Psychological Assessment
2009, Vol. 21, No. 2, 137–151

1040-3590/09/$12.00 DOI: 10.1037/a0015316

Racial Bias in Personality Assessment: Using the MMPI-2 to Predict Psychiatric Diagnoses of African American and Caucasian Chemical Dependency Inpatients

Matthew J. Monnot, Stuart W. Quirk, Michael Hoerger, and Linda Brewer
Central Michigan University

An assessment of predictive bias was conducted on numerous scales of the Minnesota Multiphasic Personality Inventory–2 (MMPI-2; J. N. Butcher, W. G. Dahlstrom, J. R. Graham, A. Tellegen, & B. Kaemmer, 1989), including the Restructured Clinical (RC) scales, in the prediction of clinical diagnostic status for African American and Caucasian male veterans seeking substance abuse treatment. Patients completed a battery of self-report instruments and were administered structured diagnostic interviews. African American patients obtained higher scores across most MMPI-2 scales compared with Caucasians with clinically meaningful elevations (T scores $>$ 5 points) on 3 scales. The RC scales demonstrated strong correlations with diagnoses, however, like other MMPI-2 scales examined in this study, they displayed a general trend of predictive bias. Step-down hierarchical regression procedures (G. J. Lautenschlager & J. L. Mendoza, 1986) indicated the presence of predictive bias for a majority of the scales examined; however, most of these effects were small to modest (accounting for 3%–5% of variance). The pattern of slope and intercept biases across types of MMPI-2 scales differs from prior research and indicates the importance of evaluating bias in various populations and settings.

Keywords: MMPI-2, RC scales, racial bias

Test bias in psychopathology assessment is an important concern due to the magnitude of decisions made as a result of such assessments. The most widely used self-report inventory of psychopathology, the Minnesota Multiphasic Personality Inventory–2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), has been criticized as potentially biased in predicting psychiatric status of racial minorities (Adebimpe, 1981; Aponte & Johnson, 2000). However, empirical efforts have yielded equivocal results with regard to both mean differences and biased predictions of clinically relevant extratest variables (Arbisi, Ben-Porath, & McNulty, 2002; McNulty, Graham, Ben-Porath, & Stein, 1997; Timbrook & Graham, 1994). Additional research of racial bias using relevant clinical criteria (such as symptom reports or diagnostic status) in various clinical settings and with differing patient populations is required in order to clarify previously discrepant findings. Such research is also necessary to help determine the importance of patient group and clinical setting with regard to predictive bias in the MMPI-2.

Early efforts to explore racial bias in the original MMPI and subsequent MMPI-2 examined group differences in mean scale elevations (e.g., Gynther, 1972). Evaluating mean differences between ethnic minorities and Caucasian groups has yielded ambiguous results. For example, researchers (Frueh, Smith, & Libet, 1996) have reported that the profiles of African American veterans indicated greater maladjustment (i.e, higher mean scores on Scales 6 and 8) compared with Caucasian patients; however, this effect was not replicated (Frueh, Gold, de Arellano, & Brady, 1997). Such inconsistency is reflected in a meta-analysis of scale differences between African American and Caucasian groups (Hall, Bansal, & Lopez, 1999). A fundamental limitation of this line of research is that the presence of mean differences between groups is insufficient for establishing test bias. Mean differences may reflect genuine differences between groups or settings rather than biases in clinical conclusions or behavioral predictions (Archer, Griffin, & Aiduk, 1995).

Tests of predictive equivalence require the inclusion of relevant external criteria against which patterns of prediction can be scrutinized (Arbisi et al., 2002; Morrison, Edwards, & Weissman, 1994; Munley, Busby, & Jaynes, 1997). Small but reliable mean differences between Caucasian and African Americans have been found without a corresponding difference in the prediction of psychopathology in normal samples (McNulty et al., 1997) and partner ratings of interpersonal behaviors and personality characteristics (Timbrook & Graham, 1994).

In one of the most comprehensive evaluations of racial bias in the MMPI-2 to date, Arbisi and colleagues (2002) collected clinically relevant criterion information from inpatient charts in a general inpatient psychiatric hospital setting and a large VA medical center to evaluate relevant MMPI-2 scales for racial bias. They examined the comparative associations between MMPI-2 scales and criterion variables by group status (tests of slope bias) as well as systematic over- or underprediction for each group (intercept bias). In this general sample of inpatients, the MMPI-2 yielded a number of small but reliable intercept bias effects indicating un-

Matthew J. Monnot, Stuart W. Quirk, Michael Hoerger, and Linda Brewer, Department of Psychology, Central Michigan University.

Linda Brewer is now at Central Arkansas Veterans Healthcare System.

Correspondence concerning this article should be addressed to Matthew J. Monnot, who is now at Organization Development, Genentech, South San Francisco, CA 94080. E-mail: monnot.matthew@gene.com

derprediction of psychopathology in the African American compared with Caucasian patients.

Important questions remain with regard to MMPI-2 utility, and differences in utility, among Caucasians and African Americans. These include the need to evaluate the possibility of biased predictions for specific psychiatric diagnoses in specific settings, to do so using the relatively new Restructured Clinical scales (RC; Tellegen et al., 2003), and to expand the types of extratest criteria used. The powerful influence of local base rates pon the conclusions, utility, and appropriateness of assessment instruments has been long discussed in assessment literature (e.g., Elwood, 1993; Meehl, 1954) but infrequently evaluated (Stukenberg, Brady, & Klinetob, 2000). Additionally, clarification of test bias (or lack thereof) in the MMPI-2 can be established by examining the instrument's performance across types of clinical settings in which the instrument is widely used. A main purpose of the present study was to evaluate the MMPI-2 in predicting diagnostic status in an inpatient substance abuse treatment setting with male veterans. Additional contributions of the present study include (a) the use of structured interviews for diagnosis as the external criterion, (b) a large inpatient sample, (c) the evaluation of the RC scales for bias, and (d) both magnitude and direction of correlations between RC scales and diagnostic status.

Method

Participants

The sample consisted of 1,411 veteran inpatients from a veteran administration medical center located in the midwestern United States. Participants were either being assessed or provided comprehensive substance abuse treatment between 1994 and 1996. The majority of the sample consisted of African Americans (61%), with the remaining participants being of Caucasian (38%) and other racial composition (1%). On average, participants possessed a high school level of education ($M = 12.5$ years, $SD = 1.7$) and were middle age ($M = 44.1$, $SD = 8.3$).

Demographic information, including age, marital status, occupation type, and length of employment, were assessed for significant difference between racial groups. The majority of respondents were divorced, whereas Caucasian inpatients were more likely to be married than African Americans. The majority of African Americans were employed in semiskilled jobs, whereas most Caucasian respondents worked in skilled manual jobs. In both groups, most respondents were employed full time in the year prior to admission (see Table 1).

All participants were administered the Structured Clinical Interview (SCID) for DSM–III–R, patient version (Spitzer, Williams, Gibbon, & First, 1992), including the psychotic screening and personality disorder sections. Patients were also administered the Substance Use Disorders Diagnostic Schedule (SUDS; Harrison & Hoffman, 1985).

Ninety percent of the sample met diagnostic criteria for alcohol abuse or dependence. Cannabis (68%) and cocaine (67%) were the most frequently abused substances, with opiates, sedatives, and hallucinogens abuse observed in 32%, 27%, and 21% of the sample, respectively. Twenty-seven percent of the sample abused only one substance, 27% abused two substances, 27% abused three or four substances, and the remaining 18% abused at least five substances. Approximately 50% of the sample ($n = 784$) also met criteria for an additional diagnosis (Axis I or Axis II). The primary diagnoses analyzed in this study were schizophrenia, bipolar disorder, major depressive disorder, posttraumatic stress disorder (PTSD), antisocial personality disorder, and borderline personality disorder. The frequency of diagnoses in the final sample are presented in Table 3. Diagnoses were coded as absent, subthreshold (indicating that the patient met a number of diagnostic criteria but not the sufficient number for the diagnosis to be given), or diagnostic criteria met. For regression analyses, each diagnosis was coded as a three-level variable (0 = absent, 1 = subthreshold, 2 = criteria met). See Quirk, Christiansen, Wagner, and McNulty (2003) for additional information regarding sample characteristics.

Table 1
Demographic Information for Both Racial Groups

	African American ($n = 735$)		Caucasian ($n = 449$)			
Variable	*M*	*SD*	*M*	*SD*	χ^2	*p*
Martial status						
Single	218	29.7	109	23.5		
Married	116	15.8	114	24.6		
Divorced/widowed/separated	400	54.5	239	51.7	15.88	<.001
Occupation type						
Manager/administrative	27	3.9	31	6.9		
Clerical and sales	53	7.2	44	9.5		
Skilled manual	162	22.0	143	30.9		
Semiskilled	263	35.8	136	29.4		
Unskilled	193	26.3	93	20.1	23.80	<.001
Length of employment						
Full time	252	35.7	191	42.5		
Part time	205	29.1	101	22.5		
Disabled/retired	127	18.0	101	22.5		
Unemployed/in a controlled environment	121	17.2	56	12.5	14.50	.002

Note. Total $N = 1148–1197$.

