Reasons for Low Inclusion Rates

Study	*Inclusion Rate (%)*	*Exclusionary Factor*	*% Eliminated*
Williams and Butcher (1989a, 1989b)	24	Invalid profiles	14.1
		Spike or WNL profiles	34.4
		Code $N < 20$	27.6
Graham, Ben-Porath, and McNulty (1999)[a]	28	Invalid profiles	16.3
		WNL profiles	7.4
		Spike or poorly defined profiles	18.4
		Code $N < 10$	30.1
Archer, Griffin, and Aiduk (1995)	54	Invalid profiles	15.2
		Spike profiles	6.8
		Code $N < 20$	23.9

Note. WNL = within normal limits. In each case, the divisor for the percentage eliminated was the initial sample size for purposes of comparison. However, it should be noted that criteria were usually applied sequentially, so the proportion of cases meeting exclusionary criteria that were later in the sequence could be higher.

[a]We thank J. L. McNulty (personal communication, March 20, 1999) for some of the information included in this portion of the table.

files. We suspect the discrepancy reflects a hidden effect of using well-defined codes on inclusion rate. Suppose a profile demonstrated a common but poorly defined 2-point code and an uncommon but well-defined 3-point code. The profile was officially excluded from consideration because of infrequency rather than poor definition.

Williams and Butcher (1989b) excluded almost 35% of their sample for spike or within-normal-limits profiles. In contrast, only 20.5% of the Archer et al. (1995) profiles were spike or within-normal-limits profiles, and only 25.8% of the Graham et al. (1999) profiles were spike, within normal limits, or poorly defined. The difference suggests a much higher proportion of relatively unelevated profiles in the Williams and Butcher sample. Whether this is a function of an adolescent rather than an adult sample, including collection sites that were not traditionally psychiatric, or some less obvious factor is unclear.

To summarize, excluding potentially invalid profiles reduced the available protocols by about 15%, and requiring at least 20 cases eliminated another 25%. within-normal-limits and spike profiles seem to have represented a larger portion of the Williams and Butcher (1989b) sample, which was comprised of adolescents from heterogeneous settings. Finally, computing the isolated effect from the information provided is not possible, but the use of well-defined codes clearly reduced the number of profiles included in code groups by at least 18%.

Unique Versus Complete Description

Another issue this literature raises is the optimal level of patient description. Most of the studies compared members of the code group to the rest of the sample, and the resulting actuarial description consisted of those variables on which the code group was distinctive. Halbower's (1955) actuarial descriptions instead consisted of the mean score for members of the code group on every Q-sort item. He made no attempt to ensure discrimination between the code groups, although eliminating low-variability items increased the probability of differences between code groups, and Meehl (1956) provided some evidence of differences between the code group descriptions and a Q-sort profile for the "average" patient.

Graham et al.'s (1999) approach is closer in spirit to Halbower's (1955). The use of within-normal-limits patients as the comparison group was an attempt at a more complete description of code group members, regardless of whether the descriptors were unique.

We find no reason why unique description and complete description cannot be considered complementary goals. If patients with 1-2 codes are depressed, but are no more depressed than patients in general, both of these pieces of information will probably interest the clinician, and both deserve to be reported. Until

now, however, researchers have opted to report one or the other, raising the issue of which is more useful. Meehl (1956) seemed to advocate for both positions. On the one hand, he referred to the task of the cookbook as "describing the person," and the methodology of the Halbower study was geared toward complete description (within the limits of the Q-sort items included) of the code groups. On the other hand, Meehl argued at length against assessment procedures that tend to produce descriptions rendered trivial by their extreme likelihood in the population of psychiatric patients, and introduced the term Barnum effect in the article.

The latter argument seems particularly compelling. The unique characteristics of the patient are most likely to be of use to the clinician because clinical decision-making depends largely on the capacity to distinguish among the members of a patient population. One can also argue that in a clinical environment influenced by managed care, the value of activities such as clinical assessment must be demonstrated if they are to continue (Meyer et al., 1998). It has therefore become particularly important to ensure that assessment procedures reveal more than what is generically true of psychiatric patients.

