Researchers Grapple with Problems of Updating Classic Psychological Test

The widely used Minnesota Multiphasic Personality Inventory, normed on depression-era midwesterners, is up for a retuning

In recent years, there has been enormous proliferation of tests for personality and psychological disorders. But the Minnesota Multiphasic Personality Inventory (MMPI), developed 45 years ago, remains the most heavily used test of its kind in the world. Now, researchers are engaged in a project to determine whether the test, whose definition of "normal" is based on a relatively small sample of white, rural, depression-era Minnesotans, has retained its validity.

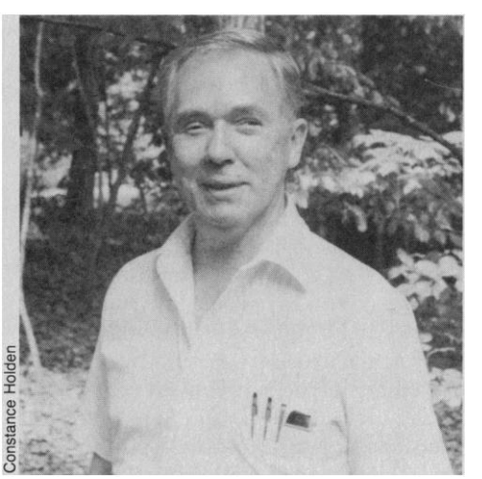
Tampering with such an instrument is a delicate project indeed, for any major change would jeopardize its comparability with the immense body of literature—over 8000 books and articles—that has accrued to the test. The MMPI is used now more than ever for psychiatric assessments, as a screening tool, and for research. The test is used in 46 countries around the world and has been translated into 124 languages. MMPI scores have been gathered on almost every kind of medical and psychiatric population as well as on the young, the old, prisoners, religious groups, and ethnic subgroups.

Over the past decade, researchers have come to feel it is time to reexamine the test to see "to what extent it is still working as it ought," according to W. Grant Dahlstrom of the University of North Carolina. Dahlstrom, who has been intimately involved with the test for over 40 years, is one of a committee established by the MMPI's owner, the University of Minnesota Press. The "restandardization" of the MMPI is a long-running project that is expected to be completed some time next year.

The MMPI has been a colossally successful instrument, thanks to the astuteness of the two clinicians who developed it, Starke R. Hathaway, a physiological psychologist, and J. Charnley McKinley, chairman of the psychiatry department at the University of Minnesota in the 1930's. The test was originally developed as an instrument to identify psychological problems in medical patients. Hathaway and McKinley had come to the recognition, precocious for their day, that a very high proportion of patients seeking medical help were in fact primarily disabled by emotional problems.

The normal population on which the test is based was represented by a group of about 700 rural Minnesotans, most of them recruited while visiting friends and relatives at the university hospital, augmented by local college students and workers in the Works Projects Administration. The typical test subject was a 35-year-old with an 8th-grade education, either a housewife or a skilled or semiskilled worker.

Hathaway and McKinley were bucking the trend of their times when they decided to construct an objective test for psychopathology. Existing tests were in disrepute, having been formulated through an "armchair" approach based on articles and textbooks, with questions that had not been well validated on clinical populations. The two doctors were in the vanguard of what was referred to as "dustbowl empiricism," says Dahlstrom. They collected over 1000 possible test items, gleaned from their own practice as well as the existing literature, and then tested them on normal groups, medical patients, and psychiatric patients. Only those questions that clearly differentiated the normals from the "criterion" groups (patients already diagnosed with the disorder under study) were retained—regardless of whether they had "face validity"—that is, regardless of whether the item seemed obviously related to the disorder under investigation. For example, "Nothing in the newspaper interests me except the comics" is frequently endorsed by people suffering from depression.



W. Grant Dahlstrom: *Former student and colleague of Hathaway is in charge of the MMPI restandardization project.*

In this way, they developed the final 550 items, simple declarative statements which the respondent marks true, false, or "cannot say." Some are related to the clinical scales which measure deviance on ten separate indicators of psychological adjustment; others are "nonworking items" that furnish information for the dozens of specialized subscales that test users have developed. The content was pegged at 6th-grade reading level and stuck for the most part to the universals of everyday experience.