Measures

The MMPI-2 (Butcher et al., 1989) is a criterion-keyed inventory designed to measure psychopathological traits and tendencies as well as indicate maladaptive behaviors (e.g., excessive drinking behavior) and affective states (e.g., general malaise). A subset of scales conceptually and, or, empirically linked to diagnoses were chosen for the present analyses. Scales were selected from the Clinical, RC, Content, and Supplementary scales. Clinical scales included 1, 2, 3, 4, 7, 8, 9, and 0, as well as related Harris and Lingoes scales (Harris & Lingoes, 1955, 1968). RC Scales d (Demoralization), 1 (Somatic Complaints), 2 (Low Positive Emotions), 3 (Cynicism), 4 (Antisocial Behavior), 6 (Ideas of Persecution), 7 (Dysfunctional Negative Emotions), 8 (Aberrant Experiences), and 9 (Hypomanic Activity) were also included. Content scales included Anxiety (ANX), Fears (FRS), Depression (DEP), Bizarre Mentation (BIZ), Antisocial Practices (ASP), Anger (ANG), Cynicism (CYN), the MacAndrew Alcoholism Scale-Revised (MAC-R), Addiction Acknowledgement Scale (AAS), and Addiction Potential Scale(APS) (see Butcher, Graham, Williams, & Ben-Porath, 1990, for an overview). Finally, Supplementary scales included PTSD (S) (Schlenger & Kulka, 1989) and PTSD (K) (Kean, Malloy, & Fairbank, 1984). Clinical Scales 5 and 6 and related Harris and Lingoes scales were excluded from regression analyses for lack of appropriate criterion diagnosis. Likewise, Content scales other than the aforementioned were excluded due to a lack of relevant diagnostic criteria.

Diagnostic classifications were made using the SCID. The SCID is a structured diagnostic interview, covering Axis I and II disorders. Each diagnostic section begins with several screening questions. If the patient endorses these initial symptoms, the interviewer asks conditional follow-up questions to determine the number of symptoms present and the level of functional impairment. In accord with previous usage (Dreessen & Arntz, 1998; Dreessen, Hildebrand, & Arntz, 1998), patients were classified as meeting no diagnostic criteria, meeting the subthreshold level or meeting the threshold for full diagnosis. The subthreshold category is assigned to the patient when symptoms cause clinically significant functional impairment but too few symptoms are present to meet criteria for full diagnosis (for more detail on subthreshold diagnoses using the SCID, see Franklin, Sheeran, & Zimmerman, 2002). The SCID displays acceptable test–retest reliability (Dreessen & Arntz, 1998; First, Spitzer, Gibbon, & Williams, 1995; Weiss, Najavits, Muenz, & Hufford, 1995) and, for diagnoses most prevalent in the present sample, exhibits excellent agreement between both joint live observation (Maffei et al., 1997) and joint classification using audio recordings (Fogelson, Nuechterlein, Asarnow, & Subotnik, 1991).

Patients were diagnosed for substance abuse using the SUDS assessment (Harrison & Hoffman, 1985). The SUDS was administered by clinical staff members with at least a bachelor's-level degree. Interviewers were trained to reliability criteria by staff psychologists. The battery of tests was typically administered 1–2 weeks after admission or as soon as patients were medically stable. Participants with extreme scores on the MMPI-2 validity scales (e.g., Variable Response Inconsistency [VRIN], True Response Inconsistency [TRIN]) and "cannot say" blank responses were excluded from analyses. Specifically, participants were not included in the sample if they omitted more than 30 items on the MMPI-2, or if the *F* scale raw score exceeded 30. Participants with VRIN or TRIN *T* scores exceeding 80 were excluded (Graham, 2000). Of the initial sample, 76 individuals were removed from further analyses due to unknown racial backgrounds or racial backgrounds identified as other than African American or Caucasian. Among African American patients, 231 had an invalid MMPI-2 profile (0 due to cannot say "?" frequency; 62 due to elevated *F*; 26 due to elevated VRIN; 34 due to elevated TRIN; and 109 due to a combination of *F*, VRIN, or TRIN) and were excluded from further analysis. In the Caucasian sample, 93 had an invalid MMPI-2 profile (4 due to cannot say "?" frequency; 24 due to elevated *F*; 2 due to elevated VRIN; 17 due to elevated TRIN; and 46 due to a combination of *F*, VRIN, or TRIN) and were also excluded from subsequent analysis (see Table 2 for a list of valid scale means). A comparison of predictor scale means between those excluded due to invalid profiles and those determined to be valid showed the invalid group to be significantly elevated on almost all scales. Lower mean-level scale scores for the invalid group included APS, Pd2 (Authority Problems), Pd3 (Social Imperturbability), Pa3 (Naiveté), and Ma3 (Imperturbability). All differences were significant at the $p = .05$ level except for Masculinity-Feminity (MF), D2 (Psychomotor Retardation), and Pd2. Differences in diagnosis are not reported, as the low number of individuals meeting diagnostic criteria among the invalid group makes interpretation difficult. Demographic differences between the valid and invalid group were marginal, as was the difference in number of African American versus Caucasian patients displaying invalid profiles.

Data Analysis

Moderated multiple regression is the most typical analysis used when studying slope and intercept differences between samples (Nunnally & Bernstein, 1994). In the present study, a variant of this approach was used. A step-down hierarchical regression procedure (Lautenschlager & Mendoza, 1986) has been used previously to identify racial bias in clinical predictions (Arbisi et al., 2002) and employment settings (Rotundo & Sackett, 1999). The presence of racial bias is tested by comparing a regression model that includes only the predictor variable against a model that includes the predictor variable, hypothesized moderator, and cross-product term of the predictor and moderator variable (full model). Significant incremental variance obtained using the full model beyond the predictor-only model indicates the presence of bias. Subsequent analyses are then conducted to identify whether the bias resulted from differences in slope, intercept, or both. Significant incremental variance associated with comparing the full model beyond a model containing the predictor and race alone is indicative of slope bias. A subsequent test of intercept bias is then conducted by determining whether the full model explains incremental variance beyond a model containing only the predictor and cross-product term.

After significance tests were conducted, an exploration of the nature of the bias was conducted by plotting the predictor scores against the diagnostic criteria using separate regression lines for each subgroup, as well as a common regression line. In the

Table 2
Comparison of Minnesota Multiphasic Personality Inventory–2 Scale Scores Between African American and Caucasian Respondents

	African American (n = 735)		Caucasian (n = 462)				
Scale	*M*	*SD*	*M*	*SD*	*d*	*t*(1195)	*p*
Basic profile							
L	50.46	09.71	51.51	10.60	0.10	1.75	.08
F	69.67	20.66	70.81	21.31	0.06	0.92	.36
K	42.82	09.73	43.83	10.48	0.10	1.70	.09
1	67.91	17.83	70.22	19.60	0.13	2.10	.04
2	67.44	15.67	72.25	16.86	0.30	5.03	.00
3	58.25	13.54	62.60	13.22	0.33	5.46	.00
4	71.51	12.80	71.94	13.51	0.03	1.14	.25
5	50.51	08.82	48.52	08.89	−0.22	3.80	.00
6	63.87	17.07	65.95	16.13	0.13	2.10	.04
7	66.45	15.73	68.12	17.18	0.10	1.72	.09
8	69.77	19.20	70.01	20.71	0.01	0.21	.84
9	62.38	11.31	57.37	11.28	−0.44	7.46	.00
0	57.78	11.88	59.81	13.95	0.16	2.69	.07
RC scales							
RCd	64.47	13.86	67.72	14.75	0.23	3.85	.00
RC1	64.92	15.90	66.35	17.40	0.09	1.46	.14
RC2	56.21	13.50	61.95	15.20	0.40	6.82	.00
RC3	62.13	11.33	57.80	12.76	−0.36	6.12	.00
RC4	67.63	10.62	66.14	12.55	−0.13	2.21	.03
RC6	62.84	19.99	55.50	20.03	−0.37	6.18	.00
RC7	62.40	14.19	61.72	15.65	−0.05	0.78	.44
RC8	63.36	15.20	59.79	14.73	−0.24	4.01	.00
RC9	57.02	11.58	53.55	11.53	−0.30	5.06	.00
H-L scales							
D1	66.72	16.52	71.94	18.27	0.30	5.11	.00
D2	55.73	11.42	58.79	11.65	0.27	4.48	.00
D3	64.76	14.87	66.32	15.51	0.10	1.73	.08
D4	67.38	17.93	71.47	20.06	0.21	3.67	.00
D5	64.72	14.83	68.05	16.41	0.21	3.63	.00
Hy1	45.86	09.85	44.86	11.08	−0.10	1.63	.10
Hy2	42.84	08.96	44.25	10.62	0.14	2.48	.01
Hy3	70.71	18.54	74.94	19.31	0.22	3.79	.00
Hy4	65.30	17.58	67.39	19.83	0.11	1.90	.06
Hy5	42.84	09.41	45.01	09.52	0.23	3.86	.00
Pd1	63.14	13.48	64.18	13.56	0.08	1.29	.20
Pd2	56.83	09.16	56.22	10.05	−0.06	1.08	.28
Pd3	46.07	09.87	45.27	10.75	−0.08	1.31	.19
Pd4	66.26	13.32	66.85	13.37	0.04	0.75	.46
Pd5	71.50	12.66	72.65	12.90	0.09	1.52	.13
Pa1	72.20	19.50	67.95	18.66	−0.22	3.73	.00
Pa2	61.74	14.16	63.90	14.16	0.15	2.56	.01
Pa3	40.73	08.07	44.76	09.98	0.44	7.66	.00
Sc1	65.80	17.37	64.17	18.01	−0.09	1.56	.20
Sc2	65.58	20.19	69.06	21.39	0.17	2.83	.01
Sc3	66.64	18.55	66.79	19.46	0.01	0.14	.89
Sc4	64.44	18.45	68.67	20.26	0.22	3.72	.00
Sc5	63.31	16.05	62.67	15.90	−0.04	0.68	.50
Sc6	69.72	20.16	69.23	21.26	−0.02	0.40	.69
Ma1	59.31	11.41	55.24	11.38	−0.36	6.02	.00
Ma2	54.90	10.23	53.35	10.13	−0.15	2.57	.01
Ma3	47.77	09.86	46.31	10.22	−0.15	2.46	.01
Ma4	60.53	11.57	55.10	11.96	−0.46	7.81	.00
Si1	53.58	10.27	55.75	12.23	0.19	3.29	.00
Si2	53.68	11.52	55.12	11.90	0.12	2.08	.04
Si3	62.22	11.92	61.55	13.35	−0.05	0.91	.36

(*table continues*)

Table 2 (*continued*)

