JOURNAL OF PERSONALITY ASSESSMENT, *78*(1), 145–160

Comparability of MMPI–2 Scales and Profiles Over Time

Patrick H. Munley

Western Michigan University and
Veterans Affairs Medical Center
Battle Creek, Michigan

This study investigated Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) scale and profile comparablilty for MMPI–2 profiles completed on 2 separate occasions by mental health patients receiving treatment at a Veterans Affairs Medical Center ($n = 114$). Patients were predominantly men (96.5%), with an average age of 44.08 and an average of 12.39 years of education at the time of initial testing. MMPI–2 tests were completed on 2 separate occasions as a routine part of treatment with a mean interval between test administrations of 688 days. Findings were analyzed for the complete sample and for 3 subsamples with different test–retest intervals. MMPI–2 scale test–retest correlation coefficients for the entire sample ranged from .48 to .69 for the Basic scales, .49 to .80 for the Supplementary scales, and .56 to .78 for the Content scales with scale high-point agreement = 38.60%, high 2-point agreement = 16.67%, and high 3-point agreement = 19.30%. High-point agreement for subsets of participants with well-defined high points, 2-points and 3-points was 41.07%, 27.50%, and 25.93% respectively. Pearson *r* correlation coefficients for *T* scores across the Basic scales for pairs of profiles averaged .78, suggesting similarity of profile shape across testing occasions. MMPI–2 profiles were also examined in relation to Skinner and Jackson's 3 modal MMPI profile types.

Given the widespread use of the Minnesota Multiphasic Personality Inventory (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) as a measure of personality and psychopathology, information on the comparablilty of test scale scores and profile configurations over time is of interest and importance to clinicians. Research on the temporal stability of the MMPI and MMPI–2 has studied both the test–retest correlations of profile scales and the configural comparability

of scale high points and code types, often with normal populations to evaluate the test–retest reliability.

With the publication of the MMPI–2, research was conducted that was relevant to the issue of the comparablilty of profile characteristics across the MMPI and MMPI–2 in the context of profile stability with the MMPI. Ben-Porath and Butcher (1989) investigated the comparablilty of MMPI and MMPI–2 scales and profile characteristics in a sample of college students who either took the original MMPI twice or the original MMPI and then the MMPI–2 for comparison purposes. The average interval between test administrations was 9 days. They found that differences between profiles obtained by participants who completed the original and revised MMPI–2 did not exceed differences obtained by those completing the original MMPI twice. In this study the percentage of participants obtaining the same high point was approximately 54% in the original–original comparison and 59% in the original–revised comparison. In the comparison of high two-point scales, agreement was around 35% for men and 44% for women in the original–original comparison group, and 36% for men and 31% for women in the original–revised comparison.

Utilizing a design similar to Ben-Porath and Butcher's (1989), Chojnacki and Walsh (1992) investigated the scores and profile configural characteristics among college students who also took the original MMPI twice or completed the MMPI and the MMPI–2 at a 1-week test–retest interval. Test–retest correlations across Basic scales for participants who took the original MMPI twice ranged from .69 to .86 for men and .61 to .89 for women. Correlations across the Basic scales for participants completing the MMPI and MMPI–2 ranged from .55 to .88 for men and .59 to .88 for women. High-point, high two-point, and high three-point consistency for all 10 clinical scales for participants who took the MMPI twice was, respectively, 62.8%, 44.2%, and 37.2% for men and 55.3%, 31.9%, and 31.9% for women. Butcher, Graham, Dahlstrom, and Bowman (1990) studied the test–retest reliability of the MMPI–2 in a sample of male and female college students and found the 1-week test–retest correlation coefficients generally high for the basic validity and clinical scales and ranged from .62 to .87 for male and .73 to .91 for female college students. These findings were very similar to test–retest correlations obtained with the normative sample for the MMPI–2 (Butcher et al., 1989), which ranged from .67 to .92 for men and .58 to .91 for women across the basic validity and clinical scales.

Putnam, Kurtz, and Houts (1996) studied test–retest coefficients and absolute score changes for the Basic, Supplementary, and Content scales of the MMPI–2 over a 4-month test–retest interval in a sample of 111 male clergy who were not receiving mental health services at the time. Test–retest coefficients were similar to reliability coefficients reported in the normative group and ranged from .51 to .89 across the basic validity and clinical scales. Mean scores across the three sets of scales did not differ for the two testing occasions. Thirty one percent of partici-

pants in the study obtained the same single-scale high-point code on retesting, and 19% obtained the exact same two-point code type. In studying the absolute value of the differences in *T* scores between the two test administrations for individual participants, Putnam et al. found that substantial changes in retest scores may occur on any MMPI–2 scale. The majority of scales showed changes averaging 3 to 6 *T*-score points over the 4-month period. Nearly half of the participants showed changes of 10 or more *T*-score points on at least five different Basic, Supplementary, or Content scales. Putnam et al. (1996) observed that the reliability coefficients for the Supplementary and Content scales tended to be somewhat higher than the coefficients for the Basic clinical scales.

