The Research Base, Psychometric Properties, and Clinical Uses of the MMPI-2 and MMPI-A

JAMES N. BUTCHER
University of Minnesota
KENNETH S. POPE
Los Angeles, California

Abstract
The Minnesota Multiphasic Personality Inventory (MMPI) has become the most widely used personality assessment instrument and is used in a broad variety of clinical assessment settings. In the decades since it was introduced, problems with the instrument, such as the datedness of the item wording, problems with some items, and narrowness of the norms, have become apparent. After considerable evaluation and discussion, these problems led to a recent revision and expansion of the instrument in order to eliminate problems and to develop a more sound psychometric instrument for clinical assessment. This article describes the revision of the instrument and summarizes some of the main features of the MMPI-2.

First, published critiques have noted problems at the item level. The sexist wording and assumptions of various items in the original instrument have been offensive to many and might introduce bias (influencing, for example, test-taking attitude and the rapport and trust established — or thwarted — between client and clinician). Other items may unintentionally embody different (e.g., racial) forms of bias, prejudice, or discrimination. Erdberg (1988), for example, reports that in one research study of a rural population, one MMPI item, taken by itself, perfectly discriminated all Black test-takers from all Caucasian test-takers. Still other items have become increasingly archaic. For example, a childhood game popular over fifty years ago may be unknown to most people taking the test in the 1980's. Still other items are objectionable to some test-takers (e.g., these items are offensive to those of certain religious faiths).

Second, critiques have noted problems with the norms. The normative sample, collected in the late 1930's, on which the MMPI scales were developed, did not include those from Asian, Black, Hispanic, Native American, and various other ethnic groups, and were unrepresentative of the American population in other significant ways. Furthermore, Colligan et al. (1983) found that the norms of the 1930's are so dated that they are inappropriate for use with contemporary individuals: "These MMPI changes are not only statistically significant but are also of clinical importance" (p. xv). Faschingbauer (1979) summarized some of the problems with norms in vivid terms:
The original Minnesota group...seems to be an inappropriate reference group for the 1980s. The median individual in that group had an eighth-grade education, was married, lived in a small town or on a farm, and was employed as a lower level clerk or skilled tradesman. None was under 16 or over 65 years of age, and all were white. As a clinician I find it difficult to justify comparing anyone to such a dated group. When the person is 14 years old, Chicano, and lives in Houston's poor fifty ward, use of original norms seems sinful. (p. 375).

Third, critiques have noted psychometric problems in the customary ways in which the T-scores are interpreted. Greene (1980), for example, called attention to inconsistencies among various raw score and T-score tables. The T-scores on the standard profile sheet are based upon the purified sample; the T-scores based on norms provided in the Handbook and most of the other MMPI references, however, were derived from the original sample. Greene (1980) notes that "this discrepancy can be the source of considerable confusion" (pp. 21-22). As another example, Colligan et al. (1983) noted the widespread misconception that the profile sheet's standard scores represented normalized T-scores (i.e., that they are based upon the normal probability curve) rather than a linear transformation. The linear transformation maintains the considerably skewed distribution of some raw scores. "Thus, scale elevations of similar degree in terms of T-scores unfortunately, do not have equivalent meanings in terms of deviation from the mean (this, is, percentile rank)" (pp. xiv-xv).

When psychologists become aware that an assessment instrument's content and wording embody sexist assumptions, that the selection process for normative reference samples was racially biased (even if such discrimination was unintentional), or that similar problems exist with the instrument and its empirical base, the need to address these issues promptly and rigorously and to implement corrections is a clear clinical, scientific, and ethical mandate (Pope & Johnson, 1987; Pope & Vasquez, 1991).

Formation of The MMPI Revision Committee
Some people have asked why it has taken so long to revise the MMPI. A number of factors prevented the instrument from being revised earlier. Of course, the cost of revising a major instrument like the MMPI is staggering and no provision for revision had been made in the original publication arrangements. Since the MMPI is an owned and copyrighted instrument, only the owner or a designated party can alter the test. The test authors themselves did not participate in the revision since one of the test authors (McKinley) was deceased and the other (Hathaway) had retired in the early 1970s.

The most likely reason why the MMPI was never revised in the past is that it was widely used and performing well. Why change something that is so successful? Given its continued success, it was apparent that any revision of the MMPI had to include the goal of keeping some aspects of the instrument intact, at least the valid and still appropriate aspects, while making necessary changes and launching an expansion of the scope of the instrument.

Once the editorial staff of the University of Minnesota Press, the copyright holder, became convinced that a revision was needed they appointed a committee composed of James Butcher, John R. Graham, W. Grant Dahlstrom, and Auke Tellegen to conduct the restandardization study and develop new norms for the MMPI-2. The project, initiated in 1982, was funded by the University of Minnesota Press out of general revenues from the MMPI and from the Minnesota Report, a computer scoring and interpretation service distributed by National Computer Systems.
NCS collaborated in the restandardization by providing test scoring and data processing.

**Goals of the MMPI Restandardization**

*First,* the Committee members unanimously agreed that the revision of the MMPI should be a "conservative" one in that the basic measures, on which the MMPI's reputation has been built — the validity and clinical scales — should be kept relatively intact in order to preserve the half century of research supporting their use. The items making up those scales, except a few objectionable items of a few scales, were kept in the revised instrument in order to maintain continuity with the original instrument.

Second, the Committee agreed to collect extensive normative and clinical data (using an expanded 704 item booklet containing all original items plus 154 new items) to evaluate and justify any changes made to the instrument and to serve as information to be used in the expansion of the MMPI.

With these two basic premises for the revision, the Committee then established a number of goals in order to:

a) Develop a new, broadly representative, normative sample for use for improving the MMPI norms and for serving as a more relevant population for new scale development.

b) Modernize the MMPI booklet by deleting objectionable, non-working, or obsolete items and expanding the item pool to include additional items addressing contemporary clinical problems such as treatment readiness, suicide, alcohol and drug problems, and so forth. The goal of keeping the final MMPI-2 booklet at about the same number of items as the original was accomplished. The MMPI-2 contains 567 items; the clinical and validity scales, however, can be obtained by administering the first 370 items.

c) Maintain the continuity of the original validity and clinical scales by keeping them intact, that is, comprising the same items as the original scales.

d) Develop new MMPI norms that would better reflect clinical problems and would resolve the problem of non-uniformity in percentile classification.

e) Collect new clinical data for evaluating changes that needed to be made in the items and scales.

f) Develop new scales that address additional clinical problems that were not covered in the original MMPI.

g) Develop an additional form of the MMPI that would be appropriate for use with adolescents. The MMPI-A, which has new norms for adolescents aged 14 through 18, includes items from the original validity and clinical scales as well as a number of new items specific to adolescent problems.

**The MMPI-2 Normative Sample**

The MMPI-2 normative sample comprises 2600 individuals (1,462 women and 1,138 men) who were selected from seven regions of the United States (California, Minnesota, North Carolina, Ohio, Pennsylvania, Virginia, and Washington). Efforts were made to balance the normative sample according to gender, age, ethnic group membership, education, and place of residence. For example, 2.4% of the male sample and 2.1% of the female sample were drawn from among Native American adults residing on a federal reservation. The final normative sample matches the demographic mix of the national population well on most variables.

The MMPI-2 normative sample is more appropriate for contemporary subjects because it is balanced for demographic variables such as ethnic group membership. The new normative sample reflects a higher educational level than the dated original sample. It is important to examine this shift carefully from two perspectives: historical and psychometric.
Fig. 1 Group mean MMPI-2 profiles for males summarizing scores for each level of education found in the MMPI-2. Restandardization sample.
Historically, there has been an increase in the educational level of the population since the 1930's (see Butcher, 1990c). The average education of the United States population is currently over 13 years (i.e., high school plus around one year of college) whereas the mean education of the population in the 1930's, when the original MMPI norms were collected, was about ninth grade. Thus, the new MMPI-2 norms more closely match the educational backgrounds of individuals taking the test today than do the original norms.