Scale	African American (n = 735)		Caucasian (n = 462)				
	M	*SD*	*M*	*SD*	*d*	*t*(1195)	*p*
Content scales							
ANX	65.33	14.17	67.36	15.08	0.14	2.36	.02
FRS	59.90	12.12	55.51	11.66	−0.37	6.19	.00
OBS	60.10	13.22	60.08	14.40	0.00	0.02	.98
DEP	71.51	17.36	74.67	19.07	0.18	2.95	.00
HEA	66.49	17.33	69.23	19.83	0.15	2.52	.01
BIZ	65.23	18.75	59.62	17.23	−0.30	5.20	.00
ANG	59.74	12.19	59.49	13.00	−0.02	0.34	.74
CYN	62.65	08.82	58.19	10.53	−0.47	7.90	.00
ASP	63.15	09.54	59.28	11.02	−0.38	6.43	.00
TPA	55.98	10.85	54.80	11.63	−0.10	1.79	.08
LSE	59.45	14.59	63.47	16.23	0.27	4.44	.00
SOD	56.24	12.09	58.40	14.06	0.17	2.81	.01
FAM	63.52	14.20	62.84	15.18	−0.04	0.79	.43
WRK	63.55	15.08	65.98	17.09	0.15	2.60	.01
TRT	65.15	15.77	67.00	17.44	0.11	1.90	.06
PTSDS	70.99	18.15	71.76	19.52	0.04	0.69	.49
PTSDK	71.86	18.34	72.40	18.93	0.03	0.49	.62
MAC–R	66.59	09.72	64.20	10.22	−0.24	4.05	.00
APS	57.07	10.25	58.19	10.91	0.11	1.79	.07
AAS	72.25	11.22	69.47	12.98	−0.23	3.92	.00

Note. d = Cohen's d statistic. p value is associated with a ± test of differences between means. L = Lie; F = Infrequency; K = Correction; 1 = Hypochondriasis; 2 = Depression; 3 = Hysteria; 4 = Psychopathic Deviate; 5 = Masculinity–Femininity; 6 = Paranoia; 7 = Psychasthenia; 8 = Schizophrenia; 9 = Mania; 0 = Social Introversion; RC = Restructured Clinical; RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; D1 = Subjective Depression; D2 = Psychomotor Retardation; D3 = Physical Malfunctioning; D4 = Mental Dullness; D5 = Brooding; Hy1 = Denial of Social Anxiety; Hy2 = Need for Affection; Hy3 = Lassitude-Malaise; Hy4 = Somatic Complaints; Hy5 = Inhibition of Aggression; Pd1 = Familial Discord; Pd2 = Authority Problems; Pd3 = Social Imperturbability; Pd4 = Social Alienation; Pd5 = Self- Alienation; Pa1 = Persecutory Ideas; Pa2 = Poignancy; Pa3 = Naiveté; Sc1 = Social Alienation; Sc2 = Emotional Alienation; Sc3 = Lack of Ego Mastery, Cognitive; Sc4 = Lack of Ego Mastery, Conative; Sc5 = Lack of Ego Mastery, Defective Inhibition; Sc6 = Bizarre Sensory Experiences; Ma1 = Amorality; Ma2 = Psychomotor Acceleration; Ma3 = Imperturbability; Ma4 = Ego Inflation Social Introversion Subscales; Si1 = Shyness/Self-Consciousness; Si2 = Social Avoidance; Si3 = Alienation–Self and Others; ANX = Anxiety; FRS = Fears; OBS = Obsessiveness; DEP = Depression; HEA = Health Concerns; BIZ = Bizarre Mentation; ANG = Anger; CYN = Cynicism; ASP = Antisocial Practices; TPA = Type-A Behavior; LSE = Low Self-Esteem; SOD = Social Discomfort; FAM = Family Problems; WRK = Work Interference; TRT = Negative Treatment Indicators; MAC–R = MacAndrew Alcoholism Scale—Revised; AAS = Addiction Acknowledgement Scale; APS = Addiction Potential Scale.

presence of slope bias, the steepness of the regression lines indicate for which subgroup the test is more predictive. In the presence of intercept bias, the subgroup fit line that is relatively lower than the comparison subgroup indicates overprediction, such that the former subgroup diagnosis is overpredicted when using a common regression line (Cleary, 1968).

The nature of over- and underprediction can become unclear when significant intercept bias is present in a crossover (nonordinal) interaction. In this situation, the test appears to overpredict for one subgroup up to the point of intersection of the two regression lines and then appears to underpredict for that same group thereafter. In the present study, when crossover interactions displayed significant intercept bias, Potthoff's (1964) extension of the Johnson and Neyman (1936) procedure for establishing simultaneous regions of significance was conducted. In the present study, step-down moderated multiple regression can be thought of as the omnibus test, whereas Potthoff's (1964) analysis offers a closer look at where the regression lines differ. For those more familiar with factorial interaction designs, this procedure can be thought of as analogous to a test of simple effects once a significant interaction is found. This procedure provides an upper and lower boundary around the point of intersection that indicates a region where predicted between-group scores are not significantly, $F(2, N\text{-}4)$, $p = .05$, different, thereby creating confidence limits around the test score that resides at the point of intersection. In other words, this provides an estimate of the region of nonsignificance. Alternately, test scores that fall below the lower boundary or above the upper boundary are considered to be within regions of significance, wherein predicted scores between groups significantly differ (see Rogosa, 1980, 1981). This estimate of the region of nonsignificance (and regions of significance) is intended to identify, in the case of nonordinal interaction, where the test stops significantly overpredicting for one subgroup and starts overpredicting for the other.

Scale Choice

The choice of scales for inclusion in regression analysis was based on both theoretical and empirical consideration. From an empirical standpoint, scales were included as predictors if results of numerous studies have shown evidence of diagnostic prediction. Additionally, each scale showing a significant correlation in previous research greater than .10 with a diagnosis

was considered for inclusion in the analyses in the present study. Research on the RC, Content, and Supplemental scales is relatively limited compared with the depth of research on the clinical scales; therefore, the authors relied on theoretical relationships between scales and diagnostic criteria in addition to available empirical evidence.

Results

Mean Differences Between Groups: MMPI-2 and Diagnoses

Given the large sample size and correspondingly narrow estimates of sampling error, reliance on statistical significance is less informative than estimates of effect size and related indicators of differences in magnitude such as *T* score differences of 5 points (Kline, 2004). Regarding clinical scales, African Americans displayed meaningfully higher group mean scores on Scale 9 (Mania, Cohen's $d = -0.44$, see Table 2). African Americans generated somewhat lower scores on Scale 2 (Depression) and Scale 3 (Hysteria), with *T*-score differences greater than 4 points and standardized mean differences of .30 or greater.

The RC scales yielded significant group differences on seven of the nine scales (see Table 2). Of the seven significant differences, the African American group displayed higher mean scores than Caucasians on five (RC3, RC4, RC6, RC8, RC9). The higher mean-level score of RC9 ($T > 5$) is the only difference to reach a magnitude suggestive of clinical significance.

African Americans generated higher scores on the scales of CYN, FRS, and ASPs and endorsed fewer concerns of low self-esteem (LSEs) with *T*-score differences greater than 4 points and standardized mean differences greater than .25.

Of the African American (*n*s = 416–510) and Caucasian (*n*s = 212–274) respondents who had both a valid MMPI-2 profile and diagnostic data available, Caucasian respondents were more likely to have been diagnosed with borderline personality disorder, major depression, bipolar disorder, PTSD, and gambling than their African American counterparts (see Table 3).

Diagnostic Prediction by Group

Results of the step-down hierarchical moderated regressions are shown in Tables 4, 5, 6, and 7, which include prediction of relevant diagnosis by Clinical, Content and Supplementary, Harris-Lingoes (including Social Introversion), and RC scales, respectively. As was mentioned earlier in this article, slope bias indicates a difference in accuracy of criterion prediction between groups. Each table contains the bivariate correlation coefficient for the entire sample and separate coefficients for each group helping to clarify for which group the predictor scale was more strongly (and in which direction) related to the diagnostic criteria. In the case of intercept bias, a significant finding is indicative of a difference between groups in diagnostic classification across a range of test scores. In the case of significant intercept bias, overprediction is present for a subgroup if that subgroup's regression line is lower relative to the other subgroup within the meaningful range of scores (Cleary, 1968). The subgroup with a regression line above the other subgroup, within the meaningful range of scores and in the case of statistical significance, is said to be underpredicted. As was mentioned previously, over- and underprediction becomes somewhat unclear in the presence of a crossover interaction. In the present study, nonordinal interactions that displayed intercept bias were analyzed further, wherein a calculation was conducted to determine the region of nonsignificance around the point of intersection (Potthoff, 1964). For each nonordinal interaction displaying significant intercept bias, the upper and lower boundary for the aforementioned region of nonsignicance is described within the text of this article, and the effect is noted in the each table.

With regard to the information presented in Tables 4–7, the full regression model includes the MMPI-2 scale, the racial background variable, and the interaction term. The prediction bias column reflects comparisons between the MMPI-2 scale and the

Table 3
Comparison of Diagnosis Between African American and Caucasian Respondents

	Participants (*n*s = 638–748)						
	African American		Caucasian				
Diagnosis	Subthreshold	Met criteria	Subthreshold	Met criteria	*d*	*t*	*p*
PTSD	23	33	34	47	0.41	5.63[a]	.000
Major depression	14	36	16	45	0.34	4.54[b]	.000
Schizophrenia	4	32	3	19	0.07	0.884[c]	.377
Bipolar disorder	1	9	3	21	0.34	4.55[d]	.001
Antisocial PD	22	82	10	35	−0.04	0.49[e]	.622
Borderline PD	15	23	15	28	0.28	3.70[f]	.000
Narcissistic PD	1	31	9	12	0.00	0.03[g]	.978
Avoidant PD	8	9	9	10	0.19	2.41[h]	.016
Gambling	7	13	4	48	0.53	7.33[i]	.000

Note. Different *n*s reflect diagnostic comorbidity. Diagnostic status was coded 0 for no diagnosis beyond substance abuse, 1 for a subthreshold diagnosis, and 2 for full diagnostic criteria met. African Americans with no diagnosis beyond substance abuse $N = 406$; Caucasians with no diagnosis beyond substance abuse $N = 193$. *d* = Cohen's *d* statistic. *p* value is associated with a *t* test of differences between means. PTSD = posttraumatic stress disorder; PD = personality disorder.
[a] $df = 734$. [b] $df = 708$. [c] $df = 655$. [d] $df = 631$. [e] $df = 746$. [f] $df = 678$. [g] $df = 656$. [h] $df = 633$. [i] $df = 669$.

