Exploring Possible Ethnic Differences and Bias in the Rorschach Comprehensive System

Gregory J. Meyer
Department of Psychology

Comprehensive System scores, they ultimately concluded “Because there are important cross-cultural differences, and because appropriate norms have not been developed, it is doubtful whether the Comprehensive System should currently be used to evaluate members of American minority groups” (p. 342).

It is not a trivial matter to suggest the Comprehensive System may not be appropriate for use with minorities. Thus, it is important to examine the data that may lead to such a suggestion.

Out of the 11 studies Wood and Lilienfeld (1999) cited to support ethnic differences, only 6 used the Comprehensive System. Sangro (1997) did not examine differences in specific scores but instead used a sample of Spanish outpatients to generate location, form quality, and popular tables that were then qualitatively (i.e., nonstatistically) contrasted with Exner’s (1993) tables. The remaining 5 studies gathered convenience samples of a targeted minority group and compared them to Exner’s nonpatient data (Aposhian, 1995; Baca, 1994; Glass, Bieber, & Tkachuk, 1996; Krall et al., 1983; Sanchez, 1993). In other contexts, Wood has strongly criticized studies that collect data from a target sample and then compare the results to Exner’s nonpatients. For instance, Wood, Nezworski, Stejskal, Garven, and West (1999) called comparisons with Exner’s nonpatient sample methodologically “flawed” and “inadequate for establishing the validity” (p. 124) of Rorschach scores. Wood et al. argued that target and control groups should only be compared when both groups were simultaneously collected by researchers. It is not clear how a design that Wood et al. believed was incapable of establishing positive evidence for Rorschach validity would nonetheless be strong enough to establish that the Comprehensive System

Although this brief review highlights how there is no consistent evidence of ethnic differences in Comprehensive System scores, a search for mean differences in minority and majority groups is a misdirected endeavor. Even if mean differences are observed (e.g., lower *R* or lower IQ in a minority group), this does not provide any specific information about test bias. In fact, test bias may exist when there are no mean differences across ethnic groups and bias may be nonexistent when there are huge ethnic differences. Consequently, even though researchers often search for ethnic differences as a way to explore test bias (e.g., Greene, 2000; Hall, Bansal, & Lopez, 1999), the strategy ultimately does not provide clear data from which inferences can be drawn.

A mean difference on its own does not indicate bias because it may accurately identify a characteristic that truly varies across the minority and majority thisinfora star

-

rectly, without the need for the more time-consuming and expensive direct determination of the test's predictive validity in terms of an external criterion" (p. 447). Because factor analysis simultaneously evaluates the relationship among a large number of variables, the presence of differences in factor structure across ethnic groups makes it more likely that slope or intercept bias will be present when the test scores are evaluated against external criteria. However, if test variables produce comparable factors and factor loadings across ethnic groups, it is also "reasonable to expect parallel or very nearly parallel regression lines" (Humphreys & Taber, 1973, p. 108) and thereby demonstrate an absence of slope bias.

This study has three goals. The first is to assess whether there are ethnic differences in the Comprehensive System. Although mean differences are not sufficient to demonstrate test bias, because it has been asserted that these differences exist, the question will be investigated using a relatively large sample of consecutively evaluated minority and majority patients.

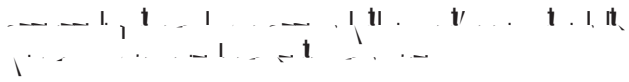
If Frank (1992) was correct and African Americans are prone to display lower degrees of self-disclosure, this response style should affect the Rorschach's first factor (see Meyer, 1999; Meyer, Riethmiller, Brooks, Benoit, & Handler, 2000). Convenient markers of the first factor are R and Lambda or Form%. Because Lambda and Form% are conceptually equivalent variables but Form% has a distribution that is more optimal for research (Meyer, Viglione, & Exner, 2001), Form% will be utilized in this study. Frank's speculations lead to the hypotheses that R should be lower and Form% higher in the minority sample. No other salient ethnic differences are expected for Comprehensive System scores.

The last statement begs the question of what constitutes a salient difference. Although there is no single correct way to define thumkefine4m,fine07i20.ih(st)ethnic-v(correct)69p

In his review of ethnic differences on the MMPI, Greene (2000) concluded that differences were less likely to emerge when investigators controlled moderator variables like test validity, age, education, gender, socioeconomic status, and intelligence. To address this, differences will be examined without correcting for any moderators and also after controlling for a number of demographic and test-related

The ethnic composition of the sample was European American = 242, African American = 157, Hispanic American = 14, Asian American = 16, Native American

nomic status, intelligence, and other variables, like referral source, were not measured in this sample and thus could not be controlled. To the extent that ethnicity is associated with some of these variables, failing to measure them will allow socioeconomic status, intelligence, or referral patterns to influence Rorschach scores, even though their impact will be attributed to ethnicity.[3]