Psychometrically, however, the educational level of the individual taking the test does not greatly influence the MMPI-2 scores. Inspection of the group mean profiles of the MMPI-2 normative subjects from different levels of education in Figures 1 and 2 clearly shows that the average profiles are almost indistinguishable across the five educational levels (Butcher, 1990a). Interpretation of profiles of individuals from different educational backgrounds does not require the special considerations needed with the original MMPI. The "mental adjustments" for education — adjustments that were not well calibrated empirically and thus weakened the actuarial or empirical basis of the instrument — that needed to be made when interpreting the original MMPI Mf and K scores do not need to be made when interpreting those two scores on the MMPI-2. The only cautions needed are when interpreting Mf scores of men with very low or very high educational levels.

Individuals were randomly solicited, initially contacted by letter, and asked to come to a prearranged testing site for completion of the test battery. In addition to the 704 item experimental version of the MMPI (Form AX), all individuals were administered a biographical history questionnaire and a life events questionnaire. In addition, a portion of the sample (822 couples) were given the Dyadic Adjustment Questionnaire (Spanier & Filsinger, 1983) and asked to complete a personality rating form on their spouse. This information has provided valuable data as initial validity correlates for the MMPI-2.

Why Does Today’s Normative Group Appear to Have More Pathology than the Original Minnesota Normative Group? When individuals in the contemporary normative group are given the MMPI and scored on the original MMPI norms, their scores are typically elevated on all of the clinical scales at about a half standard deviation above the mean. Does this mean that people today are more afflicted with pathological states, traits, or symptoms than people were fifty years ago when the test was originally standardized? Have people today deteriorated in their mental health compared to those from the last generation?

Actually, a comparison of item response differences between today’s normative sample and the original MMPI normative sample does not show many consistent item response differences to support an interpretation of increased psycho-pathology in the society at large.

Why then are there MMPI scale differences between the original Minnesota normative group and a contemporary sample of individuals drawn from the community? One contributing factor is that the instructions that the original standardization sample followed in taking the test were different from those we employ in administering the instrument today. Hathaway and McKinley (1940) allowed test-takers to omit items of which they were unsure or which they felt did not apply to them. Many individuals omitted a considerable number of items, typically more than 30 items. In practice today, we encourage those who take the inventory to try to answer all of the items. This results in the endorsement of more items that are scored on the clinical scales, hence, more scale elevation on average.
Fig. 2 Group mean MMPI-2 profiles for females summarizing scores for each level of education found in the MMPI-2. Restandardization sample.
The new MMPI-2 normative responses were collected using the same instructional set used in clinical practice. The individuals in the normative group were encouraged to try to answer all items. Most individuals omitted fewer than 2 items.

The MMPI-2 Norms
Originally, Hathaway and McKinley (1940, 1943) developed the MMPI norms using a linear T-score transformation. As noted earlier, scores were assigned a mean of 50 and a standard deviation of 10, even though the underlying distributions were somewhat skewed.

Other researchers (Colligan, et al., 1983) have recently recommended the use of a different transformation approach: normalized T-scores to compensate for the skewed scale distributions. This transformation approach, however, produces T-score distributions that are quite different from the original Hathaway and McKinley T-scores. This has the effect of pulling all scores toward the mean and not allowing outlying scores to appear as "visually extreme" on the profile as they would in the original Hathaway-McKinley T-scores.

Graham & Lilly (1986) showed that normalized T-scores produce MMPI distributions that are not comparable to the linear Ts in the original MMPI. Actually, there is nothing "wrong" with normalizing these MMPI scale distributions except that clinicians are not accustomed to interpreting such low ranging distributions and would likely not find the normalized distributions as easy to interpret as the original MMPI T-scores developed by Hathaway and McKinley.

The MMPI Restandardization Committee chose to follow, to some extent, Hathaway's linear T-score approach, except that some modification was made in order to make the distributions for the MMPI scales uniform with regard to percentile value. That is, the percentile rank would be equivalent for each T-score level for each scale. The MMPI-2 T-scores, referred to as uniform T-scores, were developed by using the eight clinical scales to form a composite distribution.

For the eight clinical scales, separated by gender, a transformation was derived for converting each scale's raw scores into the corresponding uniform T-scores. This was achieved by regressing raw scores on percentile-equivalent uniform T-scores. Uniform T-scores were developed for the 8 clinical scales and the 15 MMPI-2 content scales. Tables for converting these raw scores into uniform T-scores are provided in the MMPI-2 Manual (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989).

Interpretation of MMPI-2 Scores: A Shift in Level of Clinical Significance
When interpreting profiles with the original MMPI, clinicians usually considered a T-score of 70 to be the point at which clinically significant elevation was obtained, that is, the clinical range of scores began at the T-score of 70. Theoretically, this T-score level fell at a percentile score of 95, although as we have seen, this did not occur in practice since percentile equivalents for a given T-score were not uniform.

With MMPI-2, the 92nd percentile mark falls at a T-score of 65 on all of the clinical scales. Consequently the MMPI-2 Committee has made the recommendation that a T-score of 65 or greater demarcate the "clinical range" (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989). In clinical practice, the 65 T-score level appears to be optimal for separating known clinical groups from the MMPI-2 normative sample.

In order to illustrate the importance of the 65 T-score cut off score for the clinically interpretable range, data from two studies are presented in Figures 3 and 4. Keller & Butcher (1991) found that the optimal separation for chronic pain...
patients to be at about the 65 T level on the Hs scale (see Figure 3). Butcher (1989b) found the 65 T-score to be a valuable cut off score for separating depressed inpatients from the normative sample (see Figure 4).

What do the MMPI-2 Validity and Clinical Sales Measure?
The MMPI-2 clinical and validity scales are virtually identical to the original scales and therefore they measure what they have always measured. Recently a psychologist, using the MMPI-2 Pd scale in a study, provisionally concluded that the scale did not seem to measure "Pd behavior" in the same way as the original MMPI Pd scale. His preliminary conclusion was that the scale was possibly not as "valid" as the original Pd scale. This criticism is based on a misunderstanding of the MMPI revision. The Pd scale measures "Pd behavior" in exactly the same way as the original scale did since it is exactly the same scale. No items have been dropped and no items added.

If one makes a comparison of the original MMPI Pd raw scores and the MMPI-2 raw scores, the correlates of differences will be exactly the same. Most statistical comparisons in research studies employ raw scores for psychometric analyses (Butcher & Tellegen, 1978).

There is a difference, however, in the way in which this raw score is interpreted through comparing the scores with the new T-score distributions based upon the MMPI-2 norms. There will be a difference in the elevation of the Pd score for a given individual if the score is compared to the new versus old norms. The same raw score will produce a different T-score elevation in the old versus the new norms. This change in relative elevation using new T-scores comes about by the use of more contemporary reference groups which, as we have seen, is more appropriate for contemporary comparisons.

It is possible for practitioners to examine profiles of individuals on the original norms who were tested on the MMPI-2
version of the instrument. The MMPI-2 manual (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989) allows for the transformation of MMPI-2 raw scores into original T-scores. This transformation procedure will be of value in making comparisons on MMPI norms for individuals who were tested on the MMPI-2.

Reliability of the MMPI-2 Validity and Clinical Scales
Data were collected on the test-retest reliability of the traditional validity and clinical scales (Butcher, Dahlstrom, Dahlstrom, Graham, Tellegen & Kaemmer, 1989). An average of 8.58 days (median of 7 days) elapsed between the first and second administration. The length of the interval produced no trends. For male adults, the retest coefficients ranged from $r = 0.67$ (standard error of measurement = 1.63) for Paranoia to $r = 0.89$ (standard error = 2.24) for Psychasthenia. For female adults, the retest coefficients ranged from $r = 0.58$ (standard error = 1.98) for Paranoia to $r = 0.91$ (standard error = 2.86) for Schizophrenia.

Congruence Between MMPI and MMPI-2 Scale Scores and Profile Codes
A natural question of anyone who has grown accustomed to the MMPI and perhaps has baseline MMPI data in his or her patient or research files is: Do the MMPI-2 scores and codetypes have a low congruence with the original MMPI variables? The extensive research has shown a high congruence.