Table 4
Regression Analysis of Diagnoses on Clinical Scales, Race, and the Scale × Race Interaction Term (Men Only, ns = 639–755)

Diagnosis and clinical scale	r_{xy}	(A. A.|C.)	Full model b IV	Full model b Race	Full model b IV × Race	Full model R^2	Prediction bias R^2D	Slope bias R^2D	Intercept bias R^2D
PTSD diagnosis									
Scale 1	.316***	(.288|.314)	.015	.132	.009	.129***	.032***	.002	.030***††
Scale 3	.257***	(.222|.222)	.011	.004	.009	.089***	.028***	.002	.026***††
Scale 7	.343***	(.299|.374)	.006	.041	.010	.153***	.039***	.006*	.000
Scale 8	.361***	(.333|.387)	.008	.121	.006	.168***	.042***	.004	.038***††
Depression diagnosis									
Scale 2	.373***	(.292|.426)	.005	−.324	.018	.159***	.024***	.010**	.004
Scale 7	.365***	(.315|.414)	.006	−.037	.011	.161***	.030***	.008**	.000
Scale 8	.371***	(.335|.420)	.006	.037	.008	.170***	.034***	.007*	.000
Schizophrenia diagnosis									
Scale 7	.176***	(.217|.106)	.016	.150	−.005	.035***	.004		
Scale 8	.243***	(.270|.201)	.013	.109	−.002	.062***	.003		
Bipolar diagnosis									
Scale 2	.224***	(.179|.263)	−.006	−.209	.014	.087***	.037***	.012**	.004
Scale 4	.146***	(.050|.245)	−.019	−.386	.022	.072***	.051***	.020***††	.010*††
Scale 7	.230***	(.163|.309)	−.007	−.085	.012	.103***	.050***	.020***	.002
Scale 8	.231***	(.153|.328)	−.007	−.055	.010	.110***	.057***	.024***	.001
Scale 9	.080*	(.011|.223)	−.026	−.354	.026	.065***	.058***	.021***	.009*†
Antisocial PD diagnosis									
Scale 4	.282***	(.277|.292)	.036	.000	−.001	.080***	.000		
Scale 9	.259***	(.259|.263)	.037	.042	.001	.068***	.001		
Avoidant PD diagnosis									
Scale 7	.146***	(.115|.186)	−.001	−.012	.004	.033***	.013**	.003	.009**††
Scale 8	.165***	(.103|.249)	−.003	−.043	.006	.046***	.020***	.010*	.001
Scale 0	.235***	(.140|.343)	−.005	−.198	.009	.078***	.025***	.016***	.008*†
Borderline PD diagnosis									
Scale 4	.360***	(.280|.452)	−.004	−.526	.026	.160***	.034***	.019***	.011**†
Scale 7	.380***	(.290|.485)	−.002	−.182	.015	.179***	.035***	.021***	.005*†
Scale 8	.423***	(.333|.529)	.000	−.114	.012	.214***	.038***	.021***	.003
Narcissistic PD diagnosis									
Scale 8	.044	(.038|.017)	.002	.020	−.001	.001	.000		
Scale 9	.081*	(.100|.037)	.016	.144	−.006	.007	.001		
Gambling diagnosis									
Scale 4	.168***	(.196|.174)	.002	.068	.010	.104***	.076***	.003	.074***††

Note. Tests of slope and intercept bias were conducted only if evidence of predictive bias was indicated. The full model displays unstandardized regression coefficients, shown as b. Race was coded 1 for African American participants and 2 for Caucasian participants. Full model = scale, race, and the interaction term (IV × Race); r_{xy} = correlation between the predictor scale and diagnosis only; A. A. = correlation between predictor scale and diagnosis only for African American patients; C. = correlation between predictor scale and diagnosis only for Caucasian patients; IV = scale; R^2D = the change in the variance accounted for by the addition of the model; Prediction bias R^2D = comparison of the scale alone with the full model; Slope bias R^2D = comparison of the scale plus racial background with the full model. If slope bias is significant, then the Intercept bias R^2D = comparison of the scale plus interaction term with the full model. If there is no evidence for slope bias, then the Intercept bias R^2D = comparison of the scale alone with the scale plus racial background variable.

† = overprediction for Caucasians across test scores below the point at which regression lines intersect, while overpredicting for African Americans scores above the point of intersection. †† = overprediction for African American across the entire range of test scores.

* $p < .05$. ** $p < .01$. *** $p < .001$.

full model. The slope bias column indicates comparisons of the MMPI-2 scale plus the racial background variable against the full model. If there was evidence of slope bias, then the intercept bias was computed by comparing the MMPI-2 scale plus the interaction term against the full model. If there was no evidence for slope bias, then the intercept bias column was computed by comparing the MMPI-2 scale alone with the MMPI-2 scale plus racial background variable. Tests of slope and intercept bias were conducted only if evidence of predictive bias was indicated.

Correlations between the MMPI-2 Clinical scales and respective diagnosis ranged from .08 to .373. In terms of full regression models, the percentage of variance accounted for in diagnosis by clinical scales (showing significance) also varied greatly, ranging from 3% to 21%, with the bulk of these analyses yielding moderate to large effect sizes and an unweighted average significant R^2 of .11 (Cohen, 1988). The first test in step-down hierarchical regression, assessment of prediction bias, yielded significant results for all predictors except for Scales 7 and 8 predicting a schizophrenia diagnosis, Scales 4 and 9 predicting antisocial personality, and Scales 8 and 9 predicting narcissistic personality. Bias was evident for the prediction of PTSD with Scales 1, 3, 7, and 8. Biased prediction for depression diagnoses was present with Scales 2, 7, and 8. Clinical Scales 2, 4, 7, 8, and 9 were racially biased predictors of bipolar disorder diagnosis. Scales 7, 8, and 0 were

Table 5
Regression Analysis of Diagnoses on Both Content and Supplementary Scales, Race, and the Scale × Race Interaction Term (Men Only, ns = 639–755)

Diagnosis and clinical scale	r_{xy}	(A. A.|C.)	Full model b IV	Full model b Race	Full model b IV × Race	R^2	Prediction bias R^2D	Slope bias R^2D	Intercept bias R^2D
PTSD diagnosis									
ANX	.341***	(.300|.363)	.011	.050	.016	.149***	.038***	.006*	.000
FRS	.062	(.065|.147)	−.013	.163	.023	.055***	.050***	.003	.047***††
SOD	.326***	(.273|.358)	.013	.097	.014	.141***	.038***	.004	.034***††
PTSD (S)	.381***	(.356|.410)	.007	.091	.007	.184***	.042***	.006*	.001
PTSD (K)	.377***	(.353|.409)	.008	.075	.009	.183***	.044***	.006*	.001
Depression diagnosis									
DEP	.395***	(.332|.450)	.008	−.035	.016	.182***	.029***	.010**	.000
ANX	.341***	(.297|.377)	.011	−.004	.018	.141***	.028***	.006*	.000
Schizophrenia diagnosis									
BIZ	.342***	(.398|.260)	.055	.141	−.010	.126***	.008*	.001	.007*
Antisocial PD diagnosis									
ASP	.257***	(.225|.326)	.032	−.052	.009	.068***	.002		
ANG	.280***	(.266|.309)	.047	−.028	.002	.078***	.000		
CYN	.136***	(.098|.207)	.004	−.133	.011	.020**	.002		
MAC-R	.226***	(.236|.201)	.049	.192	−.007	.051***	.000		
AAS	.203***	(.213|.180)	.088	.133	−.019	.042***	.001		
APS	.083*	(.105|.043)	.034	.313	−.013	.009	.002		
Borderline PD diagnosis									
ASP	.199***	(.147|.327)	−.013	−.149	.030	.084***	.045***	.013**	.002
ANG	.359***	(.265|.470)	−.008	−.173	.039	.165***	.039***	.021***	.005
Avoidant PD diagnosis									
SOD	.248***	(.154|.357)	−.008	−.101	.017	.085***	.026***	.017***	.004

Note. Tests of slope and intercept bias were conducted only if evidence of predictive bias was indicated. The full model displays unstandardized regression coefficients, shown as b. Race was coded 1 for African American participants and 2 for Caucasian participants. Full model = scale, race, and the interaction term (IV × Race); r_{xy} = correlation between the predictor scale and diagnosis only; A. A. = correlation between predictor scale and diagnosis only for African American patients; C. = correlation between predictor scale and diagnosis only for Caucasian patients; IV = scale; R^2D = the change in the variance accounted for by the addition of the model; Prediction bias R^2D = comparison of the scale plus racial background with the full model. If slope bias is significant, then the Intercept bias R^2D = comparison of the scale plus interaction term with the full model. If there is no evidence for slope bias, then the intercept bias R^2D = comparison of the scale alone with the scale plus racial background variable. PTSD = posttraumatic stress disorder; ANX = Anxiety; FRS = Fears; SOD = Social Discomfort; DEP = Depression; BIZ = Bizarre Mentation; PD = personality disorder; ASP = Antisocial Practices; ANG = Anger; CYN = Cynicism; MAC–R = MacAndrew Alcoholism Scale—Revised; AAS = Addiction Acknowledgement Scale; APS = Addiction Potential Scale.
†† = overprediction for African American across the entire range of test scores.
* $p < .05$. ** $p < .01$. *** $p < .001$.

biased predictors of avoidant personality. Finally, Scale 4, when used to predict gambling, produced the largest intercept bias ($\Delta R^2 = .075$; see Table 4).