To assess associations with ethnicity, I examined 188 Comprehensive System variables. These included: (a) all the scores on Exner's (1993) structural summary (except as noted later); (b) all the individual criteria for the Schizophrenia Index (SCZI), Depression Index (DEPI), Coping Deficit Index (CDI), Suicide Constellation (S – CON), Hypervigilance Index (HVI), and Obsessive Style Index (OBS); (c) total scores for the SCZI, DEPI, CDI, S – CON, HVI, and OBS; and (d) dichotomous scores (i.e., positive vs. not) for the SCZI, DEPI, CDI, S – CON, and HVI. Dichotomous OBS scores were not included because no patient was positive on the OBS. Several score combinations dealing with form quality were also excluded. DQv/+ with FQ–, DQv with FQ–, form quality for pure form responses, and form quality for space responses (except S – %) were never entered into the original database and could not be calculated by computer from the existing information.

Two sets of ethnic comparisons were undertaken. European Americans were first compared to African Americans and then to all non-European Americans. To quantify the association between ethnicity and Rorschach scores, correlations were computed between each score and ethnic status (European American = 0; minority = 1). Nonparametric analyses were conducted using the Spearman rank order correlation (r_s) and parametric analyses were conducted using the Pearson correlation (r). Correlations were used as an effect size measure over Cohen's d for two reasons. First, in a parametric analysis, the standardized mean difference between two groups (i.e., Cohen's d) can be directly converted into a correlation, making these two statistics interchangeable (see Rosenthal, 1991). Second, it is easy to compare Spearman and Pearson correlations.[4] However, there is no nonparametric statistic that can be com-

[3]The patients were evaluated at the University of Chicago Medical Center, located on the south side of Chicago. Although the medical center serves a culturally diverse population, there are clear boundaries that separate the generally wealthy and university-affiliated (i.e., faculty, students, staff) Hyde Park and Kenwood neighborhoods from the surrounding impoverished and predominantly African American neighborhoods. In addition to serving the local community, the experts, specialty clinics, and specialty units at the Medical Center also attract patients with financial resources from distant communities.

[4]The formula for the Spearman correlation is just a simplification of the Pearson formula and exactly equivalent results are obtained when either formula is used with ranked variables (e.g., Cohen & Cohen, 1983, pp. 40–41).

pared to Cohen's *d* in an easy or direct manner. Thus, correlations provided a more optimal effect size measure for the purposes of this study.

Before using partial correlations to control for demographic variables, I examined the association between ethnicity and each Rorschach score using both r_s and *r*. If there were salient differences between the nonparametric and parametric results, it would indicate the nonnormal distributions associated with some Rorschach scores were interfering with the parametric statistical analyses. Such evidence would strongly argue against using partial correlations from this study to draw inferences about other samples. On the other hand, highly similar parametric and nonparametric coefficients would suggest that the partialled results are likely to generalize to other samples, despite the nonnormal shape of many Rorschach scores. In the full sample of patients, the differences between r_s and *r* were trivial (*M* diff = .002, *Mdn* = .000, range –.058 to .067). Similar findings were obtained from the partially matched subsamples (i.e., *M* = .001, *Mdn* = .000). Because the parametric and nonparametric results were virtually identical, the partial correlations that control for moderators are likely to generalize to other samples.

As recommended by Nunnally and Bernstein (1994), regression equations were created to simultaneously evaluate slope and intercept bias. A number of Rorschach predictor variables were evaluated against relevant criterion measures. The available criteria included education, diagnostic determinations, and MMPI–2 scales.

As a gross index of cognitive capacity, education level was viewed as a potentially relevant criterion for the following Comprehensive System predictors: organizational efforts (Zf), determinant blends (Blend), integrated perceptions (DQ+), organized resources (EA), human movement (M), and content indicative of intellectualization (Intellectualization Index). All scores were considered as percentages to control the effects of R.

The diagnosis of a psychotic disorder was seen as a relevant validity criterion for the SCZI, the percentage of responses containing good form quality (X + %; which should have an inverse relationship with the criterion), the percentage of responses with minus form quality (X – %), and disorganized or illogical thought processes as indicated by the Weighted Sum of 6 Special Scores (WSum6). A depressive disorder diagnosis was used as a suboptimal (see Jørgensen, Andersen, & Dam, 2000) criterion for the Depression Index.[5] The maximum severity of impairment associated with the diagnoses assigned to each patient (see Dawes, 1999; Meyer & Resnick, 1996) was considered a reasonable criterion for three general

[5]The *DEPI* ultimately was not significantly associated with diagnoses and so could have been dropped from the analyses. It was retained for the sake of complete reporting.

