JOURNAL OF PERSONALITY ASSESSMENT, *79*(2), 243–256

A Comparison of MMPI–2 High-Point Coding Strategies

Robert E. McGrath, Tayyab Rashid, and Judy Hayman
School of Psychology
Fairleigh Dickinson University

David L. Pogge
Four Winds Hospital
Katonah, New York

High-point coding refers to the popular practice of classifying Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1983) profiles based on which clinical scales are the most elevated. A previous review of high-point code studies (McGrath & Ingersoll, 1999a) noted marked discrepancies across studies in the rules used to define high-point codes. This study was conducted to evaluate the costs and benefits of different strategies for high-point coding. The impact of 4 rules for high-point coding on effect sizes and group sizes was evaluated. The 4 rules included requiring a minimum elevation, excluding potentially invalid protocols, restricting coding to well-defined codes, and replacing the lower scale in infrequently occurring codes with the next most elevated scale. The evidence supported the clinical utility of requiring a minimum elevation for code scales. The results were more equivocal concerning the value of well-defined coding and for not replacing the lower scale in infrequent codes. Results were surprisingly negative concerning the utility of excluding potentially invalid protocols, suggesting that guidelines developed in situations in which there is a clear motivation to distort results may not generalize to other settings.

High-point coding refers to the practice of grouping Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1983) profiles for purposes of interpretation based on which of the clinical scales are most elevated. Usually classification is based on the two-point code, though one-, three-, and even four-point codes have been studied. High-point coding has been closely associated with the MMPI since the early 1950s (Tellegen & Ben-Porath, 1993). It offered a simple and efficient approach to clustering cases long before computers made more sophisticated methods practical. Even today, most MMPI texts treat the high-point code as an important factor in the interpretation of the profile (e.g., Friedman, Lewak, Nichols, & Webb, 2000; Greene, 1999).

In a review of 10 large-sample studies that have been conducted examining high-point codes in general, McGrath and Ingersoll (1999a, 1999b) noted marked inconsistencies across studies in the rules used to define 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. … Except for basing the rules on which scales are most elevated, there is not one element of the code definition strategy that has been constant across all these studies. (McGrath & Ingersoll, 1999a, p. 162)

Tellegen and Ben-Porath (1993) provided a context for understanding these inconsistencies through their discussion of homogeneity versus inclusiveness as the conflicting goals of a high-point coding strategy. Individuals with a 2–4 code who also appear to be responding validly, and whose *T* scores on Scales 2 and 4 both exceed 64, are likely to be a more homogeneous group than all individuals with Scales 2 and 4 elevated above the other clinical scales. However, those profiles not meeting the additional criteria become unclassifiable, reducing the clinical usefulness of an interpretive strategy based on high-point codes.

In recent years only one study has been published evaluating the impact of different strategies for high-point coding. McNulty, Ben-Porath, and Graham (1998) were interested in comparing the criterion-related validity of well-defined and poorly defined high-point codes. The concept of well-defined coding emerged out of research on two-point code congruence between the MMPI and its revised version, the MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). In response to previous studies demonstrating that congruence was poor between the two forms, Graham, Timbrook, Ben-Porath, and Butcher (1991) demonstrated that congruence rates were much better among well-defined codes, those in which the less elevated code scale exceeded the next highest scale by at least five *T* points. Reasoning that high-point codes are likely to be unreliable when they are poorly defined, they recommended against using interpretive data based on the high-point code unless the code is well-defined.

McNulty et al. (1998) found that, as hypothesized, well-defined codes were on average associated with larger concurrent validity coefficients than poorly defined codes when the criterion was conceptually related to the code. The mean correlation between well-defined code variables and related criteria exceeded the mean correlation for poorly defined codes by .03, leading the authors to support Graham et al.'s (1991) recommendation. However, they did not consider the cost of well-defined coding on inclusiveness in their conclusions. We would propose that any comparison of high-point coding strategies should weigh the impact of the rules on both validity and classification rates.

Given inconsistencies in the rules used, we thought it would be worthwhile to compare several of the approaches to high-point coding described in the literature, with the purpose of developing recommendations for clinical practice. This study focused on four high-point coding rules. In each case the impact of the additional rule on both homogeneity and inclusiveness was evaluated.