CONCLUSIONS

The high-point code is perhaps not the best approach for classifying MMPI profiles. In fact, Sines (1966) described the beginnings of an actuarial interpretive system based on a linear combination of the clinical scales using squared Euclidean distances. Unfortunately, the project was never completed. The optimal strategy for profile classification continues to be a source of discussion (cf. Tellegen & Ben-Porath, 1996). Ideally, it would be a linear combination of information from more than just the two most elevated scales. It might even include scales not included in the basic profile, particularly the content scales.

For the foreseeable future, however, high-point coding will probably remain the system of choice for the classification of MMPI profiles. Additional research concerning the impact of alternative coding systems on inclusiveness and criterion-related validity is therefore probably appropriate. Learning more about the strategies practicing clinicians use for high-point coding in the absence of consistent guidelines might also be interesting.

ACKNOWLEDGMENTS

The completion of this study would not have been possible without the help of many people. In particular, we thank John Graham, Yossef Ben-Porath, John McNulty, and Malcolm Gynther.

REFERENCES

Altman, H., Gynther, M. D., Warbin, R. W., & Sletten, I. W. (1972). A new empirical automated MMPI interpretive program: The 6-8/8-6 code type. *Journal of Clinical Psychology, 28,* 495–498.

Altman, H., Warbin, R. W., Sletten, I. W., & Gynther, M. D. (1973). Replicated correlates of the MMPI 8-9/9-8 code type. *Journal of Personality Assessment, 37,* 369–371.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (1998, March). *Clinical correlates of common MMPI–2 two-point codes in psychiatric inpatients: A first look.* Presented at the 33rd Annual Symposium on Recent Developments in the Use of the MMPI–2 and MMPI–A, Clearwater Beach, FL.

Archer, R. P. (1997). *MMPI–A: Assessing adolescent psychopathology* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Archer, R. P., Griffin, R., & Aiduk, R. (1995). MMPI–2 clinical correlates for ten common codes. *Journal of Personality Assessment, 65,* 391–407.

Baer, R. A., Wetter, M. W., & Berry, D. T. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12,* 509–525.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI–2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment, 3,* 634–640.

Ben-Porath, Y. S., McCully, E., & Almagor, M. (1993). Incremental validity of the MMPI–2 content scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment, 61,* 557–575.

Berry, D. T., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analysis. *Clinical Psychology Review, 11,* 585–598.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Manual for administering and scoring.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (1995). Methodological problems and issues in MMPI, MMPI–2, and MMPI–A research. *Psychological Assessment, 7,* 320–329.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *MMPI–A manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Dahlstrom, W. G. (1992). Comparability of two-point high-point code patterns from original MMPI norms to MMPI–2 norms for the restandardization sample. *Journal of Personality Assessment, 59,* 153–164.

Dawes, R. M. (1993). Prediction of the future versus an understanding of the past: A basic asymmetry. *American Journal of Psychology, 106,* 1–24.

Fleiss, J. L. (1973). *Statistical methods for rates and proportions.* New York: Wiley.

Gilberstadt, H., & Duker, J. (1960). Case history correlates of three MMPI profile types. *Journal of Consulting Psychology, 24,* 361–367.

Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation.* Philadelphia: Saunders.

Goldberg, L. R. (1965). Diagnosticians versus diagnostic signs: The diagnosis of psychosis versus neurosis from the MMPI. *Psychological Monographs, 79*(9, Whole No. 602).

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73,* 422–432.

Gough, H. G. (1965). Conceptual analysis of psychological test scores and other diagnostic variables. *Journal of Applied Psychology, 70,* 294–302.

Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). *MMPI–2 correlates for outpatient community mental health settings.* Minneapolis: University of Minnesota Press.

Graham, J. R., Smith, R. L., & Schwartz, G. F. (1986). Stability of MMPI configurations for psychiatric inpatients. *Journal of Consulting and Clinical Psychology, 54,* 375–380.

Graham, J. R., Timbrook, R. E., Ben-Porath, Y. S., & Butcher, J. N. (1991). Code-type congruence between MMPI and MMPI–2: Separating fact from artifact. *Journal of Personality Assessment, 57,* 205–215.