Graham, Timbrook, Ben-Porath, and Butcher (1991) studied the degree of code type congruence between the MMPI and MMPI–2 in the MMPI–2 normative sample and found that congruence was greater when code types had definition of at least 5 *T*-score points. They also found that congruence was greater between the two versions of the test on one administration than between two administrations of each version of the test. In a subsample of 82 male and 111 female normative participants who took the 704-item experimental version of the MMPI (Form AX) on two occasions at approximately a 1-week test–retest interval, MMPI one-point, two-point, and three-point code type congruence was 52%, 28%, and 19%, respectively, and MMPI–2 one-point, two-point, and three-point code type congruence was 49%, 26%, and 15%, respectively. For this comparison code types were included regardless of degree of profile definition.

In addition to studies focusing on MMPI and MMPI–2 scale reliability and profile stability with normal populations, there have been a few studies that have focused on the comparablilty of MMPI and MMPI–2 scales and profile configurations over time in clinical populations. Pauker (1966) studied the stability of MMPI profiles of female psychiatric patients by comparing test–retest changes in profiles of 107 women retested on average after 42 days, with a range 13 to 176 days. Generalized distance function comparisons of their test and retest profiles indicated more similarity between their initial test and retest profiles when compared to distance functions of their initial test profiles with retest profiles obtained from other women matched on several pertinent variables. In this study, high-point agreement on retesting was 44% and high two-point agreement was 25%. Von Cleve, Jemelka, and Trupin (1991) studied the test–retest reliability of psychological test scores that were obtained during the first few weeks of incarceration in a state prison from felony offenders over a 1- to 3-week test–retest interval. MMPI reliability coefficients for the Basic scales ranged from .60 to .95.

Graham, Smith, and Schwartz (1986) studied the stability of MMPI high-point, low-point, and two-point code types in a sample of 405 male and female psychiatric patients who had completed the MMPI on two occasions with an average of 80 days between test administrations. Test–retest agreement for high-point and two-point codes was 42.72% and 27.65%, respectively. They found that code types

were somewhat more consistent when scales in the initial code-type had more extreme scores and when scores on scales in the codes were markedly different from other scales in the profile. Graham et al. (1986) suggested that psychologists should be careful about making long-term predictions from a single MMPI.

Ryan, Dunn, and Paolo (1995) studied the test–retest stability of the standard validity and clinical scales and supplemental scales of the MMPI–2 in a sample of male substance abusers. The subgroup of participants with a retest interval of about 5 months had single-, two-, and three-point code type agreements across test administrations of 31.4%, 19.6%, and 13.7%, respectively. Test–retest correlations for this group ranged from .63 to .85 across the Basic scales and .57 to .84 across the Supplementary scales. The subgroup with a retest interval of about 13 months had code-type agreement rates for one-, two-, and three-point code types of 34.7%, 12.2%, and 6.1%. Test–retest correlations across the Basic Scales ranged from .44 to .81 and ranged from .60 to .83 across the Supplementary Scales. At least 20% of participants in the 5-month retest interval showed *T*-score differences ≥ 10 points from test to retest on 12 of the 25 scales and at least 20% of the participants in the 13-month retest interval showed *T*-score changes ≥ 10 points on 18 of the 25 scales. Ryan et al. (1995) noted that because the majority of substance abusers produced different patterns of elevated scores across the two assessments, clinicians should exercise caution in making long-term predictions about an individual from a single administration of the MMPI–2.

Skinner and Jackson (1978) developed a model of psychopathology based on an integration of MMPI actuarial systems. They noted the importance of considering profile elevation, scatter, and shape when matching individual MMPI profiles to code type profiles. They identified three modal MMPI profile types: neurotic, psychotic, and sociopathic, which may serve as a three-dimensional frame of reference to consider the shape of MMPI profiles. A person's MMPI profile may be represented in this three-dimensional model by calculating the Pearson product–moment correlation coefficient between an individual's *T* scores across the 13 Basic scales with each of the three modal profiles representing the three dimensions of the model. Skinner and Jackson hypothesized that profile shape parameters may provide indexes of more enduring personality disposition whereas elevation and scatter may reflect more temporary or situational factors influencing the extent of maladjustment. Munley and Zarantonello (1990) used this model in considering profile changes for classic MMPI profiles on transformation to MMPI–2 norms; however, the model has apparently not yet been used as a frame of reference for considering MMPI or MMPI–2 profile configurations over time.

The purpose of this study was to examine MMPI–2 scale and profile comparablilty for MMPI–2 profiles obtained from mental health patients who had completed the MMPI–2 on two occasions as a routine part of their treatment. Given the limited studies on test–retest comparablilty of the MMPI–2 among psychiatric and mental health patients over extended periods of time, this study's aim

was to consider scale and profile comparablilty among mental health patients who had completed the test on two different occasions at more wide-ranging test–retest intervals than is found in many of the previously published studies on the temporal stability of the MMPI and MMPI–2. In particular, profiles completed on two different occasions as a routine part of mental health treatment that were separated by a minimum of a 1-month time interval were studied. MMPI–2 scale test–retest correlation coefficients, scale high-point agreement, and changes in profile characteristics were studied. Profile characteristics studied over time included: individual scale means, absolute values of difference scores, profile elevation, profile scatter, test–retest profile similarity, and profile similarity to Skinner and Jackson's three modal profile types.