METHOD

Participants

The sample included 752 adult inpatients who completed the MMPI–2 on admission to Four Winds Hospital, a private psychiatric facility in the New York metropolitan area. Table 1 summarizes demographic data for the sample.

Measures

In addition to the MMPI–2, scores were also available for each patient from the Symptom Checklist–90 (SCL–90; $N = 648$), the Hopkins Psychiatric Rating Scale (HPRS; $N = 696$), or both. The SCL–90 (Derogatis, 1983) is a self-report measure used extensively in psychiatric settings. The 90 five-point items are used to measure nine symptom dimensions: Somatization, Obsessive–Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. In addition, there are three general distress measures that incorporate information from all 90 items. It was completed at the same time as the MMPI–2.

The HPRS is a clinician rating scale developed specifically to parallel the SCL–90 (Derogatis, 1983). Each item is completed on an anchored 7-point scale of severity. It includes one item representing each of the nine symptom dimensions from the SCL–90 as well as nine additional items: Sleep Disturbance, Psychomotor Retardation, Hysterical Behavior, Abjection-Disinterest, Conceptual Dysfunction, Disorientation, Excitement, Euphoria, and Global Pathology (completed on a 9-point scale). The patient's primary therapist completed the HPRS within 72 hr of admission.

TABLE 1
Demographic Data

	M	*SD*	*N*	%
Age	36.1	13.9	752	
Education	14.0	3.1	699	
Gender				
Male			348	46.3
Female			404	53.7
Marital status				
Single			355	47.8
Married			215	29.0
Widowed			22	3.0
Divorced			88	11.9
Separated			62	8.4

Procedure

Each patient was first assigned a high-point code based on the following set of rules adopted from previous high-point code studies:

1. The code was based on the two most elevated clinical scales.
2. Scales 5 and 0 were ignored here and in all subsequent classifications.
3. Numeric precedence was used to resolve ties. For example, if Scales 2 and 4 were equally elevated, Scale 2 was given preference in determining the code.
4. The order of elevation was ignored, so patients were placed in the 6–8 code group whether 6 or 8 was the more elevated.

Using these rules it was possible to identify a high-point code for each patient in the sample.

Four additional coding rules were then each applied independently to the sample. The first two are typically considered part of standard practice concerning the use of high-point codes. The Minimum Elevation rule restricted high-point coding to those individuals whose code scales were both 65 *T* or higher. The Validity rule excluded patients who potentially responded in an invalid manner from the high-point code groups. Profiles were considered potentially invalid if the person omitted more than 30 items ($N = 6$), VRIN was > 80 *T* ($N = 50$), the raw score for TRIN was < 6 or > 12 ($N = 36$), or the raw score for F was > 26 in men or 28 in women ($N = 23$; Graham, Ben-Porath, & McNulty, 1999).

The Well-Defined Coding rule restricted coding to those profiles in which the code scales exceeded all the other scales by at least 5 points. Although it is not part of standard coding practice, several fairly influential MMPI researchers have made the case it should be (Graham et al., 1999; Tellegen & Ben-Porath, 1993).

The Third-Point Replacement rule was different from the others in that (a) no contemporary commentators are advocating its adoption as part of standard practice and (b) it was less restrictive than the original two-point coding system. Despite the first point, it was included in this investigation because it potentially has a profound impact on inclusiveness. Marks, Seeman, and Haller (1974) only examined high-point codes that occurred in at least 10 of their adolescent profiles. If the original code occurred less than 10 times, they replaced the lower point in the code with the third most elevated scale. Again, numerical precedence was used in cases of ties. Using this strategy they were able to place 99% of their adolescents in one of the code groups they targeted for study, the highest level of inclusiveness found in any of the studies reviewed by McGrath and Ingersoll (1999a). In this study, the code groups defined by the initial coding method were considered the restricted groups, whereas the groups defined after using the Third-Point Replacement rule were considered the unrestricted groups.

Frequency tables were then generated for each high-point code group using each rule. A high-point code was included in subsequent analyses if there were at least 10 individuals in the restricted code group and at least 10 more individuals in the unrestricted group. To meet the second condition, Marks et al.'s (1974) decision to apply the Third-Point Replacement to code groups for which $N < 10$ rule had to be modified. No code groups met both conditions for inclusion in subsequent analyses until the rule was applied to any code group where $N < 45$. For each of the four rules, Table 2 provides the high-point codes that met the criteria for additional consideration.