First, it should be remembered that the MMPI validity and clinical scales have been preserved relatively intact in the MMPI-2. Most of the scales contain exactly the same items as the original MMPI scales. Ben-Porath and Butcher (1989), in a test-retest study in which participants were given either two administrations of the original MMPI or administrations of both the MMPI and the MMPI-2, found that MMPI-2 scales have the same degree of relationship to the original MMPI scores as the original MMPI scales have.
with themselves in subsequent administrations of the test. The MMPI-2 has been shown to be a highly reliably measure of the original MMPI variables. (It is obviously important that any such comparisons utilize adequate comparison groups and procedures. An experimentally naive researcher, for example, might simply administer an MMPI and then, after a brief period, administer the MMPI-2 for comparison. Such research would leave uncontrolled or unexamined such confounding variables as test-retest effects, order effects, statistical regression, etc.).

Recently, Graham and Ben-Porath (1990) found that the MMPI-2 scales and profile codes are highly congruent (94% for men and 95% for women in the restandardization sample; 82% for men and 94% for women psychiatric patients) with the original MMPI scales and profile codes if they are well defined (i.e., if the scales or codes have a difference of 5 T-score points between the next higher score). Graham and Ben-Porath further found that in a psychiatric sample only about 14% of profile codes change when scored on the two different norms. Even when the profile codes are comprised of different scales, at least 91% had the same highest scale in the two codes, indicating that a congruent interpretation is likely to be made with the MMPI-2 and MMPI codes.

Finally, Graham and Ben-Porath (1990) found that in the few cases where a different MMPI and MMPI-2 code emerges, validity data (from personality rating studies) indicate that the MMPI-2 code tends toward being more accurate in prediction than the original MMPI code.

New MMPI-2 Content Scales
Assessing content themes has become an important part of clinical MMPI interpretation over the past twenty years. Clinicians find that homogeneous content scales are relatively easy to interpret and to explain to others (Bursich, 1984). Some content measures from the original MMPI, for example the Harris-Lingoes Scales or the Koss-Butcher Critical Items, are available, virtually intact, in the MMPI-2. Others, particularly the Wiggins Content scales, are not available in MMPI-2 since many of their items were deleted because they contained objectionable content and the Wiggins scales are no longer representative of the MMPI-2 item pool since new items were incorporated.

A new set of MMPI-2 content scales was developed by Butcher, Graham, Williams, and Ben-Porath (1989) to assess the main content dimensions in the MMPI-2. The scales were derived by a multimethod, multistage strategy involving both rational and statistical procedures to assure rational content relevance and strong statistical properties. The 15 content scales assess important personality factors (e.g., Antisocial Practices or Obsessiveness), symptomatic behavior (e.g., Depression, Anxiety) or address important clinical problem areas (e.g., Family Problems, Negative Treatment Indicators).

The 15 scales are:

1. Anxiety
2. Fears
3. Obsessiveness
4. Depression
5. Health Concerns
6. Bizarre Mentation
7. Anger
8. Cynicism
9. Antisocial Practices
10. Type A
11. Low Self-Esteem
12. Social Discomfort
13. Family Problems
14. Work Interference
15. Negative Treatment Indicators.

The MMPI-2 Content Scales have been shown to have considerable appeal for clinical practice since they cover a broad range of problems and are intuitively understandable. In addition, they show strong internal psychometric characteris-
tics and external validity. The external validity of the MMPI-2 Content Scales have been shown to be equal to or greater than the original MMPI clinical scales (Butcher, Graham, Williams, and Ben-Porath, 1990).

New Validity Measures in the MMPI-2
Assessing the validity and applicability of a particular patient's profile has always been viewed as an important facet of profile interpretation. Appraisal of the validity scales must be done to assure the adequacy of the individual's self report. The traditional MMPI validity scales (Cannot Say, L, F, and K) have been maintained in the MMPI-2 and operate in the same manner (Graham, Watts & Timbrook, 1991).

In addition, several other measures have been incorporated to assess more adequately the individual's test taking attitudes. These scales are:

THE F(B) SCALE
Since all of the F scale items of the original MMPI appear in the first 370 items in the booklet, there is not a measure of symptom exaggeration for items toward the end of the item pool. An additional infrequency measure, the F(B) scale was developed for the MMPI-2 to detect possible deviant responding to items located toward the end of the item pool. The 40 item F(B) Scale was developed following the same procedures as the original F scale, that is, by including items that had low endorsement percentages in the normative sample. This scale is particularly valuable in interpreting the MMPI-2 Content Scales that require valid response to the full item pool.

MMPI-2 CONSISTENCY SCALES
In addition to the original MMPI validity scales (L, F, and K) which enable the practitioner to assess client motivation to distort responses, two new scales have been developed for the MMPI-2 to assess inconsistency of responding. These scales, True Response Inconsistency (TRIN) and Variable Response Inconsistency (VRIN), assess the extent to which the individual has endorsed semantically related items in a consistent fashion. For example, if the following two items were both marked true or were both marked false, the individual's responses would be inconsistent: "Most of the time I feel blue" and "I am happy most of the time." The response in consistency scales will enable the clinician to determine if the individual has answered in a non-content oriented manner (e.g., using a random response set). VRIN raw scores exceeding 12 and TRIN raw scores of less than 6 or more than 12 form rough criteria for significantly inconsistent responding (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989).

CONTROVERSIAL "SUBTLE-OBVIOUS" SCALES
In the past, some investigators have relied upon the Weiner-Harmon (Weiner, 1948) "subtle" scales to assess test validity (Greene, 1980). These scales have been somewhat controversial and a number of researchers have cautioned against reliance upon them (Graham, 1987). Scoring keys for the Weiner-Harmon subtle scales are available in the MMPI-2 for those who plan to continue to research them, although the MMPI Restandardization Committee was not unanimously supportive of their inclusion (Butcher, Dahlstrom, Graham, Tellegen, and Kaemmer, 1989).

Research employing the MMPI-2 normative data and a clinical study of couple's in therapy (Hjemboe & Butcher, 1991) has shown that the subtle scales are poor psychometric instruments in that they have low alpha coefficients compared to those of the obvious items and the full MMPI score (Butcher, 1989a). Moreover, Weed, Ben-Porath, and Butcher (1990) have shown that the subtle items failed to predict external correlates (spouse behaviour ratings). The obvious items on the MMPI scales showed higher external
validities than even the full MMPI score. This study showed that the subtle items actually reduce the validity of the clinical scales.

THE REVISED ADOLESCENT FORM OF THE MMPI (MMPI-A)

Even though the original MMPI was developed for use with adults, it has been one of the most popular instruments for assessment of adolescents. Numerous studies detail the effectiveness of the instrument with adolescent populations, for example: In psychiatric settings (Archer, Gordon, Giannetti & Singles, 1988; Dudley, Mason & Hughes, 1972; Marks, Seeman & Haller, 1974), medical settings (Colligan & Osborne, 1977); correctional settings (Gold & Petronis, 1980; Hathaway & Monachesi, 1963), and drug and alcohol settings (Wisniewski, Glenwick & Graham, 1985; Wolfson & Erbaugh, 1984). Moreover, Hathaway and Monachesi (1963) conducted an extensive study examining the performance of the MMPI with normal teenagers in a public school setting.

Several limitations have been noted for the use of the MMPI with adolescent populations. The items on the original MMPI were oriented toward adult subjects both in terms of the wording of the items and in the range of relevant content; few scales were developed specific to adolescent problems; broad base norms for the MMPI clinical scales have not been available; and interpretive information has been limited and mostly involves a downward extension of adult interpretive strategies.