METHOD

Participants

The sample consisted of a group of 114 mental health patients receiving treatment at a Veterans Affairs Medical Center who had completed the MMPI–2 on two separate occasions as a routine part of their treatment. The sample was identified retrospectively from testing files from among veteran patients who had completed the MMPI–2 as a part of their evaluation for psychological and mental health services on at least two separate occasions. Participants needed to have had valid MMPI–2 profiles on both occasions. Profile validity was defined in terms of: (a) no more than 30 items omitted, (b) *F* scale score < raw score of 26, (c) FB scale score < raw score of 26, and (d) VRIN scale raw score < score of 13. Patients with F and FB scores > 25 were excluded because of the likelihood that scores in this range, even when obtained among psychiatric inpatients, indicate invalid profiles. Participants with VRIN scores > 12 were also excluded because such scores probably indicate a random response pattern (Graham, 2000).

The sample was predominantly male, with 110 (96.5%) men and 4 women (3.5%). At the time of initial testing, the sample had an average age of 44.08 (*SD* = 7.48) and an average of 12.39 (*SD* = 1.97) years of education. Marital status was as follows: 14 single (12.3%), 48 married or married/currently separated (42.1%), 52 divorced (45.6%). The participants represented the following ethnic-racial groups: White (75.4 %), African American (21.9%), Hispanic (1.8%), Native American (0.9%). At the time of the first MMPI–2 testing a majority of the participants (104) were inpatients for psychiatric treatment and/or residential treatment for specialized mental health services (e.g., residual treatment for posttraumatic stress disorder [PTSD], residential treatment for substance abuse), and 10 were outpatients. At the time of second testing 89 were receiving inpatient and/or residential treatment and 25 were outpatients. MMPI–2s completed by patients receiving inpatient

or residential treatment were usually completed near the beginning of their treatment episode. Second MMPI–2s were usually completed in the context of a different hospitalization or different episode of care. Principle diagnoses at the time of the initial testing were: adjustment reaction (n = 13, 11.4%), bipolar disorder (n = 7, 6.1%), dysthymic disorder/depression NOS (n = 8, 7%), major depression (n = 13, 11.4%), mental disorder due to organic condition (n = 3, 2.6%), other psychosis (n = 2, 1.8%), personality disorder (n = 4, 3.5%), PTSD (n = 45, 39.5%), schizophrenia (n = 6, 5.3%), substance abuse (n = 12, 10.5%), and other/not specified (n = 1, 0.9%).

Length of time between the two MMPI–2 administrations ranged from 38 to 2,576 days with a median of 431.5 days, and a mean of 688.4 days (SD = 642.77). The current sample was classified into three subgroups based on length of time between the two MMPI–2s. Subgroup 1 consisted of 54 (47.4%) patients who completed the second MMPI–2 within 1 year (M = 209.46 days, SD = 89.29). Subgroup 2 consisted of 37 (32.4%) patients who completed the second MMPI–2 within 1 to 3 years (M = 695.54 days, SD = 212.57). Subgroup 3 consisted of 23 (20.2%) patients who completed their second MMPI–2 within 3 to 8 years, 4 (3.5%) in 3 to 4 years; 11 (9.7%) in 4 to 5 years; and 8 (7%) within 5 to 8 years; M = 1801.43 days, SD = 413.15). The three subsamples did not differ significantly on initial testing in terms of age, $F(2, 111) = .54$, $p = .58$, or education, $F(2, 111) = .04$, $p = .96$. Distribution of initial primary diagnoses grouped according to six major categories across Subgroups 1, 2, and 3, respectively, was psychosis (11.1%, 5.4%, 13.04%), major mood disorder (9.26%, 18.92%, 34.78%), nonpsychotic depression (11.1%, 5.4%, 0%), personality disorder/substance abuse (18.52%, 13.51%, 4.35%), adjustment reaction (9.26%, 18.92%, 4.35%) and PTSD (40.74%, 37.84%, 39.13%). Although differences in the distribution of diagnoses for the subgroups across the six categories were not statistically significant $\chi^2(10, N = 113) = 15.80$, $p = .11$, Subgroup 3, the smallest subsample and the group with the longest test–retest interval, did have the highest percentage of psychosis and major mood disorder diagnoses and the lowest percentage of nonpsychotic depression, adjustment disorder, and personality disorder/substance abuse diagnoses.

Procedure

Means and standard deviations of the T scores with K correction from the Basic, Supplementary, and Content scales were obtained for the two test administrations.

Individual scale comparisons. For each scale, Pearson product–moment correlations and paired t tests with the Bonferroni correction were calculated. Also, the absolute T-score differences for each individual's two scores on each scale were calculated.