Subsequent tests of slope bias displayed significant results for the aforementioned relationships for depression predicted by Scales 2, 7, and 8; all relationships predicting bipolar, avoidant diagnosis when predicted by Scales 8 and 0; borderline diagnosis predicted by Scales 4, 7, and 8; and gambling predicted by Scale 4. The largest effect sizes for slope bias were observed when Scale 8 was used to predict bipolar diagnosis ($\Delta R^2 = .024$). Scales 1, 3, and 8 were used to predict PTSD-produced sizable intercept bias ($\Delta R^2 = .030, .026, .038$, respectively). Scales 4 and 9 produced significant intercept bias when used to predict bipolar diagnosis. Finally, Scale 4 displayed significant interaction bias when predicting gambling.

Scale 4 overpredicts both bipolar and borderline diagnoses for Caucasians when scores are in the lower region of significance, whereas it overpredicts for African Americans in the upper region of significance, and vice versa. With regard to Scale 4 and borderline diagnosis, the lower and upper boundaries of the region of nonsignificance are −40.25 and 4.77, $F(2, 676) = 3.71$, $p = .05$, respectively. When Scale 4 is used to predict bipolar diagnosis, the lower limit of the area of nonsignificance is −36.52 and the upper limit is 7.89, $F(2, 635) = 3.71$, $p = .05$. The regression of bipolar diagnosis on Scale 9 produced a nonordinal interaction where the raw score lower and upper boundaries of nonsignificance are 4.94 and 16.88, $F(2, 629) = 3.71$, $p = .05$, respectively. Scores in the lower region of significance overpredicted for Caucasians, whereas scores in the upper region overpredicted for African Americans. Using Scale 7 to predict borderline personality disorder also displayed this trend, wherein lower raw scores overpredict for Caucasians and high scores overpredict for African Americans; however, the lower boundary of nonsignificance was 0.97, and the upper limit was 17.76, $F(2, 676) = 3.71$, $p = .05$. Scale 0 displayed intercept bias predicting avoidant diagnosis, which was also nonordinal, with a lower and upper nonsignificance boundary of 6.58 and 28.67, $F(2, 631) = 3.71$, $p = .05$, respectively. Scale 0 overpredicts for Caucasians and underpredicts for African Amer-

icans in the lower region of significance, whereas the test underpredicts for Caucasians and overpredicts for African Americans in the upper region of significance.

Selected Content and Supplementary scales displayed correlation coefficients with relevant diagnoses ranging from .062 to .395, and nearly all full regression models were significant predictors of associated diagnoses with small to moderate effect sizes (R^2 = .020–.184). Prediction bias was present for relationships between all predictors and diagnoses except those involving antisocial personality. Slope bias was present for all of the aforementioned significant relationships except for FRS and SOD (Social Discomfort) with PTSD and BIZ with schizophrenia. Intercept bias was present for PTSD predicted by FRS and SOD as well as schizophrenia predicted by BIZ. Intercept bias evident when regressing schizophrenia diagnosis with the BIZ scale was nonordinal (crossover interaction). A region of nonsignificance was estimated using Potthoff's (1964) analysis. This resulted in a lower boundary of −50.97 and an upper boundary of 7.68, $F(2, 631) = 3.71$, $p = .05$, for the region of nonsignificance. The test overpredicted for African Americans in the lower range of BIZ scores, whereas it overpredicted for Caucasian inpatients in the higher score range. Finally, Supplementary scales PTSD (S) and PTSD (K) were significant predictors of PTSD diagnosis (R^2 = .13). Both PTSD (S) and PTSD (K) yielded a significant prediction and slope bias, but not intercept bias.

Next, test bias associated with related Harris and Lingoes scales was assessed. Correlations between scales and diagnoses ranged from −.181 to .419 (subgroups combined), with full models demonstrating R^2 that ranged from 0 to .199. Of the 92 relationships assessed, 68 displayed prediction bias. Of the significant prediction biases, 47 displayed slope bias, 7 of which also displayed intercept bias, and 20 models evidenced intercept bias only. Slope and intercept bias ranged from 0 to .029 and from 0 to .076, respectively. Several of the relationships that displayed intercept bias were nonordinal interactions. The D2 scale on Depression overpredicted for Caucasians at lower scores yet overpredicted for African Americans at higher scores, where the lower point of the nonsignificance region was −573.82 and the higher point was 5.18, $F(2, 706) = 3.71$, $p = .05$, overpredicting only for African Americans within a region of significance that is within possible scores. The regression of bipolar diagnosis on Ma2 resulted in overprediction for Caucaisians at lower scores and the opposite at higher scores, resulting in a nonsignificance region from 0.01 to 4.57, $F(2, 629) = 3.71$, $p = .05$. The trend continued for Si3 (Alienation–Self and Others) in prediction of avoidant diagnosis, with a lower nonsignificance score of 6.45 and an upper limit score of 236.52, $F(2, 631) = 3.71$, $p = .05$. The same over- and underprediction trend resulted for the nonordinal interactions between Pd3, Pd4, and Pd5 on borderline diagnosis. The lower and upper boundaries for nonsignificance were 3.90 and 8.51 for Pd3, $F(2, 672) = 3.71$, $p = .05$, 0.28 and 5.92 for Pd4, $F(2, 676) = 3.71$, $p = .05$, and 1.75 and 6.81 for Pd5, $F(2, 676) = 3.71$, $p = .05$.

Results from the step-down analyses using the RC scales with related diagnoses yielded comparable biases between groups. Significant simple bivariate correlation coefficients ranged from .035 to .390. The largest amount of variance explained by the full model in diagnosis was found for PTSD by RC7 (R^2 = .370). The remaining RC scales did not yield an R^2 greater than .179 (RCd and borderline diagnosis) on relevant diagnoses. Of the 39 relationships assessed, 24 displayed significant prediction bias, with 19 models indicating slope bias and 6 models indicating intercept bias. Scales showing slope bias included RCd and RC7 for PTSD; RCd, RC2, RC7 for depression; RCd, RC2, RC4, RC7, RC8, and RC9 for bipolar diagnosis; RCd and RC3 with borderline personality; RCd with avoidant personality disorder; and RCd and RC4 for gambling. Two models displaying both significant slope and intercept bias included RC9 with bipolar diagnosis and RC4 with gambling diagnosis. Models showing only intercept bias included the regression of RC2 and RC8 on PTSD, RC8 on depression, and RC7 on avoidant diagnosis. The largest predictive bias involved gambling diagnosis. Both RCd and RC4 demonstrated biased predictions of bipolar diagnosis (ΔR^2 = .080 and ΔR^2 = .082, respectively), the latter of which displayed both significant slope and intercept bias. The largest intercept bias was seen on the prediction of PTSD by RC8 (ΔR^2 = .047). The only crossover interaction displaying significant intercept bias involved RC9 on bipolar diagnosis, which displayed a lower limit for the area of nonsignificance of 1.69 and an upper limit of 10.98, $F(2, 633) = 3.71$, $p = .05$.

Discussion

The results of the present study reveal a pattern of predictive bias across numerous scales, albeit generally modest in effect size. These biases indicate both over- and underprediction of psychiatric disorders among African Americans on a variety of scales and suggest differential accuracy for the MMPI-2 in predicting diagnostic status between subgroups of male veteran inpatients seeking substance abuse treatment. A unique finding is that of the 46 relationships displaying significant intercept bias, none overpredicted diagnosis for Caucasians across the entire range of test scores. In all but 1 of the relationships showing intercept bias (BIZ with schizophrenia diagnosis), the test overpredicted either for African Americans across the entire range of test scores or only at higher test scores. In these relationships in which the test overpredicted for African Americans only at a higher range of test scores, the test overpredicted for Caucasians at lower test scores. Additionally, in some of these relationships, the region of significance wherein the test overpredicted for Caucasians was actually outside the range of possible test scores, making this overprediction unimportant for practical purposes.

The present results are different from those of the most similarly designed study by Arbisi et al. (2002). Arbisi and colleagues found evidence of modest underprediction for African American men in roughly half of the scales assessed. Differences in results may be attributable to the type of population, setting, and/or extratest criteria used. Arbisi et al. used a general psychiatric population from two hospitals and examined both men and women with a chart-based review of psychiatric symptoms. In the present study, we examined a male veteran, substance abusing population with structured clinical diagnoses as the extratest criterion. Whereas, the precise source(s) of the methodological or sample differences that account for the different pattern in results cannot be determined from these two studies, the results do support the importance of examining predictive bias between racial groups across various settings and populations in which the MMPI-2 is routinely used to make important clinical decisions.