The authors originally devised eight clinical scales, starting with "hypochondriasis," the simplest to define and the most immediately relevant to their medical patients. The other scales measuring pathology are depression, hysteria (which has two elements—somatic symptoms and items indicating denial), psychopathic deviance (antisocial tendencies), paranoia, psychasthenia (measuring obsessive-compulsive tendencies and anxiety), schizophrenia (eccentric thinking), and hypomania (energy and activity level). Later, two more scales were added: one measuring masculinity-femininity, and one on social introversion. In what was then an innovation, the authors added several scales that serve as validity checks on the answers to the main items. These take into account incompetence or confabulation by the test-taker, and tendencies such as exaggeration and defensiveness that have an across-the-board effect on responses.

A vast literature, complete with "cookbooks," has sprung up to aid in interpretation of MMPI scores. Each scale is scored separately, but is meaningful only in the context of the whole profile. "Code types" refer to frequently observed combinations of

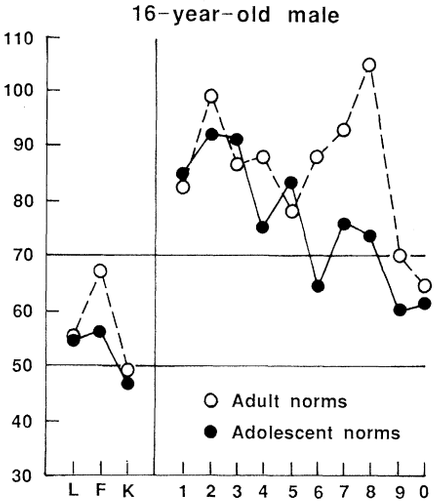
scale elevations. Perhaps the most common is 1-2-3 in any order, known as the "neurotic triad," a pattern combining hypochondriasis, hysteria, and depression. Another common code type is 4 (psychopathic deviance) and 9 (hypomania), often manifested by people with antisocial personalities.

The MMPI is widely used in employment screening as well as in clinical practice and research. Policemen, for example, are often given MMPI's as part of promotion decisions to screen out overly rigid and suspicious personalities, proneness to anxiety or indecisiveness, or a penchant for impulsive use of weapons. Nuclear power plant personnel are screened for factors such as hostility to authority, poor reactions to stress, and substance abuse (which has its own suggestive code types and subscales).

In view of the heavy reliance on the test, particularly in employment decisions, it has become important to learn whether the scores obtained from the original norming sample still supply a legitimate standard of comparison for the population at large. Much has happened since World War II—urbanization, higher levels of education, increased ethnic diversity, mobility and family breakdown, changes in the status of women, and a liberalizing of religious and moral views—all of which might be suspected to lead to significant changes in the way test-takers respond to many of the items.

According to research on the test within the past decade, Americans are now scoring somewhat higher on most of the clinical scales than they were in 1940, which means that most now look slightly deviant. The reasons for this are still not clear. James Butcher of the University of Minnesota says this is probably an artifact of the test. When Hathaway was doing his research, he tolerated an average of 30 omissions, but test-takers are now urged to respond to all items and the omission rate is only about 3.

Dahlstrom says there are other possible interpretations. The rate of distress in the general population could indeed be higher (a reading supported, at least, by statistics on postwar rates of depression), people could be construing certain items differently, or they could just have become more open about reporting their problems. Psychologist John R. Graham of Kent State University is inclined to the last theory. "In the 30's," he says, "people were so concerned about what other people thought that lots of things probably went on in their heads that nobody ever knew about." He also says that if increased social turmoil has led to more psychopathology, it may well be counterbalanced by the lessening of the stigma of mental illness and the increasing accessibility of treatment.



MMPI profile *showing differences between adolescent and adult norms on a clinically elevated profile. A score over 70 indicates deviance from about 95% of the normal population. If the adolescent norm is applied, code type changes from 8-2 to 2-3-1. L, F, and K scales serve as validity checks; 1, hypochondriasis; 2, depression; 3, hysteria; 4, psychopathic deviance; 5, masculinity-femininity; 6, paranoia; 7, psychasthenia; 8, schizophrenia; 9, hypomania; 10, social introversion. [From C. L. Williams, "MMPI profiles from adolescents: Interpretive strategies and treatment considerations," J. Child Adolesc. Psychotherap. 3 (no. 3), 179 (1986)]*

A particular impetus for reevaluating the test has come from the perceived need to include blacks and other minorities in the norm and evidence that certain groups of blacks deviate significantly from white norms on certain scales. The most extreme example of this comes from a study conducted by psychologist Malcolm Gynther of Auburn University in the small black rural community of Riverbend in Alabama. An oral version of the test was administered to sidestep the illiteracy problem. Nonetheless, the profiles of the sample of 88 young black adults made 80% of them look psychotic. One item, for example—"Evil spirits possess me at times"—was affirmed by 58%, compared with 3% of whites.