Profile characteristics comparisons. The *D* statistic (Osgood & Suci, 1952) was calculated by taking the square root of D^2, where D^2 is the sum of the squared differences between scores on each pair of Basic scales across the two MMPI–2s for each participant. The modified *D* statistic (Chojnacki & Walsh, 1992) was calculated by summing the absolute differences between scales for the two administrations across the basic validity and clinical scales. The modified *D* reflects the number of *T*-score points that a patient's profile varies across the Basic scales from one test to the other. Because some investigators have calculated *D* statistics based only on the clinical scales and others have used the 13 Basic scales, *D* and modified *D* statistics in this study were calculated for all 13 Basic scales and also calculated based only on the 10 clinical scales. Profile high-point, high two-point, and high three-point scale consistency was considered across pairs of profiles. High-point agreement occurred when the participant's most elevated scale was the same on both MMPI–2 profiles. High two-point agreement occurred when the participant's two most elevated scales were the same for both profiles. High three-point agreement occurred when the participant's three highest elevated scales were the same for both profiles. In the case of a tie between two scores for defining position within a code type (e.g., tie for high-point, two-point, or three-point), both scales were considered high points within the profile for matching purposes and the second profile had to match only one of the tied scales to be considered consistent. Scale order was not considered for two-point or three-point consistency. The high-point agreement comparison was conducted for the entire sample, the three sample subgroups classified by test–retest interval, and for subgroups that had initial well-defined high points, high two-points, or high three-points (e.g., at least 5 *T*-score points between the scale in the code type and the next highest scale in the profile).

To consider changes in profile elevation and scatter, mean profile elevations and profile standard deviations about profile means were calculated across the two tests for the Basic 13 scales for each pair of profiles. Absolute value differences in individual profile means and individual profile standard deviations were also examined. To consider extent of profile shape changes, Pearson product–moment correlation coefficients for *T* scores across the Basic 13 scales between each pair of individual MMPI–2 profiles were calculated. Profile changes were also considered in reference to Skinner and Jackson's (1978) model. Pearson product–moment correlation coefficients were calculated between each individual's MMPI–2 *T* scores across the 13 Basic scales for each testing with each of Skinner and Jackson's three modal profiles. Following calculations of the correlations between individual profiles and Skinner and Jackson's (1978) modal profiles, each profile was classified in two ways. First, each profile was classified based simply on the profile's highest correlation with a modal profile type. Second, a minimum correlation of .50 was required with the modal profile type for classification. Profile classification agreement by each method for the two sets of profiles obtained was then considered.

RESULTS

Participants in this study obtained elevated profiles across both testing occasions. For Test 1, participants averaged a mean of 6.12 (*SD* = 2.5) scales elevated ≥ *T* score 65 for the 10 clinical scales; for Test 2 participants averaged a mean of 6.32 (*SD* = 2.4) scales elevated ≥ *T* score 65 for the 10 clinical scales. Tables 1, 2, and 3 present the basic results of the MMPI–2 scale comparisons across the two testing episodes. Paired *t* tests with Bonferroni-adjusted *p* levels of .001 were not significant for any of the scale mean comparisons. Because equal increases and decreases in changes in *T* scores cancel out across participants, the means and standard deviations of the absolute value of the difference scores for individual participants reflect the average amount of *T*-score change in either direction shown by participants across testing sessions within individual scales. Although overall the scale means did not differ significantly from the first to the second testing, the means and standard deviations of the absolute value difference scores for each participant reflect substantial *T*-score changes occurring for many participants across testing sessions. Across the Basic scales, mean absolute value difference scores range from 5.90 to 13.31 with scales F, 1, 3, 6, 7, and 8 showing some of the largest mean absolute value difference scores. Across the Supplementary scales, mean absolute value difference

TABLE 1
MMPI–2 Basic Profile Scale Means, Standard Deviations, Paired *t* Tests, Scale Correlations, and Absolute Score Mean Differences and Standard Deviations

	Test 1		*Test 2*			*Absolute Difference*			*Subgroup Test Interval*		
Scale	*M*	*SD*	*M*	*SD*	*Paired t*	*M*	*SD*	*Total Sample r*	*< 1 Year*[a]	*1 to 3 Years*[b]	*> 3 Years*[c]
L	51.61	9.61	51.61	9.54	0.01	7.04	6.31	.51	.59	.45	.45
F	80.52	18.12	80.18	18.01	0.22	12.88	9.97	.59	.50	.67	.65
K	40.93	9.16	40.85	9.85	0.10	5.90	6.20	.59	.60	.64	.63
1(Hs)	71.87	14.33	73.62	14.81	–1.33	11.07	8.88	.53	.52	.63	.35
2(D)	77.91	13.79	79.82	12.62	–1.84	8.79	6.96	.65	.67	.60	.70
3(Hy)	71.35	15.47	72.78	15.65	–1.05	10.92	9.58	.57	.61	.65	.27
4(Pd)	73.27	11.98	71.61	11.91	1.67	8.71	6.22	.61	.66	.62	.42
5(Mf)	51.47	8.75	51.51	8.84	–0.05	5.95	4.68	.63	.70	.65	.32
6(Pa)	75.69	16.89	75.54	15.45	0.10	13.31	9.72	.48	.46	.47	.54
7(Pt)	78.13	15.60	79.20	14.82	–0.88	10.63	7.59	.63	.63	.73	.55
8(Sc)	81.90	16.72	82.37	17.47	–0.32	12.52	8.90	.60	.55	.72	.54
9(Ma)	58.77	12.35	57.62	13.56	1.04	8.92	7.83	.59	.54	.72	.57
0(Si)	65.20	11.65	66.61	11.59	–1.64	6.75	6.41	.69	.59	.78	.77

Note. $N = 114$.
[a]$n = 54$. [b]$n = 37$. [c]$n = 23$.