The MMPI Restandardization Committee initiated a revision and expansion of the MMPI for adolescents in 1982. The MMPI item pool was revised and broadened by incorporating new adolescent specific items. New adolescent norms have been collected on 805 boys and 815 girls from several states: Minnesota, Ohio, North Carolina, California, Pennsylvania, Virginia, Washington State, and New York and the samples were balanced for age and ethnic group membership. The revised version of the MMPI for adolescents (MMPI-A) contains 478 items with objectionable and adolescent-irrelevant items removed (Butcher, Williams, Graham, Archer, Tellegen, Ben-Porath & Kaeammer, in press). The validity and clinical scales have been kept virtually intact in MMPI-A; however, Mf and Si have been reduced in length.

The MMPI-2 content scales (Butcher, Graham, Williams & Ben-Porath, 1990) have been expanded and validated for use with adolescent populations. The MMPI-A content scales that parallel the MMPI-2 content scales are: Anxiety (anx), Obsessiveness (obs), Depression (dep), Health Concerns (hea), Bizarre Mentation (biz), Anger (ang), Cynicism (cyn), Low Self Esteem (lse), Social Discomfort (sod), Family Problems (fam), and Negative Treatment Indicators (TRT). In addition, four new content scales were developed that focus upon specific adolescent problems. They are Conduct Problems (con), Alienation (ahn), Low Aspirations (las), and School Problems (sch). Although these scales were developed according to a combined rational and statistical scale development strategy, they were found to have strong external validities when evaluated against external correlates (Child Behavior Checklist, Deveraux Adolescent Behavior Rating Scale, and clinical behaviors) in a clinical validation study (Williams, Butcher, Ben-Porath & Graham, in press).

The MMPI-A is likely to be useful to adolescent clinical assessment programs since the inventory is shorter than the original instrument, contains more relevant adolescent content, and incorporates several new scales that focus upon adolescent problems in addition to the traditional validity and clinical scales.
Clinical and Ethical Issues
As with any assessment instrument, there are pitfalls likely to plague the inadequately trained or careless clinician. Unfortunately, the process of assessment itself forms the basis of a disappointingly significant percentage of the malpractice, ethics, and licensing complaints against psychologists (Pope, 1989a). Pope and Vasquez (1991) discuss some of the more general problem areas. In this article, we will highlight a few issues directly relevant to the MMPI-2, although they are relevant also to the more general use of psychological tests.

Keeping Pace with Developments
Psychology takes justifiable pride in itself as an empirically-based profession. Approaches to clinical work are constantly subjected to research and are validated, refined, or discredited. Those who enter the field assume a crucial though sometimes burdensome responsibility to keep abreast of empirical findings relevant to their areas of practice. This professional responsibility is especially significant in the field of testing in which data regarding norms, validity, reliability, and applicability are constantly redefining our understanding of effective assessment. Erdberg (1988), Fowler (1988), and Weiner (1988), for example, have each stressed that anyone whose knowledge of assessment instruments and procedures has not been significantly updated in the last five years is practicing in an incompetent and unethical manner.

Clinicians utilizing the MMPI need to obtain a solid working knowledge of the current norms and psychometric properties as well as the evolving empirical base of this inventory. Continued use of a dated version of a test—particularly one involving such problems as sexist wording of items and the exclusion (even if unintentional) of racial minorities from the normative sample—seems hard to justify scientifically, clinically, and ethically when a revised version that addresses these problems is available.

The Required Reading Level
Although the inventory provides internal checks to ensure that this criterion is met, clinicians should ensure that individuals to whom they administer the test currently have at least a sixth grade reading level. The method of reading level determination originally used in the evaluation of the MMPI-2 items suggested that an eighth grade reading level was needed (Stenner, Horabin, Smith & Smith, 1988). This conclusion has recently been questioned, however, and other reading level determination programs have suggested that a 6th grade reading level is more appropriate.

Careful Administration
It is tempting for any of us, as busy clinicians, to look for short cuts in the testing process. This inventory, however, requires the same rigorous care appropriate to any standardized instrument. Supervision of the testing process is especially important. Among the reasons for carefully monitoring a client’s completion of the inventory—rather than, for example, allowing the client to take it home and complete it whenever he or she finds time—are the following.

First, individuals both in the original MMPI normative sample and in the current version’s contemporary normative sample took the test under carefully monitored conditions. Altering the administration process can significantly influence the results of a standardized test, the power of which is dependent upon maintaining the standard format of administration. As Faschingbauer (1979), for example, wrote:

We know, for example, that the MMPI can provide false information when the usual mental health testing environment is altered. Hence, we were taught as students to be
concerned about the settings in which tests are administered. Some of us had this point personally accentuated when we sneaked off to take the MMPI in our offices only to be startled by the unusually high elevations this private setting produced. To be reliable and valid any test should be administered in a setting close to that employed in the normative studies. (p. 384).

Second, in reporting assessment results psychologist should explicitly address testing circumstances that may affect the validity or reliability of the test results. How could psychologists who fail to monitor the circumstances under which the client completes the inventory fulfill this obligation?

Third, clients who are directed to fill out the inventory in an unsupervised setting may be influenced by others while they are pondering the questions. In some cases, test data obtained in the ordinary course of clinical work may later become crucial (e.g., as baseline data establishing the client's condition prior to a trauma) in civil or criminal litigation. If the psychologist cannot establish clearly that the assessment was monitored, the data may be discredited or excluded and the client may be deprived of acceptable and reliable clinical information to which he or she is entitled.

Fourth, when psychologists fail to monitor the administration of a standardized test such as the MMPI, they are violating the published finding of APA's Committee on Professional Standards (1984). When a complaint was filed with the Committee regarding a psychologist who had given his client an MMPI to take home, the Committee found that whenever a psychologist does not have direct, first-hand information as to the condition under which the test is taken, he or she is forced (in the above instance, unnecessarily) to assume that the test responses were not distorted by the general situation in which the test was taken (e.g., when the client consulted others about test responses). Indeed the psychologist could have no assurance that this test was in fact completed by the client. In the instance where the test might be introduced as data in a court proceeding it would be summarily dismissed as hearsay evidence. (p. 664).

One of the authors observed a patient taking the MMPI in an outpatient waiting room while the psychologist worked in his office. Frequently when the patient marked down a response, the patient's spouse, who was reading along, commented, "now that's not you!" or "That's not what you believe...Change that answer!" The patient would re-read the item, reconsider, and they dutifully change the answer.

John R. Graham (personal communication) reported an intriguing phenomenon at a psychiatric hospital. He noticed that at frequent but irregular intervals a large group of patients who had assembled in the dayroom raised their hands high in the air. When he investigated, he found that a psychologist had given an MMPI to one of the patients, telling him to complete the form and then return it to the psychologist's office. The patient had asked for help from his fellow residents. As the patient read aloud each of the items, the assembled group would raise their hands to vote on whether the item should be answered true or false.

Pope, Tabachnick, and Keith-Spiegel (1987, 1988) have published national survey data concerning both the practices and beliefs of psychologists in regard to "having clients take tests (e.g., MMPI) at home" as well as to related assessment issues such as "not allowing a client access to a testing report."

**MMPI-2 Test Results as Hypotheses**

Clinicians must constantly keep in mind that the results of an actuarially-based instrument such as the MMPI do not
constitute conclusively authoritative pronouncements but rather generate hypotheses that the clinician must then evaluate in light of other sources of data.

**CLINICAL USES AND REFERENCE GROUPS**

Clinicians must be acutely aware of the degree to which certain clients to whom they administer the inventory may differ (e.g., in terms of background experiences) from those who have participated in either the original normative sampling or the subsequent research. Similarly, they must be alert to the ways in which the uses to which they put the MMPI-2 may differ from the uses validated or explored in the accumulated research. For example, Pope and Bouhoutsos (1986) described the utilization of the MMPI in screening patients (who had been sexually abused by a prior therapist) for admission to group therapy in UCLA treatment programs. The traditional MMPI indicators for suitability for group therapy appeared to lack predictive validity for this particular subset of patients. In some of the most extreme instances, sexually abused patients whom the MMPI results seemed to indicate were unable to participate meaningfully in — let alone benefit from — outpatient group therapy were nevertheless able to respond positively in this treatment modality.