The first and second authors then reviewed the interpretive statements provided by Greene (1999) for each 2-point code in Table 2. They selected by consensus those SCL–90 scales and HPRS items they expected to correlate with each code. The list was limited to those criteria they expected to be higher in the code group, as predictions of negative relationships were more tentative and could have attenuated the findings. The number of criteria identified for each high-point code may be found in the last column of Table 2. Subsequent analyses focused on comparisons of restricted versus unrestricted high-point codes as predictors of relevant criteria.

RESULTS

Two comparison groups were identified for each code group. The first included all patients in the sample who did not meet the unrestricted criteria for the code. This approach is consistent with the method used to define comparison groups in most of

TABLE 2
High-Point Codes Used to Evaluate Each Rule

Rule	*High-Point Code*	*Restricted Code Group N*	*Unrestricted Code Group Additional N*	*Unrestricted Code Group Total N*	*No. of Expected Correlates*
Minimum Elevation					
	2–4	56	16	72	8
	4–6	38	12	50	8
Validity					
	6–8	40	25	65	7
	7–8	46	16	62	6
Well-Defined Coding					
	1–3	17	15	32	2
	2–3	29	44	73	10
	2–4	26	46	72	8
	2–6	12	31	43	7
	2–7	30	41	71	8
	2–8	13	29	42	11
	4–6	17	33	50	8
	4–8	14	28	42	7
	4–9	13	17	30	7
	6–8	42	23	65	7
	7–8	32	30	62	6
	8–9	10	11	21	10
Third-Point Replacement					
	2–3	73	30	103	10
	2–4	72	24	96	8
	2–7	71	32	103	8
	4–6	50	23	73	8
	6–8	65	24	89	7
	7–8	62	26	88	6

the studies reviewed by McGrath and Ingersoll (1999a). Patients were included in the second comparison group if their profiles did not meet the unrestricted criteria for the code and if all eight clinical scales were < 65 *T*. This second group was similar to the comparison group used by Graham et al. (1999), who noted that a more complete set of the correlates for the codes could be generated if the comparison group consisted of nonpathological individuals. Because they found it impractical to collect MMPIs and clinical data for nonpathological individuals, they used outpatients who generated within normal limits (WNL) profiles as the next best alternative. The resulting effect sizes were on average larger than those found in any of the other high-point code studies examined by McGrath and Ingersoll (1999b), at least in part because of the difference in the comparison groups used. Across subsequent analyses, the size of the WNL comparison groups varied between 34 and 50 patients ($M = 42.3$); comparison groups comprised of all other patients varied between 557 and 676 patients ($M = 618.7$).[1]

Based on the preceding steps, four dichotomous variables were created for each code group: restricted code group versus WNL comparison group, unrestricted code group versus WNL comparison group, restricted code group versus all others, and unrestricted code group versus all others. For each expected relationship between code group and criterion measure a point-biserial correlation was then computed.

The point-biserial correlation has several weaknesses as an indicator of the strength of empirical relationships. To reduce the impact of these weaknesses on the interpretation of the results, two other statistics commonly used to gauge effect sizes were also computed. First, the relationship between the Pearson correlation and the strength of association is generally considered monotonic but not linear (however, see Ozer, 1985). Because much of the analysis depends on comparisons of effect sizes, this issue was addressed by squaring the correlation to generate the coefficient of determination. Second, the point-biserial correlation declines substantially when the dichotomous variable is seriously skewed, as was particularly common when all other patients were used as the comparison group. Cohen's *d* statistic was also computed because it is not systematically affected by skew in the dichotomous variable.

In addition, because prior studies relied heavily on significance tests as the arbiter for identifying important relationships between codes and criteria, the significance of each relationship was also examined. The earliest code studies tended to use an alpha level of .05, but more recent studies have generally used .01 based on a suggestion by Green (1982). In these analyses significance was examined using both standards.

To summarize, there were five pieces of information generated for each combination of code and criterion variables. The first three are commonly used measures of effect size, in the broad sense of this term: the point-biserial correlation, the coefficient of determination, and Cohen's *d* statistic. The last two indicators were whether the relationship significantly differed from zero at .05 and .01 alpha levels.