Table 6
Regression Analysis of Diagnosis on Harris-Lingoes (and Social Introversion) Scales, Race, and the Scale × Race Interaction Term (Men Only, ns = 639–745)

			Full model						
			b						
Diagnosis H & L scale	r_{xy}	(A. A.|C.)	IV	Race	IV × Race	R^2	Prediction bias R^2D	Slope bias R^2D	Intercept bias R^2D
PTSD diagnosis									
Hy1	−.149***	(−.110|−.174)	.000	.372	−.033	.062***	.042***	.002	.040***††
Hy2	−.181***	(−.134|−.260)	.007	.478	−.037	.084***	.051***	.006*	.028***††
Hy3	.368***	(.313|.393)	.012	.021	.029	.165***	.037***	.008*	.000
Hy4	.269***	(.248|.256)	.029	.186	.010	.102***	.033***	.001	.032***††
Hy5	−.132***	(−.164|−.160)	−.040	.359	−.026	.066***	.048***	.001	.047***††
Sc1	.325***	(.279|.397)	.008	.115	.030	.157***	.055***	.009**	.003
Sc2	.335***	(.257|.380)	.017	.092	.053	.145***	.037***	.008*	.002
Sc3	.334***	(.331|.357)	.033	.175	.026	.156***	.044***	.004	.041***††
Sc4	.353***	(.301|.373)	.024	.105	.025	.154***	.033***	.005*	.002
Sc5	.260***	(.236|.305)	.016	.139	.043	.114***	.045***	.005*	.004
Sc6	.307***	(.333|.280)	.042	.236	.004	.131***	.037***	.000	.037***††
Depression diagnosis									
D1	.399***	(.314|.458)	.007	−.136	.022	.180***	.025***	.011**	.002
D2	.272***	(.223|.293)	.016	−.122	.046	.095***	.025***	.005	.020***†
D3	.205***	(.174|.216)	.026	.074	.029	.066***	.026***	.002	.024***††
D4	.419***	(.356|.463)	.024	−.004	.030	.196***	.024***	.008**	.000
D5	.375***	(.295|.440)	.011	−.053	.052	.167***	.029***	.012**	.000
Sc1	.321***	(.290|.401)	.008	.074	.033	.148***	.045***	.010**	.001
Sc2	.383***	(.340|.408)	.046	.069	.047	.166***	.026***	.006*	.001
Sc3	.360***	(.325|.462)	.019	.079	.040	.163***	.037***	.009**	.001
Sc4	.389***	(.345|.416)	.028	.046	.029	.171***	.024***	.006*	.000
Sc5	.235***	(.197|.298)	−.001	.075	.052	.091***	.036***	.007*	.001
Sc6	.273***	(.244|.305)	.016	.111	.020	.103***	.028***	.004	.025***††
Schizophrenia diagnosis									
Sc1	.253***	(.268|.245)	.036	.078	−.001	.068***	.004		
Sc2	.143***	(.131|.160)	.026	.015	.009	.021***	.001		
Sc3	.221***	(.228|.214)	.039	.049	.001	.051***	.002		
Sc4	.153***	(.130|.187)	.011	−.010	.010	.025***	.002		
Sc5	.235***	(.283|.154)	.092	.130	−.026	.060***	.005		
Sc6	.239***	(.309|.119)	.066	.170	−.025	.066***	.009*	.008*	.009*††
Bipolar diagnosis									
D1	.250***	(.190|.302)	−.009	−.082	.018	.103***	.041***	.016***	.002
D2	.149***	(.124|.162)	−.011	−.046	.030	.054***	.032***	.004	.028***††
D3	.146***	(.153|.143)	.007	.073	.018	.052***	.031***	.001	.030***††
D4	.245***	(.162|.324)	−.022	−.045	.035	.109***	.049***	.022***	.001
D5	.241***	(.169|.308)	−.025	−.032	.044	.103***	.045***	.018***	.000
Ma1	.033	(.040|.096)	−.023	.087	.031	.038***	.037***	.003	.035***††
Ma2	.085*	(−.017|.212)	−.067	−.211	.065	.062***	.054***	.021***	.006*†
Ma3	−.082*	(−.061|−.113)	.018	.254	−.029	.041***	.035***	.003	.032***††
Ma4	.045	(.026|.148)	−.040	.004	.045	.046***	.044***	.008*	.000
Pd1	.119**	(.027|.220)	−.058	−.049	.062	.064***	.050***	.018***	.001
Pd2	−.001	(−.035|.028)	−.028	.073	.020	.033***	.033***	.001	.032***††
Pd3	−.126***	(−.129|−.147)	.003	.248	−.025	.051***	.035***	.003	.033***††
Pd4	.111**	(.012|.215)	−.048	−.161	.049	.062***	.050***	.018***	.004
Pd5	.159***	(.090|.241)	−.033	−.162	.043	.073***	.047***	.017***	.004
Sc1	.175***	(.105|.295)	−.027	−.002	.035	.093***	.062***	.024***	.000
Sc2	.238***	(.197|.285)	−.019	.036	.049	.097***	.041***	.013**	.001
Sc3	.233***	(.158|.339)	−.034	.000	.050	.115***	.061***	.028***	.000
Sc4	.263***	(.205|.324)	−.017	−.007	.035	.114***	.045***	.018***	.000
Sc5	.170***	(.105|.262)	−.042	.006	.056	.081***	.052***	.019***	.000
Sc6	.151***	(.086|.214)	−.017	.041	.024	.064***	.042***	.011**	.001
Antisocial PD diagnosis									
Pd1	.265***	(.262|.271)	.036	.097	−.013	.070***	.000		
Pd2	.218***	(.210|.246)	.112	−.055	.005	.049***	.000		
Pd3	.016	(.052|−.052)	.067	.118	−.044	.003	.003		
Pd4	.128***	(.115|.149)	.027	−.077	.008	.016**	.000		
Pd5	.170***	(.140|.235)	.015	−.213	.024	.031***	.002		

(*table continues*)

Table 6 (*continued*)

Diagnosis H & L scale	r_{xy}	(A. A.|C.)	Full model b IV	Full model b Race	Full model b IV × Race	Full model R^2	Prediction bias R^2D	Slope bias R^2D	Intercept bias R^2D
Ma1	.239***	(.236|.251)	.114	.025	.005	.058***	.001		
Ma2	.183***	(.171|.209)	.048	−.081	.013	.034***	.000		
Ma3	.012	(.058|−.082)	.086	.166	−.060	.005	.005		
Ma4	.136***	(.123|.164)	.040	−.029	.012	.019**	.000		
Avoidant PD diagnosis									
Sc1	.143***	(.095|.232)	−.012	−.010	.020	.041***	.022***	.009*	.000
Sc2	.185***	(.075|.313)	−.042	−.059	.054	.062***	.028***	.020***	.002
Sc3	.148***	(.143|.184)	.002	.038	.014	.036***	.013*	.003	.001
Sc4	.168***	(.097|.251)	−.014	−.032	.023	.046***	.018**	.010**	.001
Sc5	.073	(.029|.150)	−.024	.001	.028	.021**	.016**	.006	.010*††
Sc6	.120**	(.095|.142)	.000	.036	.008	.019**	.019*	.002	.019*††
Si1	.195***	(.097|.308)	−.018	−.089	.026	.062***	.025***	.017***	.004
Si2	.215***	(.142|.310)	−.018	−.062	.036	.068***	.023***	.014**	.002
Si3	.121***	(.080|.193)	−.008	−.027	.014	.031***	.017**	.006	.011**†
Borderline									
Pd1	.360***	(.288|.451)	−.008	−.125	.075	.162***	.034***	.018***	.003
Pd2	.057	(.089|.017)	.054	.267	−.023	.024***	.021***	.001	.020***††
Pd3	−.172***	(−.085|−.294)	.060	.427	−.084	.067***	.035***	.016***	.032***†
Pd4	.291***	(.212|.390)	−.021	−.266	.062	.118***	.036***	.018***	.006*†
Pd5	.330***	(.249|.430)	−.018	−.321	.061	.143***	.036***	.021***	.009**†
Sc1	.385***	(.286|.527)	−.013	−.083	.047	.199***	.054***	.029***	.002
Sc2	.372***	(.301|.442)	.006	−.030	.063	.162***	.027***	.014***	.000
Sc3	.384***	(.293|.501)	−.012	−.054	.058	.188***	.043***	.025***	.001
Sc4	.381***	(.300|.459)	.001	−.070	.041	.169***	.028***	.016***	.001
Sc5	.335***	(.242|.459)	−.030	−.086	.081	.154***	.045***	.026***	.002
Sc6	.343***	(.291|.403)	.010	.014	.025	.142***	.024***	.008*	.000
Narcisistic									
Sc1	.050	(.058|.021)	.013	.030	−.005	.002			
Sc2	.033	(.021|.017)	.007	.002	−.001	.000			
Sc3	.017	(−.004|.020)	−.005	−.013	.004	.000			
Sc4	.017	(−.010|.039)	−.009	−.031	.007	.001			
Sc5	.026	(.025|−.032)	.019	.035	−.013	.001			
Sc6	.072	(.092|−.022)	.028	.074	−.015	.006			
Ma1	.023	(.009|.067)	−.017	−.050	.020	.001			
Ma2	.010	(.005|.013)	.000	−.012	.002	.000			
Ma3	.080*	(.110|.038)	.057	.073	−.023	.009			
Ma4	.079*	(.095|.055)	.041	.070	−.013	.007			
Gambling diagnosis									
Pd1	.159***	(.112|.212)	−.042	.111	.063	.108***	.084***	.011**	.002
Pd2	−.093*	(−.011|−.145)	.072	.657	−.075	.089***	.080***	.008*	.028***††
Pd3	−.006	(.027|−.047)	.033	.420	−.027	.076***	.072***	.001	.075***††
Pd4	.164***	(.172|.194)	−.011	.090	.037	.107***	.081***	.006*	.001
Pd5	.141***	(.142|.164)	−.011	.106	.030	.098***	.079***	.004	.074***††

Note. Tests of slope and intercept bias were conducted only if evidence of predictive bias was indicated. The full model displays unstandardized regression coefficients, shown as b. Race was coded 1 for African American participants and 2 for Caucasian participants. Full model = scale, race, and the interaction term (IV × Race); r_{xy} = correlation between the predictor scale and diagnosis only; H & L scale = Harris and Lingoes scale; A. A. = correlation between predictor scale and diagnosis only for African American patients; C. = correlation between predictor scale and diagnosis only for Caucasian patients; IV = scale; R^2D = the change in the variance accounted for by the addition of the model; Prediction bias R^2D = comparison of the scale alone with the full model; Slope bias R^2D = comparison of the scale plus racial background with the full model. If slope bias is significant, then the Intercept bias R^2D = comparison of the scale plus interaction term with the full model. If there is no evidence for slope bias, then the Intercept bias R^2D = comparison of the scale alone with the scale plus racial background variable. PTSD = posttraumatic stress disorder; Hy1 = Denial of Social Anxiety; Hy2 = Need for Affection; Hy3 = Lassitude-Malaise; Hy4 = Somatic Complaints; Hy5 = Inhibition of Aggression; Sc1 = Social Alienation; Sc2 = Emotional Alienation; Sc3 = Lack of Ego Mastery, Cognitive; Sc4 = Lack of Ego Mastery, Conative; Sc5 = Lack of Ego Mastery, Defective Inhibition; Sc6 = Bizarre Sensory Experiences; D1 = Subjective Depression; D2 = Psychomotor Retardation; D3 = Physical Malfunctioning; D4 = Mental Dullness; D5 = Brooding; Ma1 = Amorality; Ma2 = Psychomotor Acceleration; Ma3 = Imperturbability; Ma4 = Ego Inflation Social Introversion Subscales; Pd1 = Familial Discord; Pd2 = Authority Problems; Pd3 = Social Imperturbability; Pd4 = Social Alienation; Pd5 = Self- Alienation; Si1 = Shyness/Self-Consciousness; Si2 = Social Avoidance; Si3 = Alienation–Self and Others.