Another example of a subset who would warrant special consideration are psychologists who have engaged in sexual intimacies with a patient (see Pope, 1989b, 1990a, 1990b; Pope & Vetter, 1991); individuals who are familiar with the test and may have administered and scored it literally hundreds or thousands of times may differ in crucially significant ways from the sample of individuals involved in the normative and subsequent research. When used with appropriate reference groups, the MMPI-2 has shown exceptional usefulness in assessment and treatment planning (Butcher, 1990b).

**USE OF AUTOMATED SCORING AND INTERPRETATION SERVICES**

Research indicates that computer-based test scoring and interpretation services have been widely accepted by psychologists; less than 40% report never "using a computerized test interpretation service" (Pope, Tabachnick, and Keith-Spiegel, 1987, 1988). Those who rely upon such services should clearly understand the decision rules by which raw data are transformed into interpretive statements and the evidence regarding the validity, reliability, and related psychometric properties of the instrument and program.

Those providing the computerized services have a clear ethical, scientific, and professional responsibility to provide a written presentation of this information. Psychologists who provide services include scoring and interpretation of tests must be able to demonstrate (i.e., through producing appropriate evidence) the validity of the programs and procedures upon which the service is based. Test users must be given the original scores from which interpretive statements are derived, must also be given access to the matrix of original scores, and the manual or interpretive report must set forth how the interpretive statements are developed from the original scores. Clinicians must adequately understand the tests they are using in their practice and the ways in which inferences are drawn from those tests. If they lack this understanding, they are functioning outside their area of competence (Pope & Vasquez, 1991).

**Conclusion**

Change is inherent in an empirical discipline, a fact recognized by those who initiated and supported the creation of the inventory. Both the original developers of the MMPI and the holder of the copyright recognized the need to update the instrument and to address certain serious problems with the original version. MMPI-2 is
the long-overdue revision of what has been perhaps the most widely used and most widely researched standardized psychological assessment instrument.

Sections of this article were previously published as "MMPI-2: A Practical Guide to Clinical, Psychometric, and Ethical Issues" in the Independent Practitioner (1990, vol. 9, #1, pp. 33-40).

Correspondence concerning this article may be addressed to James N. Butcher, Ph.D. Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455.

References


Butcher and Pope's article offers a concise and valuable description of the revision of one of the most important psychometric instruments available for psychological practice and research. The MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) promises to be a substantial improvement over its predecessor and is a major contribution to the field of personality assessment. Much of the former research and the clinical experience that one has accumulated over the years will continue to be applicable to the MMPI-2. I am somewhat surprised that the revision of the basic scales was not a more substantive one. Based on the research literature, psychometric considerations, and clinical needs, a more fundamental revision may have been appropriate.

The Improvements in the MMPI-2

Many weaknesses in the MMPI have been eliminated in the MMPI-2, including revision of some items that may have introduced bias, the extension of the representativeness of the normative sample, and the development of current norms to transform raw scores into T scores. Researchers have kept intact the original clinical and validity scales but objectionable and obsolete items have been removed. A new and more broadly representative normative sample was balanced on a number of key variables, including ethnic membership. Education level does not appear to influence the Mf and K scores and the interpretation of these scales, therefore, will be easier. Two new validity scales are available to assess consistency of responding that, in conjunction with the other validity scales like F, may make it easier to detect faking bad and other types of invalidity that have plagued clinical interpretation. The scale T scores now have uniformity in percentile distribution which also promises to make interpretation among scales more comparative.

The MMPI-2 scales and profile codes appear to be highly congruent with the original MMPI scales and profile codes, permitting application of the wealth of research and information on the MMPI to the new instrument. A new set of 15 content scales have been developed that address clinical problems not previously covered. A new adolescent form, the MMPI-A was developed with adolescent specific items and new norms for youth aged 14 through 17.

Possible Limitations in the MMPI-2

The renormalization of the MMPI-2 has resulted in narrower ranging distributions for all scales that may initially present some difficulties for clinical interpretation. A T score of 65 or greater is now interpreted to be clinically significant. Making the distribution of T scores uniform, however, assumes that the different scales measure similar types of psychological phenomena with similar distributions in the population. The clinical scales are mixtures of traits, states, and clinical
symptoms. These are very different types of psychological phenomena and ordinarily can be expected to show different types of distributions. It is only because the scales of the MMPI are a mix of such variables that the uniformity of T-score distributions may be viable methodologically. The underlying theoretical assumptions, however, are questionable.

The limitations of the revised inventory are not much dealt with in the MMPI-2 Manual or in the Butcher and Pope article. Issues of validity, such as empirical correlates and clinical applications of the basic scales, need detailed examination. Some discussion of such issues is appropriate because they put the revisions into the larger context. It can also lead to a useful re-evaluation of the role of psychological inventories in clinical assessment and diagnosis.

The authors of the MMPI-2 state that most of the items of the validity and clinical scales were retained in order to preserve the vast amount of research that has accumulated on the MMPI. Was this decision, however, justified on an empirical basis? What do the findings of past or present research point to, a revised instrument based on much the same items, or a more updated and empirically based selection of new items? The authors of the MMPI-2 obviously chose the former route, as others have recommended, in order to preserve the wealth of information accumulated on the MMPI. The authors fail to discuss the issue of revising the items based on empirical keying. This is surprising in view of the fact that the original MMPI broke new ground because of its empirical keying approach to item selection. Are there substantial reasons why a more fundamental revision of the inventory in some way, or an item selection based on an empirically derived set of new items, should have been considered or at least discussed? I will try to outline some of the issues that I feel could have been considered in this regard.

Changes in Diagnoses and Clinical Utility of the MMPI
In some areas there have been notable changes in symptom frequencies and diagnoses in the population in the past sixty years (Blum, 1978; Morrison, 1974; Silverstein, Warren, Harrow, Grinker, & Pawelski, 1982). Is it possible that a new set of empirically based items might have been more successful in achieving the original purpose of the MMPI, which was to diagnose psychiatric symptoms and syndromes? Problems with the MMPI as an instrument for diagnosis have led to substantial changes in its use. Differential diagnosis proved to be difficult based on single high point scales. New codes have been developed but these codes are still not exclusionary, that is, they are not specific to single diagnostic categories.

The use of the MMPI for clinical diagnosis has presented problems. In some studies the correspondence of MMPI basic scales and code types (Holland, Levi, & Watson, 1981; Winters, Weintraub, & Neale, 1981); and of MMPI decision rules (Klinger, Johnson, Giannetti, & Williams, 1977; Rogers, Wasyliew, & Dolmetsch, 1982) to clinical diagnosis has been less than satisfactory. Other studies have found the MMPI to reach clinical utility for the diagnosis of some disorders, such as schizophrenia (Patrick, 1988). The usefulness of the MMPI as a tool for clinical diagnosis has been somewhat disappointing because in practice the inventory is often used to address differential diagnosis among various possible syndromes, which was not the basis of its construction. In the literature, particular MMPI high point and code types are often common to a number of different diagnoses (see Vincent et al., 1983). Explanations for some relationships are not always obvious and can become somewhat of a stretch in rationalization. It must be pointed out, however, that the MMPI is probably still better than most of
Clinical Diagnosis and Personality Theory

In their current assigned clinical tasks, the MMPI basic scales and codes have been relatively more successful in the differential diagnosis of the broader categories of DSM-III personality disorders (e.g., anxious and fearful vs. dramatic, emotional, or erratic), of affective disorders with codes in which scale 2 appears, of schizophrenia, and of global psychopathology (Patrick, 1988; Vincent et al., 1983). The success of the MMPI in the differential diagnosis of personality disorders (see Morey, Blashfield, Webb, & Jewell, 1988; Morey, Waugh, & Blashfield, 1985; Vincent et al., 1983) may be related to the fact that responses to scales that are mixtures of symptoms, traits, and states, are likely to be distributed along continuous dimensions. Consequently, the MMPI may be more compatible with conceptualizations of psychopathology as extremes along continuous or normal dimensions (Eysenck & Eysenck, 1975) than with conceptualizations of psychopathology as discrete and qualitatively different aberrations from normal as, found in psychiatric diagnoses (Foulds, 1965). The factor structure of the MMPI-2 also suggests that several scales load on both personality and symptom factors (Butcher et al., 1989). Thus, there may be an inherent clash between the basic scales of the MMPI and differential diagnosis of clinical syndromes because they are conceived within different universes of discourse. This may not be the case for the measurement and diagnosis of personality disorders.