[1]Other definitions were considered for the comparison groups that might have led to somewhat different outcomes in this study. To be more consistent with prior practice, the comparison groups for each restricted group could have included those individuals in the unrestricted group not included in the restricted group. For example, under the Minimum Elevation rule, unrestricted 2–4s that did not meet the rule would be included in the comparison groups for the restricted 2–4 code group. Second, in the case of the Validity rule, potentially invalid profiles could have been eliminated from the comparison groups as well as for the restricted code groups. However, either of these options would have complicated the interpretation of the findings because of differences in the comparison groups used for the restricted and unrestricted codes.

Table 3 provides summary statistics from these analyses. For each of the four combinations of code and comparison groups described previously, a mean was computed for r, r^2, and d in which each statistic was weighted by the size of the sample on which it was based. The proportion of outcomes significant at each alpha level and mean sizes of the code groups are also indicated.

The first aspect of this table worth noting is how much larger the means for the three effect size measures are when the code groups were compared to the WNL group versus all other patients. Given that this difference is evident for d as well as for the other two statistics, it is probably due more to the differences in the comparison groups than differences in the degree of skew.

The second noteworthy finding is that significance proved to be a poor indicator of the value of more restrictive rules. The proportion of significant outcomes was frequently higher for the unrestricted code group than for the restricted group. This pattern tended to occur even when the mean correlation was larger for the restricted than for the unrestricted codes. Because the values of r and t are closely related, this finding suggests the finding probably reflects the greater power of the unrestricted code tests due to increased sample size.

As expected, more restrictive strategies produced larger mean effect sizes in the majority of cases. However, the differences were often small. They were also often associated with sizeable decreases in the number of cases classified, as indicated by the last row in the table.

To compare information on homogeneity and inclusiveness directly, an additional statistic was generated using the following steps. First, the degree of improvement provided by the restrictive strategy was computed for each of the five measures of association. For example, the degree of improvement in r was computed by:

$$\frac{\text{Mean } r \text{ for restricted codes} - \text{Mean } r \text{ for unrestricted codes}}{\text{Absolute value of mean } r \text{ for unrestricted codes}}.$$

In one case where both the numerator and denominator were zero the degree of improvement was set to zero. A similar formula was used to compute the degree of loss in inclusiveness:

$$\frac{N \text{ for unrestricted codes} - N \text{ for restricted codes}}{\text{Absolute value of } N \text{ for restricted codes}}.$$

The ratio of the first formula to the second was then used to provide information about the relative costs and benefits of using the restricted code. The resulting ratio x can be interpreted using the following guidelines:

TABLE 3
Outcomes for Each Rule

	Minimum Elevation		*Validity*		*Well-Defined Coding*		*Third-Point Replacement*	
	Restricted	*Unrestricted*	*Restricted*	*Unrestricted*	*Restricted*	*Unrestricted*	*Restricted*	*Unrestricted*
Compared to all others								
$M\ r$	0.015	–0.012	0.099	0.122	0.034	0.038	0.045	0.048
$M\ r^2$	0.001	–0.001	0.013	0.018	0.004	0.005	0.006	0.007
$M\ d$	0.054	–0.044	0.423	0.434	0.192	0.148	0.156	0.142
p (.05)[a]	0.000	0.067	0.538	0.769	0.191	0.292	0.340	0.404
p (.01)[b]	0.000	0.000	0.462	0.615	0.124	0.169	0.191	0.298
Compared to WNL								
$M\ r$	0.309	0.257	0.341	0.353	0.293	0.291	0.317	0.303
$M\ r^2$	0.137	0.094	0.197	0.190	0.134	0.131	0.143	0.133
$M\ d$	0.704	0.527	0.879	0.806	0.731	0.570	0.752	0.675
p (.05)[a]	0.533	0.600	0.769	0.692	0.551	0.584	0.702	0.723
p (.01)[b]	0.467	0.400	0.538	0.615	0.360	0.494	0.574	0.638
$M\ N$	42.750	54.813	39.385	58.385	19.000	46.022	59.872	84.383

Note. If r and d were negative, r^2 was also averaged as a negative number. WNL = within normal limits comparison group.
[a]Proportion of correlations that were significant at .05. [b]Proportion of correlations that were significant at .01.