TABLE 2
MMPI–2 Supplementary Scale Means, Standard Deviations, Paired *t* Tests, Scale Correlations, and Absolute Score Mean Differences and Standard Deviations

	Test 1		*Test 2*			*Absolute Difference*			*Subgroup Test Interval*		
Scale	*M*	*SD*	*M*	*SD*	*Paired t*	*M*	*SD*	*Total Sample r*	*< 1 Year*[a]	*1 to 3 Years*[b]	*> 3 Years*[c]
A	71.61	11.40	72.11	12.04	–0.59	6.27	6.48	.71	.72	.76	.65
R	53.89	11.70	55.44	11.76	–2.06	6.23	5.24	.77	.82	.81	.57
ES	33.01	6.37	32.67	6.17	0.68	2.62	4.68	.64	.52	.75	.83
MACR	60.19	11.00	58.93	11.43	1.42	7.33	6.12	.64	.75	.68	.35
OH	46.47	9.78	47.46	9.95	–1.06	7.57	6.54	.49	.65	.43	.12
DO	35.11	7.71	35.03	7.18	0.13	4.01	5.16	.62	.73	.52	.33
RE	38.76	9.02	38.88	8.84	–0.21	3.96	4.07	.80	.88	.72	.72
MT	75.82	11.12	75.51	11.73	0.32	6.68	7.55	.61	.53	.68	.73
GM	37.32	8.39	36.11	7.83	1.66	5.02	6.07	.54	.44	.49	.80
GF	44.43	8.69	44.19	8.58	0.31	5.97	5.36	.57	.80	.43	.12
PK	82.88	15.76	83.02	15.76	–0.12	9.68	7.72	.69	.65	.74	.72
PS	83.98	15.41	84.53	16.51	–0.46	9.82	8.06	.68	.62	.77	.70

Note. $N = 114$.
[a]$n = 54$. [b]$n = 37$. [c]$n = 23$.

TABLE 3
MMPI–2 Content Scales Means, Standard Deviations, Paired *t* Tests, Scale Correlations, and Absolute Score Mean Differences and Standard Deviations

	Test 1		*Test 2*			*Absolute Difference*			*Subgroup Test Interval*		
Scale	*M*	*SD*	*M*	*SD*	*Paired t*	*M*	*SD*	*Total Sample r*	*< 1 Year*[a]	*1 to 3 Years*[b]	*> 3 Years*[c]
ANX	74.97	11.91	74.57	12.39	0.47	7.16	5.69	.72	.66	.77	.75
FRS	58.32	12.60	59.66	11.73	–1.46	8.04	5.70	.68	.68	.69	.65
OBS	64.72	12.24	66.02	12.26	–1.44	7.60	6.03	.69	.77	.63	.63
DEP	78.15	12.17	77.96	12.48	0.18	8.74	7.56	.56	.66	.40	.55
HEA	73.62	14.29	75.49	14.04	–1.50	10.03	8.83	.56	.46	.63	.60
BIZ	67.57	15.86	66.02	15.60	1.20	10.45	9.07	.62	.52	.67	.65
ANG	67.47	13.29	67.50	13.13	–0.03	6.54	5.95	.78	.84	.84	.53
CYN	62.11	11.82	60.39	11.85	1.95	7.44	6.00	.68	.74	.71	.59
ASP	59.70	12.32	57.86	11.57	2.00	7.75	6.28	.66	.76	.52	.70
TPA	58.26	13.16	58.39	13.20	–0.13	7.68	7.15	.68	.74	.85	.31
LSE	68.71	13.11	68.86	13.54	–0.13	8.57	8.25	.60	.73	.62	.32
SOD	65.09	13.59	67.73	13.69	–2.80	7.54	7.17	.73	.64	.86	.73
FAM	65.44	12.74	65.68	13.24	–0.24	8.61	6.24	.66	.70	.59	.69
WRK	71.65	12.71	72.25	12.03	–0.61	7.82	7.13	.64	.60	.74	.63
TRT	72.61	13.81	73.67	14.49	–0.93	9.68	7.49	.63	.68	.59	.63

Note. $N = 114$.
[a]$n = 54$. [b]$n = 37$. [c]$n = 23$.

scores ranged from 2.62 to 9.82; for the Content scales mean absolute value difference scores ranged from 6.54 to 10.45. Tables 1, 2, and 3 also present the test–retest Pearson *r* correlation coefficients for scale scores for the entire sample and for the three sample subgroups defined by test–retest interval. For the sample as a whole, test–retest correlations ranged from .48 to .69 for the Basic scales, .49 to .80 for the Supplementary scales, and .56 to .78 for the Content scales. Correlations across the scales for the three sample subgroups tended to be similar, however, for some scales the correlations were notably lower for the subgroup with the test–retest interval of more than 3 years.