Green, S. B. (1982). Establishing behavioral correlates: The MMPI as a case study. *Applied Psychological Measurement, 6,* 219–224.

Greene, R. L. (1991). *The MMPI–2/MMPI: An interpretive manual.* Boston: Allyn & Bacon.

Groth-Marnat, G. (1997). *Handbook of psychological assessment* (3rd ed.). New York: Wiley.

Gynther, M. D., Altman, H., & Sletten, I. W. (1973a). Development of an empirical interpretive system for the MMPI: Some after-the-fact observations. *Journal of Clinical Psychology, 29,* 232–234.

Gynther, M. D., Altman, H., & Sletten, I. W. (1973b). Replicated correlates of MMPI two-point code types: The Missouri Actuarial System. *Journal of Clinical Psychology, 29,* 263–289.

Gynther, M. D., Altman, H., & Warbin, R. W. (1972a). A new actuarial–empirical automated MMPI interpretive program: The 4-3/3-4 code type. *Journal of Clinical Psychology, 29,* 229–231.

Gynther, M. D., Altman, H., & Warbin, R. W. (1972b). A new empirical automated MMPI interpretive program: The 2-4/4-2 code type. *Journal of Clinical Psychology, 28,* 498–501.

Gynther, M. D., Altman, H., & Warbin, R. W. (1973a). Behavioral correlates for the Minnesota Multiphasic Personality Inventory 4-9, 9-4 code types: A case of the emperor's new clothes? *Journal of Consulting and Clinical Psychology, 40,* 259–263.

Gynther, M. D., Altman, H., & Warbin, R. W. (1973b). Interpretation of uninterpretable MMPI profiles. *Journal of Consulting and Clinical Psychology, 40,* 78–83.

Gynther, M. D., Altman, H., & Warbin, R. W. (1973c). A new empirical automated MMPI interpretive program: The 2-7/7-2 code type. *Journal of Clinical Psychology, 29,* 58–59.

Gynther, M. D., Altman, H., & Warbin, R. W. (1973d). A new empirical automated MMPI interpretive program: The 6-9/9-6 code type. *Journal of Clinical Psychology, 29,* 60–61.

Gynther, M. D., Altman, H., Warbin, R. W., & Sletten, I. W. (1972). A new actuarial system for MMPI interpretation: Rationale and methodology. *Journal of Clinical Psychology, 28,* 173–179.

Gynther, M. D., Altman, H., Warbin, R. W., & Sletten, I. W. (1973). A new empirical automated MMPI interpretive program: The 1-2/2-1 code type. *Journal of Clinical Psychology, 29,* 54–57.

Halbower, C. C. (1955). A comparison of actuarial versus clinical prediction to classes discriminated by the Minnesota Multiphasic Personality Inventory (Doctoral dissertation, University of Minnesota, 1955). *Dissertation Abstracts International, 15*(06), 1115.

Hathaway, S. R., & McKinley, J. C. (1983). *Manual for the administration and scoring of the MMPI.* Minneapolis, MN: National Computer Systems.

Kelley, C. K., & King, G. D. (1978). Behavioral correlates for within-normal-limit MMPI profiles with and without elevated *K* in students at a university mental health center. *Journal of Clinical Psychology, 34,* 695–699.

Kelley, C. K., & King, G. D. (1979a). Behavioral correlates of infrequent two-point MMPI code types at a university mental health center. *Journal of Clinical Psychology, 35,* 576–585.

Kelley, C. K., & King, G. D. (1979b). Behavioral correlates of the 2-7-8 MMPI profile type in students at a university mental health center. *Journal of Consulting and Clinical Psychology, 47,* 679–685.

Kelley, C. K., & King, G. D. (1979c). Cross validation of the 2-8/8-2 MMPI code type for young adult psychiatric outpatients. *Journal of Personality Assessment, 43,* 143–149.

Kelley, C. K., & King, G. D. (1980). Two- and three-point classification of MMPI profiles in which Scales 2, 7, and 8 are the highest elevations. *Journal of Personality Assessment, 44,* 25–33.

King, G. D., & Kelley, C. K. (1977a). Behavioral correlates for spike-4, spike-9, and 4-9/9-4 MMPI profiles in students at a university mental health center. *Journal of Clinical Psychology, 33,* 718–724.