The MMPI also has a prominent role as a measure of personality based on its behavioural correlates. As a personality measure, however, it has substantial theoretical and methodological weaknesses. As already mentioned, many of the basic scales are mixtures of symptoms, traits, and states. The intercorrelations among scales are substantial. The meaning of low scores is unclear.

The MMPI-2 also has some shortcomings because of its close similarity to the MMPI. For example, it retains some of the unresolved ambiguities of the MMPI in its development, scoring, and interpretation (Horvath & Jonsdottir-Baldursson, 1990). As instruments of personality and behaviour the scales of the MMPI, and consequently the MMPI-2, are not based on a consistent rationale for scale construction. The items of the basic scales of the MMPI were not selected to measure specific personality and behavioural criteria, but to differentiate among rather broad categories of abnormality or clinical syndromes. A theoretical or rational strategy was not used in the construction of the basic scales of either the MMPI or MMPI-2, as seen in the construction of some other measures of psychopathology (e.g., Jackson, 1989).

External and Construct Validity

Overall, the MMPI accounts for a moderate and adequate amount of variance in external criteria (Atkinson, 1986; Parker, Hanson, & Hunsley, 1988). Some studies, however, have found that the size and specificity of the behavioural correlates of MMPI-based categories are rather low (Hedlund, 1977; Lane & Lachar, 1979). Other studies have been more successful in demonstrating expected behavioural or clinical correlates (Keane & Gibbs, 1980; Williams & Butcher, 1989). The MMPI scales can have strong specific behavioural correlates such as the length of hospital stay (Glosz & Grant, 1981). Many of the scales, therefore, do seem to show adequate criterion and construct validities. Inconsistencies in the correlation of the scales with gender appear to be frequent enough for concern. Perhaps the major problem, expressed by Hedlund
(1977), is that symptoms and behaviours tend to correlate with several basic scales rather than show specificity to a single or few scales. The application of clinical descriptors to older populations has also been questioned (King, 1978).

The MMPI-2 scales present moderate criterion validity coefficients in expected directions (Butcher et al., 1989). Data presented by Graham (1990), however, suggest that among the restandardization samples, the size of the external correlates of the MMPI-2 were modest to low and not always as expected. The same pattern is observed among psychiatric patients and their symptom descriptors. There were also many and important differences between the correlates of males and females in both samples. These findings suggest that the MMPI-2 may present some similar concerns about validity as did the MMPI.

Conclusions
All of the above issues suggest that changes in the traditional format of the MMPI should be considered in the future and a more fundamental revision may have been appropriate. The question is how is this to be done? The authors of the MMPI-2 faced a number of difficult choices, one of which was whether to alter substantially an instrument with so much research, use, and an adequate amount of reliability and validity overall. They seem to have chosen two difficult and perhaps incompatible tasks, namely, to retain and improve a popular instrument and also to create a new one. To a large extent they appear to have succeeded. They have improved the inventory and the new content scales they developed show much promise and may point the way toward the future assessment of psychological disorders. The new MMPI-2 has so many advantages that, no doubt, it will have a great impact on the field of applied and clinical psychology.

References


The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) has long been a dominant instrument in the field of personality assessment. So it was with great interest that practitioners and psychometricians alike watched the unveiling of its first major revision in 1989. In a timely paper entitled "The Research Base, Psychometric Properties and Clinical Uses of the MMPI-2", Butcher and Pope guide test users and researchers toward an understanding of the updated MMPI by summarizing its key features.

A central reason for revising the MMPI, according to Butcher and Pope, was the need to edit specific items for ambiguity, complexity, and outmoded or sexist wording. At the same time, the MMPI revision committee committed itself to a "conservative" revision with the goal of keeping the basic scales relatively intact. The authors attribute the adoption of a conservative strategy to a desire to ensure the continued relevance of the vast MMPI research base. They could have argued further that the success of the original MMPI attests to the clinical usefulness and heuristic value of the basic scales. Despite its merits, however, the conservative approach would also appear to have certain drawbacks to which MMPI users should be alerted.

The conservative approach prevented the MMPI revision committee from correcting some well-known psychometric shortcomings associated with the basic scales. First, some items are still keyed in opposite directions on different scales, such as item 32 which is keyed both true on the Schizophrenia scale and false on the Social Introversion scale. Hence, any valid response a test-taker might give to such an item will be interpreted as indicative of psychopathology. Second, the degree of item overlap among the basic scales is essentially unchanged. Thus, scale intercorrelations on the revised MMPI remain difficult to interpret and not entirely appropriate for multivariate analytical techniques. Third, apparently no formal attempts were made to deal with response styles on the MMPI-2. Scales are still not balanced with respect to keying; many items still tend to elicit responding in terms of social desirability. The cost of correcting these sorts of problems would be to reduce the one-to-one comparability of the MMPI and MMPI-2; the potential benefit would be further increases in the relative discriminant validity of the MMPI-2 scales.

Butcher and Pope might also have commented directly on the implications of revising the MMPI items from the perspective of test construction theory. The MMPI is probably the preeminent example of an instrument developed according to an empirical approach to test construction (Burisch, 1984; Wiggins, 1973). That is, items were originally selected for an MMPI scale (e.g., the Psychopathic Deviate or Pd scale) because the items discriminated between a "normal" group and a criterion group, such as diagnosed psychopaths. Given that the overall validity of an empirically derived scale is determined by the relationship of its items with an external criterion, chang-
ing the items on a scale should necessarily affect the scale's validity.

Butcher and Pope address the validity issue head on by offering this illustration: "The $Pd$ scale measures "$Pd$ behavior" in exactly the same way as the original scale did since it is exactly the same scale." But the $Pd$ scale is not exactly the same scale. Adding or deleting items from a scale is not the only issue, as the authors imply. Four of the items on the $Pd$ scale were modified (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). From the empirical viewpoint, an item that discriminates between psychopaths and "normals" due to its outdated wording or complex grammar is as appropriate as an item that discriminates on the basis of content. Any item modification has the potential to affect the correlates of the item, and hence, the validity of the scale.

Whether modifications such as item changes do affect the validity of the MMPI-2 scales is in fact an empirical question. For example, using a group of university students who had responded to both the MMPI and the MMPI-2, Duckworth (1990) found some of the largest mean scale score differences occurred on the $Pd$ scale. In fairness, the use of different normative groups for scoring the MMPI and MMPI-2 may have contributed to this finding, and a single study cannot confirm or disconfirm the validity of a scale. Nonetheless, such data do not support the assertion that $Pd$ scales on the MMPI and MMPI-2 scales measure the same thing, and changes to the items may be part of the explanation.

Butcher and Pope breathe a "sigh of relief" because research has generally shown a high congruence between scale scores and code types on the original MMPI and those on the MMPI-2. They report data to support their point based on well-defined profile codes. Unfortunately, high congruence is not always reported in the literature (e.g., Duckworth, 1990) for MMPI - MMPI-2 profile comparisons based on the unselected profiles test users and researchers are most likely to encounter. Nevertheless the authors need not despair; they have set themselves a very difficult task in looking for congruence between MMPI and MMPI-2 profiles. The data show that the test-retest stability of code types based on two administrations of the original MMPI itself is at best modest, especially for higher order code types (Graham, 1987). This comes as no surprise because code types are based on extreme scale scores which will be affected by regression towards the mean as well as the well-known tendency of retest profiles to show less psychopathology than the original profile (Windle, 1954).