Table 4 presents the profile characteristics for the first and second MMPI–2s. High-point agreement, high two-point agreement, and high three-point agreement for the entire sample were 38.60%, 16.67%, and 19.30%, respectively. When subsets of participants with well-defined high points, two-points, and three-points on initial testing were considered separately, their respective high-point agreement figures with their second MMPI–2 were somewhat higher: 41.07%, 27.50%, and 25.93%. Percentages of initial well-defined high points, high two-points, and high three-points for each subsample are also presented in Table 4. High-point scale agreement was not significantly different for the three subsamples with different test–retest intervals, $\chi^2(2, N = 114) = 3.82$, $p = .15$. Similarly, high two-point agreement, $\chi^2(2, N = 114) = 1.20$, $p = .55$, and high three-point agreement, $\chi^2(2, N = 114) = 2$, $p = .28$, were not significantly different across the three subsamples. Table 4 also presents the *D* and modified *D* statistics. Modified *D* statistics indicate that individual profiles changed an average of 123.39 *T*-score points across the 13 Basic scales and an average of 97.56 across the 10 clinical scales.

Considering profile elevation and scatter and shape characteristics, mean profile elevation across the 13 Basic scales for the first and second MMPI–2 averaged around 68 *T*-score points with an average standard deviation of profile *T* scores around the profile mean of about 17. Across the two sets of test profiles, individual profile absolute value mean elevation change averaged 5.67 ($SD = 4.12$) and absolute value profile standard deviation change averaged 2.94 ($SD = 2.66$). Profile elevation means and standard deviations about profile means were similar for the entire sample and for the three subsamples with the different test–retest intervals. Shape characteristics for Test 1 and Test 2 profiles were similar with Test 1 and Test 2 *T* scores across Basic scale profiles correlating on average .78 ($Mdn = .84$, $SD = .19$); and with a similar pattern of correlations seen between individual profiles and Skinner and Jackson's (1978) modal profile types. Classifying Test 1 and Test 2 profiles based simply on their highest correlation with one of Skinner and Jackson's modal profile types indicated 23 (20.2%) of the Test 1 profiles correlated highest with the neurotic modal profile type, 87 (76.3%) correlated highest with the psychotic modal profile type, and 4 (3.5%) correlated highest with the sociopathic modal profile type. For Test 2 profiles, 20 (17.5%) correlated highest with neurotic, 93 (81.6%) with psychotic, and 1 (0.9%) with the sociopathic modal profile types. Requiring a mini-

TABLE 4
Profile Characteristics of Test 1 and Test 2

Profile Similarity Characteristics	*Total Sample*	*Test Interval*		
		< 1 Year[a]	*1 to 3 Years*[b]	*> 3 Years*[c]
Profile highest scale agreement[d]				
High point agreement	38.60	33.33	51.35	30.43
Well-defined initial high points[d]	49.12	42.59	59.46	47.83
High two-point agreement	16.67	12.96	21.62	17.39
Well-defined initial high two-points	35.09	37.04	37.84	26.09
High three-point agreement	19.30	14.81	18.92	30.43
Well-defined initial high three-points	23.68	27.78	13.51	30.43
Profile distance statistics				
10 clinical scales				
D statistic				
M	37.60	36.77	37.07	40.38
SD	14.08	13.03	14.59	15.85
Modified *D*				
M	97.56	94.94	96.22	105.87
SD	38.64	35.88	38.50	45.26
13 Basic scales				
D statistic				
M	42.59	42.14	41.65	45.17
SD	14.98	15.01	14.65	15.81
Modified *D*				
M	123.39	121.74	120.19	132.39
SD	45.81	46.11	43.12	50.07
Profile elevation, scatter, and shape characteristics				
Profiles Test 1				
Mean profile elevation across 13 Basic scales	67.59	67.46	67.51	68.02
Mean *S*D of *T* scores around profile mean across 13 Basic scales	16.87	16.06	17.67	17.48
Mean correlation with Skinner's modal neurotic profile	0.23	0.26	0.16	0.28
Mean correlation with Skinner's modal psychotic profile	0.62	0.60	0.66	0.62
Mean correlation with Skinner's modal sociopathic profile	0.01	0.00	0.00	0.04
Profiles Test 2				
Mean profile elevation across 13 Basic scales	67.95	66.13	70.09	68.78
Mean *SD* of *T* scores around profile mean across 13 basic scales	16.91	15.86	17.91	17.79
Mean correlation with Skinner's modal neurotic profile	0.28	0.29	0.22	0.35
Mean correlation with Skinner's modal psychotic profile	0.64	0.63	0.67	0.61
Mean correlation with Skinner's modal sociopathic profile	–0.04	–0.08	0.00	0.01
Pearson *r* Profile 1 with Profile2				
M	0.78	0.77	0.81	0.75
SD	0.19	0.19	0.18	0.22
Skinner modal profile classification agreement (highest correlation)[d]	78.07	77.77	83.79	69.57
Skinner modal profile classification agreement (.50 criteria)[d]	73.68	72.22	75.67	73.92

Note. $N = 114$.
[a]$n = 54$. [b]$n = 37$. [c]$n = 23$. [d]Given in percentages.

mum correlation of .50 to be considered similar to a modal profile type changed the findings to 14 (12.3%), 84 (73.7%), and 1 (.9%) respectively for Test 1 profiles, with 15 (13.2%) then considered indeterminant; and 18 (15.8%), 87 (76.3%), and 1 (0.9%), respectively, for Test 2 profiles, with 8 (7%) considered indeterminant. Using the highest correlation with one of the modal profile types to classify profiles, agreement across the two testing occasions was 78% with 89 of the 114 pairs of profiles receiving the same classification on both occasions: neurotic 11(9.7%), psychotic 78 (68.4%), and sociopathic 0 (0%); $\chi^2(4, N = 114) = 23.32, p < .001$, Cohen's $\kappa = .36$. Requiring a minimum .50 correlation to be considered similar to a modal profile type resulted in the following profile pair agreement figures: 6 (5.3%) neurotic, 75 (66%) psychotic, 0 (0%) sociopathic, and 3 (2.6%) indeterminant with a total of 84 or 73.7% receiving the same classification on both profiles; $\chi^2(9, N = 114) = 37.07$, $p < .0001$, Cohen's $\kappa = .36$.