$x > 1.0$: The improvement in strength of association provided by using the rule outstripped the loss in inclusiveness.
$x = 1.0$: The improvement in strength of association matched the loss in inclusiveness.
$0 < x < 1$: The improvement in strength of association was exceeded by the loss in inclusiveness.
$x = 0$: There was no improvement in the strength of association.
$x < 0$: The rule reduced the strength of association.

Values for x are provided in Table 4.

COSTS AND BENEFITS

The results for the significance tests will be considered first. The finding that most of the ratios were negative even when the ratios for the effect size indicators were positive again highlights the limitations of significance as the basis for making judgments of this type. This is in itself an interesting finding because previous high-point code studies have relied so heavily on significance testing alone as an indicator of important associations. The vagaries of power limit the value of comparing

TABLE 4
Cost-Benefit Statistics

	Minimum Elevation	*Validity*	*Well-Defined Coding*	*Third-Point Replacement*
Compared to all others				
$M\ r$	7.98	–.41	–.07	–.15
$M\ r^2$	7.09	–.60	–.14	–.35
$M\ d$	7.90	–.05	.21	.24
M	7.66	–.35	.00	–.09
p (.05)[a]	–3.55	–.65	–.24	–.39
p (.01)[b]	0.00	–.54	–.19	–.87
Compared to WNL				
$M\ r$	0.72	–.07	.01	.11
$M\ r^2$	1.62	.08	.02	.18
$M\ d$	1.19	.19	.20	.28
M	1.18	.07	.08	.19
p (.05)[a]	–0.39	.23	–.04	–.07
p (.01)[b]	0.59	–.26	–.19	–.24

Note. Rows labeled "M" provide the means of $M\ r$, $M\ r^2$, and $M\ d$. WNL = within normal limits comparison group.

[a]Proportion of correlations that were significant at .05. [b]Proportion of correlations that were significant at .01.

proportions of significant outcomes as a basis for evaluating different approaches to high-point coding.

Putting aside the statistics for significance tests, the comparison of costs to benefits offered clear support only for the Minimum Elevation rule. The six x values for r, r^2, and d associated with the rule are the largest in the table, and five of the six exceed 1.0. Looking at Table 3, the Minimum Elevation rule on average excluded fewer cases than any other rule but was consistently associated with increases in mean validity. This combination of outcomes supports the conclusion that the positive outcomes outweigh the negative outcomes when high-point code interpretation is limited to profiles in which both code scales exceed 64 *T*.

Five of the six x values were positive for Well-Defined Coding ($M = .07$), while four of six were positive for Third-Point Replacement ($M = .05$). Remember that the latter rule is less restrictive than the original coding; hence, larger values in the table argue against using the Third-Point Replacement rule. It seems that with regard to these two rules the decision might best be left to the clinician. Although there is likely to be an enhancement in predictive power resulting from limiting code interpretation to well-defined codes or not using Third-Point Replacement, these are more than offset by the loss in coverage. On the other hand, this loss is not associated with a very high cost. Even if a profile is not considered appropriate for high-point code interpretation because it is poorly defined, it can still be interpreted on a scale-by-scale basis. To date, there has never been a study published comparing the validity of interpreting MMPI profiles based on high-point code versus individual scale elevations, but there is not a prior reason to presume the latter is markedly less valid than the former. The clinician who appreciates the additional reliability offered by limiting high-point code interpretation to well-defined codes might be considered justified in this decision. At the same time, a case can be made against modifying standard practice in this way if inclusiveness in a high-point code interpretive strategy is valued highly.

The implications of the loss of inclusiveness are even more benign in the case of not using the Third-Point Replacement rule. In this case, the profile can still be coded. However, the resulting code is more rare, and therefore probably has been studied less frequently, than the code to which Third-Point Replacement would have reassigned the profile. These findings do not provide a particularly strong case for incorporating Third-Point Replacement into standard high-point coding practice.