† = overprediction for Caucasians across test scores below the point at which regression lines intersect, while overpredicting for African Americans scores above the point of intersection. †† = overprediction for African American across the entire range of test scores.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 7
Regression Analysis of Diagnoses On Restructured Clinical (RC) Scales, Race, and the Scale × Race Interaction Term (Men Only, ns = 639–755)

			Full model						
			b						
Diagnosis and scale	r_{xy}	(A. A.|C.)	IV	Race	IV × Race	R^2	Prediction bias R^2D	Slope bias R^2D	Intercept bias R^2D
PTSD diagnosis									
RCd	.335***	(.275|.361)	.008	.070	.014	.143***	.035***	.006*	.001
RC1	.300***	(.279|.299)	.017	.158	.010	.122***	.034***	.002	
RC2	.354***	(.291|.355)	.025	.075	.020	.144***	.025***	.003	.021***††
RC7	.300***	(.249|.362)	.004	.085	.019	.137***	.049***	.008*	.001
RC8	.253***	(.297|.238)	.035	.269	.005	.109***	.047***	.000	.047***††
Depression diagnosis									
RCd	.385***	(.329|.423)	.010	−.002	.016	.171***	.025***	.008*	.000
RC2	.386***	(.294|.433)	.015	−.057	.032	.163***	.019***	.008*	.000
RC7	.199***	(.247|.356)	.005	.053	.019	.120***	.038***	.007*	.000
RC8	.065	(.215|.221)	.016	.185	.014	.074***	.035***	.002	.033***††
Schizophrenia diagnosis									
RCd	.145***	(.147|.136)	.012	.039	−.001	.022**	.001		
RC2	.084*	(.051|.124)	−.003	−.035	.011	.008	.002		
RC6	.354***	(.341|.317)	.060	.087	.005	.133***	.007		
RC7	.191***	(.228|.139)	.027	.126	−.007	.041***	.004		
RC8	.304***	(.359|.195)	.064	.151	−.018	.101***	.008		
Bipolar disorder									
RCd	.230***	(.139|.304)	−.011	−.035	.017	.100***	.047***	.020***	.000
RC2	.249***	(.197|.275)	−.007	−.005	.024	.094***	.032***	.010**	.000
RC4	.128***	(.042|.225)	−.024	−.134	.028	.066***	.049***	.015***	.003
RC7	.181***	(.112|.272)	−.013	−.003	.019	.085***	.052***	.017***	.000
RC8	.125**	(−.040|.183)	−.012	.087	.020	.058***	.042***	.007*	.004
RC9	.076	(−.048|.234)	−.031	−.226	.028	.068***	.063***	.027***	.008*†
Antisocial PD diagnosis									
RCd	.141***	(.121|.184)	.008	−.091	.005	.021***	.001		
RC3	.109**	(.080|.150)	.011	−.059	.008	.012*	.000		
RC4	.357***	(.341|.400)	.069	.024	−.002	.128***	.000		
RC6	.183***	(.209|.102)	.075	.076	−.023	.035***	.002		
RC8	.131***	(.138|.105)	.031	.015	−.005	.017**	.000		
RC9	.219***	(.187|.284)	.018	−.112	.009	.049***	.001		
Borderline PD diagnosis									
RCd	.390***	(.296|.477)	.000	−.113	.020	.179***	.028***	.018***	.003
RC3	.165***	(.080|.308)	−.026	−.142	.038	.071***	.045***	.015***	.002
RC4	.311***	(.213|.422)	−.008	−.231	.035	.132***	.036***	.016***	.005
RC7	.339***	(.253|.454)	−.005	−.076	.025	.154***	.041***	.019***	.001
RC8	.290***	(.277|.346)	.006	.064	.027	.117***	.033***	.008*	.001
Narcissistic PD diagnosis									
RCd	.035	(.023|.027)	.002	.000	.000	.001	.000		
RC3	−.044	(−.049|−.036)	−.012	−.044	.004	.002	.000		
RC8	.059	(.054|.011)	.013	.025	−.006	.002	.000		
RC9	.036	(.024|.070)	−.001	−.049	.004	.002	.000		
Avoidant PD diagnosis									
RCd	.175***	(.121|.240)	−.004	−.029	.009	.045	.015**	.007*	.000
RC7	.097*	(−.041|.163)	−.006	.001	.009	.024	.015**	.005	.010**††
Gambling diagnosis									
RCd	.210***	(.155|.257)	−.011	.107	.019	.125	.080***	.013***	.002
RC4	−.081*	(.088|−.127)	.040	.658	−.031	.087	.082***	.011**	.038***††

Note. Tests of slope and intercept bias were conducted only if evidence of predictive bias was indicated. The full model displays unstandardized regression coefficients, shown as b. Race was coded 1 for African American participants and 2 for Caucasian participants. Full model = scale, race, and the interaction term (IV × Race); r_{xy} = correlation between the predictor scale and diagnosis only; A. A. = correlation between predictor scale and diagnosis only for African American patients; C. = correlation between predictor scale and diagnosis only for Caucasian patients; IV = scale; R^2D = the change in the variance accounted for by the addition of the model; Prediction bias R^2D = comparison of the scale alone with the full model; Slope bias R^2D = comparison of the scale plus racial background with the full model. If slope bias is significant, then the Intercept bias R^2D = comparison of the scale plus interaction term with the full model. If there is no evidence for slope bias, then the Intercept bias R^2D = comparison of the scale alone with the scale plus racial background variable. PTSD = posttraumatic stress disorder; PD = personality disorder.
† = overprediction for Caucasians across test scores below the point at which regression lines intersect, while overpredicting for African Americans scores above the point of intersection. †† = overprediction for African American across the entire range of test scores.
* $p = .05$. ** $p = .01$. *** $p = .001$.

The slope bias and intercept bias findings in the present study were consistently small to moderate in size (Cohen, 1988). Of those scales that displayed slope bias, or a less accurate relationship with related diagnoses for African American inpatients as compared with their Caucasian counterparts, Clinical Scales 2 and 9 and bipolar diagnosis (including their respective Harris and Lingoes scales) displayed the most disparity, whereas both ASP and ANG showed the greatest bias predicting borderline personality. Those Clinical scales with the largest degree of intercept bias resulting in overprediction for African American inpatients included Scales 1, 3, and 8 with PTSD and Scale 4 with gambling diagnosis. The Content scales that displayed the greatest overprediction for African Americans were FRS and SOD in the prediction of PTSD diagnosis.

Harris and Lingoes scales aid in identification of specific aspects of each Clinical scale contributing to the type and magnitude of observed bias. For example, the relationship between Scale 2 and major depression appears to result in a moderate level of intercept bias; however, when the Harris and Lingoes scales are examined, one can see that D2 and D3 display intercept bias, whereas D1, D4, and D5 produce slope bias. Alternately, the relationship between Clinical Scale 8 and schizophrenia displays no predictive bias; yet, the Scale 8 Harris and Lingoes Scale SC6 indicates a small yet significant slope and intercept bias.

Nearly all of the step-down hierarchical regressions produced significant predictive bias. In other words, the MMPI-2 Clinical, RC, Content, Supplementary, and Harris and Lingoes scales analyzed in this study display a systematically biased relationship with diagnosis between African American and Caucasian inpatients. Distinct from previous predictive bias studies (i.e., Arbisi et al., 2002) is the particular type of bias; whereas Arbisi and colleagues found primarily the presence of intercept bias, the present study indicates both slope and intercept bias. The exception includes the RC scales, which displayed somewhat uniform slope bias (19 instances of slope bias vs. six instances of intercept bias).

The RC scales used to predict both schizophrenia and antisocial personality diagnoses offer promise in that they display significant moderately sized (on average) bivariate relationships between predictor and diagnoses. As an additional distinction from traditional scales, the RC scales appear to display relatively more slope bias than intercept bias. Half of the relationships assessed in the present study demonstrate predictive bias, and most are biased in terms of accuracy rather than under- or overprediction of diagnostic status. In almost all instances of slope bias, the scale is less accurate for African Americans, as can be seen by the subgroup differences in bivariate correlation coefficients (see Table 7).

The importance of examining differential prediction of diagnosis is highlighted by an examination of predictions of bipolar diagnosis using the Ma2 scale. Calculating a simple bivariate correlation for the entire group might have caused a researcher or practitioner to infer that Ma2 is a weak predictor ($r = .085$). Although we see that as a predictor for Caucausian participants Ma2 fairs well ($r = .212$), the overall correlation was attenuated by the African American subsample, which displayed a small inverse relationship ($r = -.017$). Subsequent analyses show predictive bias of both slope and intercept. Certainly using a test to make classifications that offers zero (or inverse) predictive validity for one group, but adequately predicts for another group, should raise concern among practitioners.

That the results from the present study are at odds with a number of prior studies suggests the importance of evaluating potential biases specific to the setting, patient group, and type of criterion examined. Whereas generalizability of assessment knowledge is a laudable goal, numerous theorists and researchers (e.g., Meehl, 1954; Stricker & Trierweiler, 1995) warn of clinical situations that fall outside the purview of generalization. Results across various studies searching for biased predictions from MMPI-2 results may be understood as examples highlighting the importance of specific criteria, populations, and settings on conclusions regarding racial bias and the MMPI-2.