A 1976 survey of 882 middle-class blacks in Alabama, Michigan, and North Carolina (known as the tri-state sample), showed some consistent, if far more modest, deviations. The most common elevations were on scales 8 (schizophrenia) and 9 (mania), and on the F scale that contains items very infrequently endorsed by normal people. This includes statements like "My soul sometimes leaves my body." There were also certain items, particularly relating to what Gynther called "social cynicism," that distinguished the response of blacks of both sexes. In addition there was a higher black response on items indicating exaggerated self-importance, such as "If given a chance, I would do things of great benefit to the world," and statements indicating higher levels of alienation and unhappiness. Gynther is inclined to believe that these differences have more to do with differences in attitudes and values—as well as being reflections of expectable differences in social well-being—than in levels of psychopathology. For example, far fewer blacks than whites endorse the statement "the police are usually honest."

There have been efforts to develop revised black norms, based on the tri-state sample, to overcome whatever cultural bias is in the test. David Lachar, a psychologist at Good Samaritan Medical Center in Phoenix, Arizona, says "I thought new norms would be a quick fix and correct a problem that seems to be substantially demonstrated in the literature." But he found the process neither quick nor a fix. An experiment applying a black-based norm to a black psychiatric population at Lafayette Clinic in Detroit resulted in erasing most psychotic patterns from the test even though clinical evaluations clearly indicated otherwise. With the black norm, 30.7% of the patients appeared normal, compared with 7.6% on the white norm. What's more, says Lachar, the new norm not only raised the threshold for pathology, but altered the profiles, so that 80% of the code types were changed. Lachar said he became convinced that any change at all in the test norm would wreak havoc with interpretations.

Dahlstrom, who extensively reviews the literature in a new book, *MMPI Patterns of American Minorities*,* says he is satisfied that the test is not less valid for blacks than for whites. In other words, although there are different patterns in some groups of blacks, these reflect the same traits they do in whites and do not result from different interpretations of the items by blacks. He also says, the tri-state study (in which he participated) notwithstanding, that when corrections are made for education and socioeconomic status, black-white differences become insignificant.

One population for which separate norms do seem to be advisable is adolescents. Use of the MMPI has posed a long-standing problem here. The test was not used on adolescents until 1963 after it had been renormed on a population of 9th graders tested during the 1940's and 1950's. Now, according to psychologist David Pellegrini of Catholic University, clinicians are faced

*By W. G. Dahlstrom, D. Lachar, and L. E. Dahlstrom, to be published this fall by the University of Minnesota Press.

with a confusing situation. They can score according to adult norms, score according to adolescent norms and use interpretive literature based on adults, or they can totally base interpretations on the (more limited) adolescent literature. These choices make a big difference: many adolescents show pervasive and striking scale elevations compared with adults, particularly on scales 4 and 9, reflecting their turbulent natures. Scoring them like adults, therefore, could make many of them look as though they have severe adjustment problems when they are just acting like adolescents. Says Graham: "they are very resistant, passive-aggressive, very rebellious as a group."

The restandardization project is coming to grips with the problem by developing a new adolescent norm based on a sample of about 3000 contemporary adolescents, using an expanded test with 124 additional questions. The new items do not affect the clinical scales but supply more detailed information on family and school relationships, substance abuse, eating disorders, identity problems and personality disorders not deeply tapped in the original test.

The nature of changes to the adult norm is yet to be determined. An expanded adult questionnaire has been developed containing a total of 704 questions which, like the new adolescent test, has new items relating to family relationships, substance abuse, and personality disorders, as well as additional ones on "Type A" personality, and depression and suicidality. There are also new items predicting the subject's responsiveness to therapy. This form is being tested on a national sample of 2700 normal adults in six states whose composition reflects the 1980 census data, which includes 11.2% blacks. It is also being tested on several psychiatric samples and a group of alcoholic inpatients.

Data have not yet been analyzed. Item-by-item comparison with Hathaway's original test data will reveal whether there have been any significant shifts in response patterns over time. But so far, the largest change from the 1930's sample appears on the masculine-feminine scale, which researchers acknowledge is the most culturally determined of the scales.