King, G. D., & Kelley, C. K. (1977b). MMPI behavioral correlates of spike-5 and two-point code types with Scale 5 as one elevation. *Journal of Clinical Psychology, 33,* 180–185.

Lachar, D. (1968). MMPI two-point code-type correlates in a state hospital population. *Journal of Clinical Psychology, 24,* 424–427.

Lachar, D. (1974). *The MMPI: Clinical assessment and automated interpretation.* Los Angeles: Western Psychological Services.

Levitt, E. E. (1989). *The clinical application of MMPI special scales.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lewak, R. W., Marks, P. A., & Nelson, G. E. (1990). *Therapist guide to the MMPI and MMPI–2: Providing feedback and treatment.* Muncie, IN: Accelerated Development.

Lewandowski, D., & Graham, J. R. (1972). Empirical correlates of frequently occurring two-point MMPI code types: A replicated study. *Journal of Consulting and Clinical Psychology, 39,* 467–472.

Loevinger, J. (1963). Conflict of commitment in clinical research. *American Psychologist, 18,* 241–251.

Marks, P. A., & Seeman, W. (1963). *The actuarial description of abnormal personality: An atlas for use with the MMPI.* Baltimore: Williams & Wilkins.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults.* Baltimore: Williams & Wilkins.

McNulty, J. L., Ben-Porath, Y. S., & Graham, J. R. (1998). An empirical examination of the correlates of well-defined and not defined MMPI–2 code types. *Journal of Personality Assessment, 71,* 393–410.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1956). Wanted—A good cookbook. *American Psychologist, 11,* 262–272.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W., Moreland, K. L., Eisman, E. J., & Dies, R. R. (1998). *Benefits and costs of psychological assessment in healthcare delivery: Report of the Board of Professional Affairs Psychological Assessment Work Group, Part I.* Washington, DC: American Psychological Association.

Piotrowski, C. (1997). Use of the Millon Clinical Multiaxial Inventory in clinical practice. *Perceptual and Motor Skills, 84,* 1185–1186.

Powis, D., McGrath, R., Pogge, D., Borgaro, S., & Stokes, J. (1998, March). *Actuarial use of the MMPI–A: Generation of clinical correlate data for frequently occurring codetypes in an adolescent inpatient sample.* Presented at the 33rd Annual Symposium on Recent Developments in the Use of the MMPI–2 and MMPI–A, Clearwater Beach, FL.

Rogers, R., Sewell, K. W., & Salekin, R. T. (1994). A meta-analysis of malingering on the MMPI–2. *Assessment, 1,* 227–237.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50,* 965–974.

Sines, J. O. (1966). Actuarial methods in personality assessment. In B. A. Maher (Ed.), *Progress in experimental personality research: Vol. 3* (pp. 133–193). New York: Academic.

Tellegen, A., & Ben-Porath, Y. S. (1993). Code-type comparability of the MMPI and MMPI–2: Analysis of recent findings and criticisms. *Journal of Personality Assessment, 61,* 489–500.

Tellegen, A., & Ben-Porath, Y. S. (1996). Evaluating the similarity of MMPI–2 and MMPI profiles: Reply to Dahlstrom and Humphrey. *Journal of Personality Assessment, 66,* 640–644.

Warbin, R. W., Altman, H., Gynther, M. D., & Sletten, I. W. (1972). A new empirical automated MMPI interpretive program: 2-8 and 8-2 code types. *Journal of Personality Assessment, 36,* 581–584.

Wiggins, J. S. (1972). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Williams, C. L., & Butcher, J. N. (1989a). An MMPI study of adolescents: I. Empirical validity of the standard scales. *Psychological Assessment, 1,* 251–259.

Williams, C. L., & Butcher, J. N. (1989b). An MMPI study of adolescents: II. Verification and limitations of code type classifications. *Psychological Assessment, 1,* 260–265.

Robert E. McGrath
School of Psychology
Fairleigh Dickinson University
Teaneck, NJ 07666
E-mail: mcgrath@alpha.fdu.edu

Received April 28, 1999