Important caveats regarding the present study are warranted. First, whereas the use of structured diagnostic interviews partially limits the judgment required by interviewers, judgment is still required in appreciating verbal reports, choosing prompts to obtain clarifying information, and decontextualizing patient narratives to conform to diagnostic nosology. Even if minimized by structured interview, implicit models of symptom attributions are still relevant and pose a threat to validity of diagnostic status in the present study. In other words, structured interviews involve clinical judgment, which means these are not bias-free evaluations. In addition, whereas previous empirical research suggests the adequacy of the SCID's psychometric properties (e.g., Arntz et al., 1992; Dreessen & Arntz, 1998; First et al., 1995; Maffei et al., 1997; Weiss et al., 1995), we were unable to estimate the reliability of diagnoses in the present study due to use of archival data. This may cause concern about accuracy of the diagnoses in the present study. Subsequent research would benefit from estimating the reliability of the diagnostic instrument locally before conducting a similar study of predictive bias.

In addition, our use of an ordinal tripartite diagnostic criterion is not ideal (Helmchen & Linden, 2000). The diagnostic data available were coded in a three-level format, a format that has been supported in previous validation studies (e.g., Dreessen & Arntz, 1998; Dreessen et al., 1998). And although used in similar studies (e.g., Quirk et al., 2003), it is clear that psychiatric criteria are more continuous than categorical. The use of cut scores, or trichotomization in this case, leads to the loss of information (Widiger & Trull, 2007). It is plausible that such loss of criterion variance reduced the observed estimates of bias both in prevalence and in magnitude (Wilcox, 1998). Future research would benefit from diagnostic criteria in the form of symptom counts or more continuous rating formats.

As noted in the discussion of generalizability, differences between type and magnitude of biases found in the present study and those found by Arbisi et al. (2002) may be due in part to differences in sample characteristics. There is the probability that some patients in the present sample were motivated to dissimulate. Patients may have been motivated to overreport symptoms to secure limited treatment resources, whereas others may have been motivated to deny the severity of any psychiatric dysfunction beyond substance use. As a final note, overreporting of substance abuse symptoms and denial of severity of psychiatric dysfunction may be responsible for the relatively low (compared with previous research) comorbidity in the present study.

Aside from these considerations, the results from the present study demonstrate consistent group differences in prediction of

select psychiatric diagnoses by race using the most relevant MMPI-2 scales within this sample. Such results add to the body of research findings regarding racial bias in the use of the MMPI-2. As such, these results may support the contention that clinical setting, patient population, and criterion type are important causes of this observed inconsistency (rather than merely representing error variance). Research on the predictive bias of the MMPI-2 would benefit from additional sample characteristics. As one example, with regard to veteran inpatients, it would be useful to measure the influence of combat exposure (e.g., Combat Exposure Scale; see Keane, Fairbank, Caddell, Zimering, & Mora, 1989) on predictive efficiency and bias. Also, the practical impact of modest-sized intercept and slope biases on clinical outcome is an understudied area that is important in appreciating the practical impact of bias on clinical decisions. Finally, the presence of racially biased diagnostic predictions across the majority of scales (including the RC scales) raises serious concerns with regard to using the MMPI-2 to predict diagnostic status within the male veteran substance-abusing population. A necessary area of future research is the practical impact of racial bias when using the MMPI-2 to classify and subsequently treat patients.

References

Adebimpe, V. R. (1981). Overview: White norms and psychiatric diagnosis of black patients. *American Journal of Psychiatry, 138,* 279–285.

Aponte, J. F., & Johnson, L. R. (2000). *The impact of culture on the intervention and treatment of ethnic populations.* Needham Heights, MA: Allyn & Bacon.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2002). A comparison of MMPI-2 validity in African American and Caucasian psychiatric inpatients. *Psychological Assessment, 14,* 3–15.

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

Arntz, A., van Beijsterveldt, B., Hoekstra, R., Hofman, A., Eussen, M., & Sallaerts, S. (1992). The interrater reliability of a Dutch version of the Structured Clinical Interview for DSM-III-R Personality Disorders. *Acta Psychiatrica Scandinavica, 85,* 394–400.

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.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales.* Minneapolis: University of Minnesota Press.

Cleary, A. T. (1968). Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement, 5,* 115–124.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dreessen, L., & Arntz, A. (1998). Short-interval test-retest interrater reliability of the structured clinical interview for DSM-III-R personality disorders (SCID-II) in outpatients. *Journal of Personality Disorders, 12,* 138–148.

Dreessen, L., Hildebrand, M., & Arntz, A. (1998). Patient-informant concordance on the structured clinical interview for DSM-III-R personality disorders (SCID-II). *Journal of Personality Disorders, 12,* 149–161.

Elwood, R. W. (1993). Psychological tests and clinical discriminations: Beginning to address the base rate problem. *Clinical Psychology Review, 13,* 409–419.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The Structured Clinical Interview for DSM-III-R personality disorders (SCID-II): II. Multi-site test-retest reliability study. *Journal of Personality Disorders, 9,* 92–104.

Fogelson, D. L., Nuechterlein, K. H., Asarnow, R. F., & Subotnik, K. L. (1991). Interrater reliability of the structured clinical interview for DSM-III–R, axis II: Schizophrenia spectrum and affective spectrum disorders. *Psychiatry Research, 39,* 55–63.

Franklin, C. L., Sheeran, T., & Zimmerman, M. (2002). Screening for trauma histories, posttraumatic stress disorder (PTSD), and subthreshold PTSD in psychiatric outpatients. *Psychological Assessment, 14,* 467–471.

Frueh, B. C., Gold, P. B., de Arellano, M. A., & Brady, K. L. (1997). A racial comparison of combat veterans evaluated for PTSD. *Journal of Personality Assessment, 68,* 692–702.

Frueh, B. C., Smith, D. W., & Libet, J. M. (1996). Racial differences on psychological measures in combat veterans seeking treatment for PTSD. *Journal of Personality Assessment, 66,* 41–53.

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

Gynther, M. D. (1972). White norms and black MMPIs: A prescription for discrimination? *Psychological Bulletin, 78,* 386–402.

Hall, G. C. N., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment, 11,* 186–187.

Harris, R., & Lingoes, J. (1955). *Subscales for the Minnesota Multiphasic Personality Inventory* [Mimeographed materials]. The Langley Porter Clinic.

Harris, R., & Lingoes, J. (1968). *Subscales for the Minnesota Multiphasic Personality Inventory* [Mimeographed materials]. The Langley Porter Clinic.

Harrison, P. A., & Hoffman, N. G. (1985). *The Substance Use Disorders Diagnostic Schedule.* St. Paul, MN: Ramsey Clinic.

Helmchen, H., & Linden, M. (2000). Subthreshold disorders in psychiatry: Clinical reality, methodological artifact, and the double-threshold problem. *Comprehensive Psychiatry, 41,* 1–7.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs, 1,* 57–93.

Keane, T. M., Fairbank, J. A., Caddell, J. M., Zimering, R. T., & Mora, C. A. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment, 1,* 53–55.

Keane, T. M., Malloy, P. F., & Fairbank, J. A. (1984). Empirical development of an MMPI subscale for the assessment of combat-related post-traumatic stress disorder. *Journal of Consulting and Clinical Psychology, 52,* 888–891.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133–139.

Maffei, C., Fossati, A., Agostoni, I., Barraco, A., Bagnato, M., Deborah, D., et al. (1997). Interrater reliability and internal consistency of the structured clinical interview for DSM-IV axis II personality disorders (SCID- II), version 2.0. *Journal of Personality Disorders, 11,* 279–284.

McNulty, J. L., Graham, J. R., Ben-Porath, Y. S., & Stein, L. A. R. (1997). Comparative validity of MMPI-2 scores of African American and Caucasian mental health center clients. *Psychological Assessment, 9,* 464–470.

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

Morrison, T. L., Edwards, D. W., & Weissman, H. N. (1994). The MMPI and MMPI-2 as predictors of psychiatric diagnosis in an outpatient sample. *Journal of Personality Assessment, 62,* 17–30.

Munley, P. H., Busby, R. M., & Jaynes, G. (1997). MMPI–2 findings in schizophrenia and depression. *Psychological Assessment, 9,* 508–511.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika, 29,* 241–256.

Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. L. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the revised NEO personality inventory. *Psychological Assessment, 15,* 311–325.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88,* 307–321.

Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement, 41,* 73–84.

Rotundo, M., & Sackett, P. R. (1999). Effect of rater race on conclusions regarding differential prediction in cognitive ability test. *Journal of Applied Psychology, 84,* 815–822.

Schlenger, W. E., & Kulka, R. A. (1989). *PTSD scale development for the MMPI-2*. Research Triangle Parn, NC: Research Triangle Institute.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III–R (SCID): I. history, rationale, and description. *Archives of General Psychiatry, 49,* 624–629.

Stricker, G., & Trierweiler, S. J. (1995). The local clinical scientist: A bridge between science and practice. *American Psychologist, 50,* 995–1002.

Stukenberg, K., Brady, C., & Klinetob, N. (2000). Use of the MMPI-2's VRIN scale with severely disturbed populations: Consistent responding may be more problematic than inconsistent responding. *Psychological Reports, 86,* 3–14.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 restructured clinical (RC) scales: Development, validation, and interpretation.* Minneapolis: University of Minnesota Press.

Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2? *Psychological Assessment, 6,* 212–217.

Weiss, R. D., Najavits, L. M., Muenz, L. R., & Hufford, C. (1995). Twelve-month test-retest reliability of the structured clinical interview for DSM-III-R personality disorders in cocaine-dependent patients. *Comparative Psychiatry, 36,* 384–389.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62,* 71–83.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53,* 300–314.

Received December 12, 2007
Revision received December 10, 2008
Accepted December 15, 2008 ■