The very search for high congruence between MMPI and MMPI-2 profiles is to justify generalization of the validity data on the original MMPI to the MMPI-2. Then again, the revision of the MMPI was presumably undertaken to produce an MMPI-2 that would be more valid than the original. Validity data on the MMPI-2 will take years to accumulate but an overview of the initial validity studies (e.g., those reported in the MMPI-2 Manual) would have helped to convince readers directly that MMPI-2 scales measure what they are supposed to measure. Moreover, some direction on how to compare the anticipated body of research on the validity of the MMPI-2 with the existing body of research on the validity of the MMPI would be a welcome addition to the literature.

The internal consistency of an empirically derived scale is generally considered to be secondary to its test-retest reliability because such a scale is constructed to predict a criterion, not to measure a homogeneous construct. Butcher and Pope report retest coefficients that clearly show the MMPI-2 scales have considerable stability. In practice, however, individual differences in MMPI scale scores have always been interpreted to differentiate
among individuals, and one may expect that MMPI-2 scale scores will be used in a similar fashion. Thus, the internal consistency of the scales provides relevant information to test users as well and bears mention.

The MMPI-2 Manual (Butcher et al., 1989) shows coefficients alpha for the basic scales that range from .34 to .85 for males and from .37 to .87 for females. At least three basic scales have coefficients below .60 for both males and females, attesting to the heterogeneity of the items making up the scale. These scales may be useful for diagnosis but are unlikely to yield unambiguously interpretable individual differences. Again, a conservative approach to test revision would have precluded any attempts to enhance the content of these scales. Nonetheless, in view of the practice of interpreting scale scores, some cautions by the authors in their section on the reliability of the MMPI-2 would seem to have been warranted.

In addition to discussing the basic scales, Butcher and Pope state that the MMPI-2 includes 15 scales constructed to measure the main content dimensions of the MMPI. More published information on their derivation would be desirable because the presence of these scales on the MMPI-2 is likely to determine in part what dimension practitioners assess and what dimensions researchers study. An explanation of how the content dimensions were conceptualized would help users compare the construct measured by the scales to other indices of the construct. For example, the items on the MMPI-2 Type A scale appear not to tap the three components traditionally associated with Type A behaviour, but rather appear to focus on only those aspects of Type A behaviour identified by meta-analyses as toxic for cardiac problems.

By merely listing the names of the content scales, Butcher and Pope miss an opportunity to add to the overview of the content scales presented in the manual. How were the 15 content dimensions on the MMPI-2 determined? Were they empirically derived, for example, from the factor structure of the original MMPI? Were they rationally selected to reflect dominant themes in personality research? Given the popularity of the MMPI and given the interactive relationship between assessment and theory, the MMPI-2 content scales may well have a gradual influence on what dimensions are seen as comprising psychopathology or even personality. A description of the evolution of these dimensions would help keep things in perspective.

Personality assessors have generally come to depend on validity and consistency indices to help them establish the interpretability of an MMPI profile for an individual. To that end, the MMPI-2 includes an alternate form of the F scale and two new consistency measures in addition to scoring keys for the Weiner-Harmon subtle scales. The authors describe the limitations of the Weiner-Harmon scales, expressing reservations about their use; but by their lack of comment they imply that the other three measures will work. Establishing useful measures of test protocol adequacy is more difficult than would appear. At a conceptual level, elevated validity indices are not readily interpretable. For example, high scores on the F scale may reflect exaggerated responding or high levels of psychopathology; low levels of consistent responding to pairs of semantically related items may reflect confusion or perhaps weak verbal skills, indecision or random responding.

At a practical level, test users might be relatively unconcerned about the meaning of validity and consistency indices if such indices accurately identified uninterpretable protocols. The success of consistency indices is mixed, however. To illustrate: Neither of the two indices on the original MMPI, namely, the TR index and
the Carelessness scale, could predict the retest stability of profiles based on the MMPI basic scales, although both were marginally related to the number of MMPI items changed from test to retest (Fekken & Holden, 1987). Restriction of range on these original MMPI indices surely contributed to their low predictive utility. It is encouraging that the consistency indices on the MMPI-2 are considerably longer than the original indices. Moreover, the MMPI-2 indices were constructed to reflect recent theorizing about the nature of consistent responding (see Tellegen, 1988). Still, their usefulness needs to be demonstrated empirically and until then, some cautions are in order.

Change within a profession, including psychology, is always slow. Each professional will have evolved a set of procedures that match his or her theoretical persuasions and practical needs. To replace the popular MMPI even with a conservative revision will require psychologists to make various procedural and conceptual adjustments. Papers like the one written by Butcher and Pope remind us of our obligation to remain current in our practice and research. We may wish that they had given us even more information in their overview of the MMPI-2, but this should not detract from the obvious: It is time to usher out the MMPI and to ring in the MMPI-2.

Correspondence concerning this article should be addressed to Dr. G.C. Fekken, Department of Psychology, Queen's University, Kingston, Ontario, CANADA K7L 3N6

References
Horvath and Fekken have made a substantial contribution to the evolution of MMPI-based assessment by providing thoughtful evaluations of key issues in the revision. We appreciate both their endorsement of the improved aspects as well as the questions and critiques they skillfully raise. In the limited space allotted to us, we will try to respond to the major issues.

First, both Horvath and Fekken, while noting the ways in which MMPI-2 improves the original instrument, point out apparent deficiencies in the original instrument that were not eliminated. For example, Horvath observes that "most of the items of the validity and clinical scales were maintained" and Fekken observes that "the degree of item overlap among the basic scales is essentially unchanged." Both authors place these observations in proper context by accurately stating that the choices (to retain many of the aspects of the original instrument) were made, in Horvath's words, "as others have recommended, in order to preserve the vast amount of information accumulated on the MMPI." Regardless of its psychometric ugliness and somewhat arcaic aspects, the MMPI worked extremely well as a screening instrument for psychopathology. The Restandardization Committee strongly believed that the test earned its place as the most frequently used measure (Lubin, Larsen & Matarazzo, 1984) through its demonstrated validity, and that this extensive validational base was well worth preserving. There needed to be a clear tie to the past but there was also ample room, with the addition of new content, to use advances in the impressive psychometric technology that has evolved over the past 50 years. Unfortunately, some information about these developments, much of which has been recently published and some of which is still in press, was not available to Horvath and Fekken when they wrote their commentary. For example, many of Fekken's suggestions regarding new content scales were actually incorporated in the development of the MMPI-2's content scales (e.g., Butcher, Graham, Williams & Ben-Porath, 1990). Horvath accurately predicted that the MMPI-2 Content Scales would add significantly to the Clinical Scales' ability to identify and distinguish various patient populations. Ben-Porath, Butcher and Graham (in press), for example, found that the MMPI-2 Content Scales enable more accurate discrimination between depressed and schizophrenic individuals in inpatient psychiatric settings and actually outperform the clinical scales in differential diagnosis.

Second, Fekken raises an important concern about the validity:

Drs. Butcher and Pope address the validity issue head on by offering this illustration: The Pd scale measures "Pd behavior" in exactly the same way as the original scale did since it is exactly the same scale. But the Pd scale is not exactly the same scale. Adding or deleting items from a scale is not the only issue... Four of the items on the Pd scale were modified... From the empirical viewpoint, an item that
discriminates between psychopaths and "normals" due to its outdated wording or complex grammar is as appropriate as an item that discriminates on the basis of content. Any item modification has the potential to affect the correlates of the item, and hence, the validity of the scale.

Two published studies address the issue of modified items (i.e., those in which the content is retained while outdated wording or complex grammar is eliminated). Ben-Porath and Butcher (1989b) examined the psychometric stability of the 68 rewritten MMPI-2 items by comparing their contribution to the relevant scales with the contribution of the originally worded items. A subsequent study evaluated the effects of changes introduced in the MMPI-2 on the stability of the 13 basic scales and 8 supplemental scales that were retained in the MMPI-2, and found the clinical scales of the two versions of the test to be basically equivalent (Ben-Porath and Butcher, 1989a).