The circumstances are very different for the Validity rule. The findings indicate using the validity criteria generally reduced effect sizes. Only two of six x values were positive. The largest positive value was .20, that is, the improvement in mean effect size was only 1/5th the mean loss in code group size. Furthermore, there is an important difference in the implications of applying the Validity rule versus the other three rules. If code classification is only allowed when a minimum elevation or a minimum difference from other scales occurs, the profile not meeting the criterion is still interpretable. The only consequence is that high-point code interpretive data cannot be applied to the protocol. In contrast, failure to meet validity criteria is taken as evidence the profile should be considered uninterpretable, an outcome that should be associated with a very high cost. These results do not support the use of validity criteria in this type of setting.

DISCUSSION

There are several statistical points worth noting about these results. First, the inclusion of multiple measures of effect size seemed helpful. Given that the values in Table 4 sometimes changed markedly across statistics, the ability to examine global trends made for firmer conclusions about the relationship between code rule and effect size. Second, the differences in conclusions one would draw from significance tests versus effect size indicators suggests that the frequency of significant outcomes is not an acceptable basis for comparing different approaches to high-point coding. The information about significance tests provided in the tables is included for the sake of completeness; it should not be used to make judgments about the comparative value of alternate coding strategies.

If one considers the outcomes associated with using each of these rules, as well as the appropriate values to be assigned to each of those outcomes, a fairly clear set of clinical guidelines emerge from this study. First, application of the Minimum Elevation rule was clearly supported. The benefit in terms of enhanced predictive validity more than offsets the cost in reduced inclusiveness. Second, application of the Validity rule was clearly not supported, at least in circumstances in which there is no clear and consistent motivation for distorting results. Third, there is no strong justification for modifying standard practice by the inclusion of the Third-Point Replacement rule. Finally, the case for incorporating the Well-Defined Coding rule into standard practice is also equivocal.

The results for the Validity rule may in some ways be the most important, as they contradict standard clinical practice. The finding is surprising in at least two ways. First, item omission should be expected to dampen validity coefficients simply by reducing the variance in scores. However, there were only six individuals who omitted more than 30 items from the MMPI, so item omission was not a substantial contributor to exclusion by the Validity rule. The findings should not be taken as evidence that patients can omit more than 30 items and still generate an acceptable profile.

Second, there is compelling evidence from several meta-analyses that the validity scales are effective indicators of inaccurate self-representation (Baer, Wetter, & Berry, 1992; Berry, Baer, & Harris, 1991; Rogers, Sewell, & Salekin, 1994). This inconsistency may have something to do with context. The meta-analyses were based on studies in which participants had a clear motivation to distort. In such situations, those motivating factors may largely determine variability in the scores on the validity scales. This study was conducted using psychiatric inpatients after hospitalization, where there was little clear motivation to fake results. In such circumstances responses to validity scale items could be largely determined by the nature of the patient's psychopathology, so that removing so-called invalid protocols would actually attenuate concurrent validity. This finding suggests that, although the validity scales can be very important when the motivation to distort exists, more research is needed on the value of the validity scales in settings where the contextual motivation to distort is trivial.

ACKNOWLEDGMENTS

We thank Samuel Klagsbrun, Martin Buccolo, Janet Segal, and the administration of Four Winds Hospital, Katonah NY, for their support in the collection of data for this article. The authors are solely responsible for its content.

REFERENCES

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.

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.

Derogatis, L. R. (1983). *SCL–90–R administration, scoring, and procedures manual*. Towson, MD: Clinical Psychometric Research.

Friedman, A. F., Lewak, R., Nichols, D. S., & Webb, J. T. (2000). *Psychological assessment with the MMPI–2.* Mahwah NJ: Lawrence Erlbaum Associates, Inc.

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., 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. (1999). *The MMPI–2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

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

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

McGrath, R. E., & Ingersoll, J. (1999a). Writing a good cookbook. I: A review of MMPI high-point code studies. *Journal of Personality Assessment, 73,* 149–178.

McGrath, R. E., & Ingersoll, J. (1999b). Writing a good cookbook. II: A synthesis of MMPI high-point code study effect sizes. *Journal of Personality Assessment, 73,* 179–198.

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.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

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

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.

Robert McGrath
School of Psychology T–WH1–01
Fairleigh Dickinson University
Teaneck NJ 07666
E-mail: mcgrath@fdu.edu

Received April 17, 2001
Revised September 15, 2001