DISCUSSION

Current findings comparing the individual MMPI–2 Basic, Supplementary, and Content scales across two different test administrations among mental health patients over extended periods of time seem similar to other research findings on the temporal stability of the MMPI–2 (Putnam et al., 1996; Ryan et al., 1995). Pearson *r* correlations across the basic scales between the two test administrations are lower than test–retest correlations found for the normative sample retested at a 1-week interval (Butcher et al., 1989), lower than those obtained in other studies of the MMPI–2 with college students (Butcher et al., 1990), and lower than male clergy not in mental health treatment retested at a 4-month interval (Putnam et al., 1996), but seem somewhat similar to findings by Ryan et al. (1995) for male substance abusers, particularly their group retested at a 13-month interval. For the Basic scales, correlations appear to be in the typical range for psychiatric patients retested with the MMPI at an interval of 1 year or more (.50 to .60), as indicated by Schwartz's 1977 study (as cited in Graham, 2000). As also observed by Putnam et al. (1996), test–retest correlation coefficients tended to be somewhat higher for the Supplementary and Content scales than the Basic scales. Although correlation coefficients for the MMPI and MMPI–2 may be expected to be lower over time, particularly for mental health patients and individuals being evaluated during times of stress, current correlations do suggest similar personality features and clinical symptom presentation features as measured by the MMPI–2 over extended periods of time.

Absolute value difference scores at the individual scale level of the individual participants serve, however, as an indication and reminder of the extent of *T*-score change that may take place, on average over time on retesting, at the level of individual participants individual scale scores. In this study mean absolute difference scores for the Basic scales ranged from 5.90 to 13.31. Absolute difference score

means and standard deviations found in this study are noticeably larger than those found by Putnam et al. (1996). Also, *D* and modified *D* statistics obtained in this study appear notably larger than those obtained by Chojnacki and Walsh (1992) in their findings on the consistency of the MMPI taken twice by college students. These differences may reflect significantly more *T*-score change and variability over time at the individual scale and profile level for individuals retested at longer intervals and for persons receiving inpatient and outpatient mental health treatment. As Ryan et al. (1995) observed, the amount of change seen in *T* scores at the level of an individual person completing the test is greater than clinicians might expect from a consideration of test–retest coefficients and mean group differences across time. The extent of absolute *T*-score differences perhaps highlights the sensitivity to change of profile code types defined by scale high points.

Current findings for agreement on one-point (38.60%), two-point (16.67%), and three-point (19.30%) highest scale agreement are, as might be expected, generally someone lower than respective code type agreement of 49%, 26%, and 15% obtained in the normative MMPI–2 sample at a 1-week test–retest interval (Graham et al., 1991) and somewhat lower than code type agreement reported by Graham et al. (1986) for the MMPI with psychiatric patients at an average 80-day retest interval (42.72% high-point and 27.65 % high two-point agreement), but similar to code type agreement with the MMPI–2 reported by Ryan et al. (1995) for substance abusers retested at 5 months (31.4%, 19.6%, and 13.7%, respectively) and 13 months (34.7%, 12.2%, and 6.1%). When subsets of the current sample were considered that had initial well-defined high-point, two-point, and three-point profiles, highest scale agreement increased to 41.07%, 27.50%, and 25.93%, respectively, which, although somewhat higher, still indicated only small to modest highest scale agreement across the two profiles. Scale high-point agreement did not differ significantly statistically across the three subsamples with different test–retest intervals. However, the relatively higher percentage of single-scale agreement for the subsample with the 1- to 3-year test–retest interval and the relatively higher percentage of three-point agreement for the subsample with the test–retest interval of more than 3 years is interesting to note. This relatively higher agreement may relate in part to the higher percentage of well-defined high points and high three-points, respectively, obtained by these subsamples on their initial MMPI–2s. Profile definition on initial testing may be a factor in high-point agreement over time. Overall, both well-defined and unrestricted initial profiles in this study showed only small to modest highest scale agreement over time. As noted by Graham et al. (1986) for the MMPI and Ryan et al. (1995) for the MMPI–2, a substantial proportion of highest scale code types change over time, and care is indicated in making longer term inferences based on a single MMPI–2.

Although MMPI–2 profile code types, as defined by highest scale scores, do show a substantial amount of change over time, significant profile similarity still seems to be evident on repeat testing in terms of similarity of overall profile shape.