Third, Fekken cites the work of Duckworth which reports apparent "differences" between the profiles of some participants who took both the original and revised MMPI. According to the written report of this research (Duckworth, in press), there were 85 participants. This diverse group included 13 clients, 31 undergraduates, 23 counseling psychology graduate students, and 18 friends. Some but not all of the testing was conducted by students in an advanced level assessment course. Unfortunately, Duckworth's report fails to address a fundamental prerequisite of experimental design: the test-retest variable. That is to say, in order to discover the amount of variance due to the two forms of the test, one must first discover and control for (e.g., use as a baseline) the (test-retest) variance due to two administrations of the same form. This potential confounding variable was taken into account in the previously cited research conducted by Ben-Porath and Butcher (1989a), in which 189 participants completed both the original and revised forms of the MMPI while 188 other participants completed the original form twice. (Note: research data concerning expected differences in administering the MMPI-2 on subsequent occasions was published in the MMPI-2 Manual.) Conditions of administration were held constant for all participants. Had Duckworth used a more complete research design, it is likely that she would have seen that responses to the same form varied in the way that she found MMPI and MMPI-2 responses varied.

Fourth, Horvath, in his critique of the original MMPI, questions validity generalization. For example, he cites a study by Winters, Weintraub, and Neale (1981) which concluded that the MMPI was ineffective in diagnosing schizophrenia. However, this research did not study the original MMPI instrument but rather an abbreviated derivation or approximation of the test, the Minimult, which utilizes only 81 of the original MMPI items. The Minimult is indeed notoriously inaccurate in predicting MMPI clinical scores (Hoffman & Butcher, 1975). Research utilizing the MMPI itself has been quite heartening in this regard. Cross-cultural research — in which an instrument is translated into another language and culture — poses one of the most gruelling and telling tests of an instrument's generalization validity. How has the MMPI fared? Research in Greece by Manos (1985) and in Turkey by Savacir and Erol (1990) has shown substantial cross-cultural usefulness of MMPI scales in clinical diagnosis. A recent review of the use of the MMPI in China (Cheung & Song, 1989), for example, reported 20 recent studies on the MMPI for schizophrenia in the Peoples Republic of China; most showed strong cross-cultural generalizability of Scale 8 among Chinese individuals suffering from schizophrenia.
Fifth, both Horvath and Fekken note accurately that there are additional aspects about the revision process, the psychometric properties, and the applicability of the MMPI-2 that were not addressed in our original article. These are significant issues that deserve a much more detailed response than we can render in this brief response. However, these issues have been and continue to be addressed in the research literature. In addition to those already cited, the following works, representing only a small sample, may be of interest to readers seeking information:

References
Butcher, J. N. (1990d). What interpretive changes do the educational differences between the original MMPI and MMPI-2 normative samples require? MMPI-2 News and Profiles, 1, (1) 2.


Roper, B., Ben-Porath, Y. S., & Butcher, J. N. (1991) Comparability of computerized adap-
Le "Minnesota Multiphasic Personality Inventory" (le MMPI) est devenu un des instruments d'évaluation de la personnalité auquel on a le plus souvent recours et dont on se sert dans une grande variété de contextes cliniques d'évaluation. Depuis son introduction, il y a quelques décennies, plusieurs problèmes, tels que la nature démodée de certains termes employés, sont devenus apparents. Après évaluation et discussion, ces problèmes ont résulté en la révision et le développement récents de l'instrument afin d'éliminer certains problèmes et de créer un instrument psychométrique plus adéquat pour l'évaluation clinique. L'article de Butcher et Pope décrit la révision de l'instrument et offre un résumé de quelques-uns des principaux éléments du MMPI-2.

Butcher, Pope, Graham, Williams, et Ben-Porath (1989) ont développé une nouvelle série d'échelles de contenu afin d'évaluer les principales dimensions de contenu du MMPI-2. Ils ont développé ces échelles à partir de procédures à la fois objectives et statistiques. Les 15 échelles de contenu évaluent les facteurs importants de la personnalité, le comportement symptomatique, ou s'appliquent à des domaines cliniques problématiques importants. Les 15 échelles sont:

1. l'anxiété
2. les craintes
3. l'obsession
4. la dépression
5. les inquiétudes sur la santé
6. les pensées bizarres
7. la colère
8. le cynisme
9. les activités contre la société
10. la personnalité du type A
11. une mauvaise opinion de soi
12. la gêne en société
13. les problèmes familiaux
14. les problèmes au travail
15. les mauvaises réponses aux traitements

Les échelles de contenu du MMPI-2 se sont montrées très utiles dans le domaine de la pratique clinique car elles couvrent une vaste gamme de problèmes et peuvent se comprendre de façon intuitive.

Les chercheurs qui ont développé le MMPI-2 avaient remarqué que l'instrument original était un des instruments les plus populaires pour l'évaluation des adolescents malgré le fait que la plupart des éléments de cet instrument s'adressaient aux adultes et qu'il n'y avait que très peu d'échelles s'adressant...
aux problèmes des adolescents. La nouvelle version adolescente de l'échelle, le MMPI-A, comprend des éléments ayant rapport à l'adolescence. En plus des échelles cliniques du MMPI-2, la version adolescente contient quatre nouvelles échelles de contexte qui mettent l'accent sur les problèmes des adolescents. Ce sont: les problèmes de comportement, l'aliénation, le manque d'ambition et les problèmes associés à l'école.

Butcher et Pope sont d'avis que les psychologues cliniques devraient se servir du nouveau MMPI-2 plutôt que de l'instrument original. Ils maintiennent qu'il est contraire à la déontologie de se servir d'une version démodée d'un test, surtout un test tel que le MMPI qui présentait des problèmes de langage sexiste et d'exclusion de minorités raciales de l'échantillon original.

Dans leurs critiques du MMPI-2, Horvath et Fekken se demandent tous deux pour quelles raisons les révisions n'étaient pas plus importantes, étant donné les multiples faiblesses qu'on avait identifiées dans le MMPI original. Butcher et Pope répondent que l'instrument original était extrêmement efficace en tant qu'instrument de dépistage de la psychopathologie, et qu'il était important de garder les aspects du MMPI qui s'étaient montrés utiles.

Horvath fait remarquer que les échelles cliniques mesurent en fait des mélanges de traits, d'états et de symptômes cliniques. Ce sont des types de phénomènes psychologiques très différents et, d'après lui, les hypothèses sous-jacentes qui permettent ces mélanges sont contestables. Il croit que l'emploi du MMPI comme instrument de mesure de la personnalité n'est pas sans faiblesses: les corrélations entre les échelles sont importantes et la signification de résultats peu élevés n'est pas claire. Il fait remarquer que le MMPI est un instrument décevant en ce qui concerne les diagnostics cliniques. Il est peu probable que ces problèmes puissent s'éliminer par une stratégie conservatrice de révision. Horvath soutient que ceux qui ont révisé le MMPI semblent avoir entrepris deux tâches difficiles et peut-être incompatibles: retenir et améliorer un instrument populaire et, en même temps, créer un nouvel instrument. Néanmoins, il tire la conclusion que le MMPI-2 est destiné à avoir un impact important dans le domaine de la psychologie appliquée.

Fekken maintient que l'approche conservatrice à la révision a laissé subsister dans le MMPI-2 plusieurs des défaillances de l'instrument original. Pour certaines questions, des réponses "oui" et "non" comptent sur des échelles différentes, de sorte que n'importe quelle réponse est interprétée comme un indice de la psychopathologie. Elle conteste également l'affirmation de Butcher et Pope selon laquelle l'instrument révisé retiendrait les avantages du MMPI original. Fekken fait remarquer que toute modification, même à la terminologie démodée et au langage sexiste, pourrait changer la validité de l'échelle. Elle soutient que la validité du MMPI-2 dépendra des recherches entreprises à l'avenir et qu'on ne peut pas prendre cette validité pour acquise, seulement parce que les révisions faites à l'instrument original ont été de nature conservatrice.

Dans leur réponse aux critiques de Horvath et de Fekken, Butcher et Pope terminent en disant que le MMPI s'est montré très fiable dans le domaine des recherches interculturelles, en Grèce, en Turquie et en Chine.