The authors of the test originally conceived of deviant sexual orientation as part of character pathology and first attempted to base the scale on a sample of homosexual men. But they found the sample too diverse to serve that purpose and also discovered that the inverse of the scale did not work for "masculine" women. This is "the one scale where contemporary thinking was excessive," says Dahlstrom. The scale, because of its heterogeneity of items, can serve as an indication of sexual problems, gender identity problems, or as a more general measure of interests. As a consequence, educated men score higher on femininity because the items measuring femininity include aesthetic tastes and emotional sensitivity. Dahlstrom says that at the University of North Carolina researchers have found that the healthiest subjects show high ratings on both masculinity and femininity, indicating a combination of assertiveness, self-management, sensitivity, and nurturance. Ultimately, he says, the scale might be altered to become a positive measure of health, or "androgyny." But for now, the restandardizers do not intend to tamper with it.

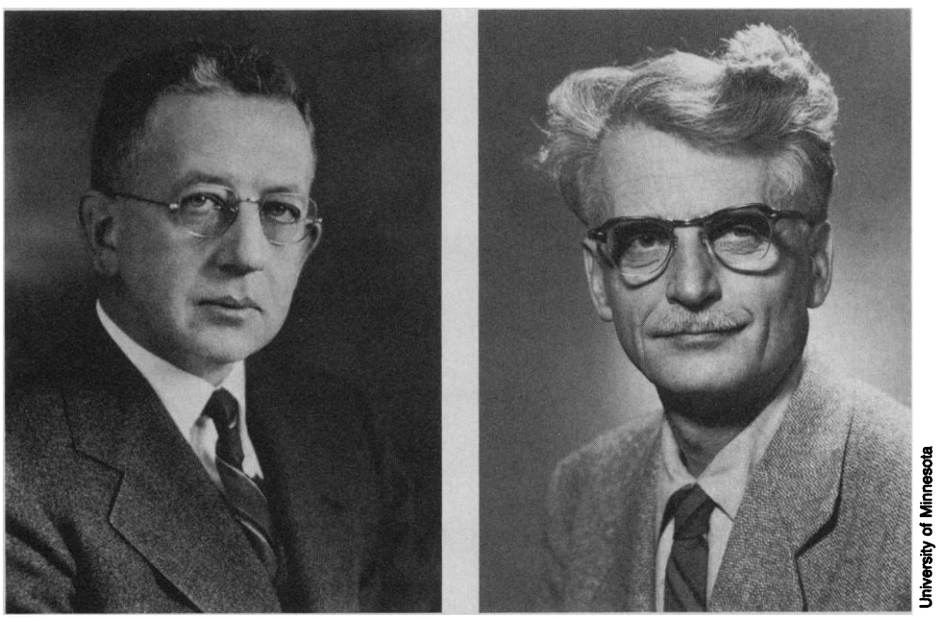
On the level of individual items, there is one for which response rates have undergone a striking change in the past 40 years: "I am an important person." Originally designed as a measure of self-importance and grandiosity, its connotations have changed with the political and human potential movements of the past two decades. Whereas 9% of the Minnesota sample endorsed this item, 49% of males and 69% of the females in a contemporary white sample endorsed it. The change was even more striking for black males, 80% of whom endorsed it in the 1976 black sample.

There are no plans at present to drop any of the original items. Of the original 550, 14% have been reworded to weed out anachronisms and sexist references. For example, a reference to playing "drop the handkerchief" may be changed to "tag games." "I would enjoy beating a criminal at his own game" is being reworded so it is not sex specific. A reference to the Bible is changed to "holy scriptures." There are very few items, though, whose meaning has been changed or lost with time. One exception is: "I think Lincoln was greater than Washington." Researchers say this does not apply to the clinical scales in any case, and may be relegated to the end where it can be administered at the discretion of the test-giver.

Notwithstanding the changing profile of American life, the MMPI has proved a remarkably durable instrument over the 45 years of its use, and it seems to be much less vulnerable to cultural variation than IQ tests, which require periodic revalidation. The Russians use the MMPI—with minor adjustments, such as the omission of some religious items—on their cosmonauts and in the preparation of their Olympic teams. An assessment of the test in Nigeria revealed only 33 items that were irrelevant or culturally inappropriate.

If anything, the restandardization project affirms the remarkable stability of personality traits and particularly of patterns of human psychopathology, which, says Graham, are "amazingly stable over time and across cultures." The test is extremely adaptable, as demonstrated by the proliferation of subscales on everything from lower back pain to post-traumatic stress disorder. According to Robert C. Colligan and his colleagues, authors of a 1980 normative study unrelated to the standardization project, the success of the test has outstripped its authors' expectations. Hathaway and McKinley "clearly viewed it as representing a first generation product, which would, and should, be superseded by improved assessment tools. This has not happened." ■

CONSTANCE HOLDEN



University of Minnesota

Fathers of the MMPI: *J. Charnley McKinley (left) and Starke R. Hathaway.*