This is reflected by the moderately high Pearson *r* correlation coefficients comparing individual pairs of test–retest profiles ($M = .78$, $Mdn = .84$, $SD = .19$), the similarity in pattern of Pearson *r* correlations between individual profiles and Skinner and Jackson's (1978) three modal profile types, and the 73.7% to 78% classification consistency with one of the three modal types across time. Overall, profile shape across the two testing episodes seems to reflect more profile consistency than indicated by scale high-point agreement percentages alone. Some of the difficulties with simply utilizing scale high points to specify code types are well known. The potential role of measurement error in defining high-point code types, when code types are based simply on the elevation of the highest 1, 2, or 3 clinical scales in a single MMPI–2 administration, has been recognized (Munley, 1991), and Graham (2000) recommended only interpreting defined code types. In considering some of the indications noted in this study of profile shape similarity over time, particularly in the context of high-point code type changes observed over time in this and other studies (Graham et al., 1986; Graham et al. 1991, Ryan et al., 1995) and in the context of the amount of absolute *T*-score differences seen for individuals within individual scales on repeat testing in this study and in other studies (Putnam et al., 1996; Ryan et al., 1995), Skinner and Jackson's (1978) hypothesis concerning profile shape parameters seems important. The basic idea that profile shape parameters might provide indexes of a patient's more enduring personality features, whereas elevation and scatter measures might reflect more temporary and situational factors influencing extent of maladjustment, seems intriguing and worthy of further study.

This study has a number of limitations that should be noted. The sample consists primarily of inpatients, who were likely experiencing significant distress on both testing occasions. In addition to patients with major mental disorders including psychosis, schizophrenia, bipolar disorder, and major depression, the sample included a high percentage of patients with PTSD who are also known to obtain very elevated MMPI–2 profiles. The predominance of inpatients and patients with elevated MMPI–2 profiles may have influenced current findings in terms of the overall pattern of predominance of elevated profiles correlating highest with the Skinner and Jackson (1978) psychotic modal profile type and correlating highly with subsequent respective second test profiles obtained during a second episode of care. Because participants in this study were often undergoing a second or subsequent hospitalization or period of residential mental health care at the time of their second testing, they may also have suffered from more severe or chronic forms of their mental health disorders. MMPI–2 profiles obtained from persons experiencing less severe mental health disorders or distress, or from persons obtaining profiles that show less elevation and less scatter about profile means, may show smaller Pearson *r* profile test–retest correlations and different classification agreement over time in terms of similarity to the Skinner and Jackson modal profiles types. In addition, this investigation did not control for actual changes in clini-

cal conditions and/or symptoms over time. Actual symptom and clinical condition changes that may accompany MMPI–2 profile changes over time are also very important components to understanding profile change. Future research may help clarify some of these questions. Understanding the nature and extent of MMPI–2 profile configural changes seen over time is important for clinicians. Clinicians who may assess patients over extended time periods may be helped by appreciating the nature and extent of both MMPI–2 profile changes and MMPI–2 profile similarities that may be seen over time in patients presenting for evaluation at different times in their treatment history.

ACKNOWLEDGMENTS

This article is based on research supported by the Veterans Affairs Medical Center, Battle Creek, MI. Portions of this article were presented as a poster session at the 109th Annual Convention of the American Psychological Association, San Francisco, August 2001.

REFERENCES

Ben-Porath, Y. S., & Butcher, J. N. (1989). The comparability of MMPI and MMPI–2 scales and profiles. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 4,* 345–347.

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

Butcher, J. N., Graham, J. R., Dahlstrom, W. G., & Bowman, E. (1990). The MMPI–2 with college students. *Journal of Personality Assessment, 54,* 1–15.

Chojnacki, J. T., & Walsh, W. B. (1992). The consistency of scores and configural patterns between the MMPI and MMPI–2. *Journal of Personality Assessment, 59,* 276–289.

Graham, J. R. (2000). *MMPI–2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

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

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

Munley, P. H. (1991). Confidence intervals for the MMPI–2. *Journal of Personality Assessment, 57,* 52–60.

Munley, P. H., & Zarantonello, M. M. (1990). A comparison of MMPI profile types with corresponding estimated MMPI–2 profiles. *Journal of Clinical Psychology, 46,* 803–811.

Osgood, C. E., & Suci, G. J. (1952). A measure of relation determined by both mean difference and profile information. *Psychological Bulletin, 49,* 251–262.

Pauker, J. D. (1966). Stability of MMPI profiles of female psychiatric inpatients. *Journal of Clinical Psychology, 22,* 209–212.

Putnam, S. H., Kurtz, J. E., & Houts, D. C. (1996). Four-month test–retest reliability of the MMPI–2 with normal male clergy. *Journal of Personality Assessment, 67,* 341–353.

Ryan, J. J., Dunn, G. E., & Paolo, A. M. (1995). Temporal stability of the MMPI–2 in a substance abuse sample. *Psychotherapy in Private Practice, 14,* 33–41.

Skinner, H. A., & Jackson, D. N. (1978). A model of psychopathology based on an integration of MMPI actuarial systems. *Journal of Consulting and Clinical Psychology, 46,* 231–238.

Von Cleve, E., Jemelka, R., & Trupin, E. (1991). Reliability of psychological test scores for offenders entering a state prison system. *Criminal Justice and Behavior, 18,* 159–165.

Patrick H. Munley
Department of Counselor Education and Counseling Psychology
3102 Sangren Hall
Western Michigan University
Kalamazoo, MI 49008–5226
E-mail: patrick.munley@wmich.edu