impairment variables that can be derived from the Comprehensive System: the Human Experience Variable (HEV; see Burns & Viglione, 1996), the Ego Impair-

dicated the matrix was reasonable for analysis. To determine the proper number of factors to extract, several criteria were considered. Parallel analysis (Zwick & Velicer, 1986) suggested up to eight factors. However, parallel analysis tends to over extract factors when the matrix contains complex variables with loadings on more than one factor and it also tends to retain poorly defined factors (Glorfeld, 1995; Zwick & Velicer, 1986). A plot of the eigenvalues (the first 10 of which were: 15.21, 5.06, 3.91, 3.72, 2.74, 2.11, 1.81, 1.53, 1.43, 1.34) suggested four factors should be extracted. Consistent with this, Guadagnoli and Velicer's (1988) criteria for factor retention (i.e., 4+ variables with loadings > .60 or 10+ variables with loadings > .40), indicated that four factors should be retained. Extracting any additional factors caused some to be poorly defined. Thus, four factors were extracted from each of the ethnic samples. For European Americans the first 10 eigenvalues were 16.36, 5.18, 3.80, 3.37, 2.69, 2.13, 1.90, 1.65, 1.56, and 1.54. Parallel analysis suggested retaining seven factors, although a scree plot suggested three or four factors. A four-factor solution met Guadagnoli and Velicer's criteria. For non-European Americans, the first 10 eigenvalues were 13.09, 5.23, 4.29, 3.91, 3.04, 2.28, 1.94, 1.67, 1.57, and 1.48. Parallel analysis again suggested seven factors, whereas the scree plot suggested three or four. Because the four-factor solution met Guadagnoli and Velicer's criteria for factor retention, the four-factor solution was considered optimal in this sample as well.

Jensen (1980) noted how the statistical test for ethnic bias in factor structure could be confounded when using rotated factor solutions because rotation algorithms can capitalize on chance associations to orient the factor axes, producing seeming differences across groups that are not genuine. Although this is a concern, for the present analyses I examined cross-ethnic correspondence using both unrotated and varimax rotated solutions.

Table 1 provides an overview of ethnic differences on demographic, diagnostic, and Rorschach response style variables for the complete samples. It can be seen that both of the minority samples were less educated, more often inpatients, more often female, and less often married than the European American patients. There were no differences in age. In seeming support of Frank's (1992) hypothesis, the minority patients produced fewer responses and a higher percentage of pure form responses. Among the subset of patients with external diagnoses, the minority patients more often received a psychotic disorder diagnosis and less often received a depressive disorder diagnosis. Overall diagnostic severity was higher in the minority samples, but this only reached statistical significance when the contrast was with African Americans.

	EA			AA			NEA			t Value	
	M	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*EA Versus AA*	*EA Versus NEA*
Demographic and Rorschach variables[a]											
Education level	3.08	1.23		2.48	1.13		2.58	1.23		4.95**	4.22**
Inpatient			48.35			65.61			66.32	–3.43**	–3.79**
Female			50.83			64.97			60.53	–2.80**	–2.02*
Are married			31.82			22.29			21.58	2.07*	2.38*
Age	35.38	12.25		35.29	12.32		33.99	12.27		0.08	1.17
R	24.68	10.70		20.57	7.59		21.21	7.86		4.18**	3.75**
Form%	0.362	0.182		0.426	0.196		0.416	0.187		–3.35**	–3.04**
Diagnostic variables[b]											
Psychotic disorder			37.13			58.33			58.49	–3.72**	–4.11**
Depressive disorder			74.25			62.88			64.15	2.12*	1.98*
Diagnostic severity	3.43	1.03		3.68	1.16		3.65	1.16		–2.03*	–1.87

Note. Education Level 1 = ≤ 11 years, Level 2 = 12 years, Level 3 = 13 to 15 years, Level 4 = 16 years, and 5 = > 16 years; diagnostic severity, 1 = mild, 5 = severe; dichotomous variables, *t* values were computed by treating the variable as numeric (e.g., male = 1, female = 2); results were equivalent to those obtained from chi-square. EA = European American; AA = African American; NWA = Non-European American.

[a]European American: N = 242; African American: N = 157; Non-European American: N = 190. [b]European American: N = 167; African American: N = 132; Non-European American: N = 159.

*p < .05. **p < .01.

Table 2 provides information about the subsamples after patients were matched

	EA			AA				EA			NEA			
	M	*SD*	%	*M*	*SD*	%	*t* Value	*M*	*SD*	%	*M*	*SD*	%	*t* Value
Demographic and Rorschach variables[a]														
Education level	2.61	1.07		2.61	1.07		0.00	2.76	1.17		2.76	1.19		0.00
Inpatient			55.00			55.00	0.00			56.46			56.46	0.00
Female			56.67			60.83	–0.65			55.10			55.78	–0.12
Are married			25.00			25.00	0.00			25.85			24.49	0.27
Age	35.81	12.66		35.43	12.64		0.23	34.51	12.48		34.28	12.47		0.16
R	23.21	9.08		21.25	8.12		1.76	23.31	8.78		22.01	8.41		1.29
Form%	0.392	0.182		0.425	0.191		–1.35	0.382	0.178		0.416	0.178		–1.64
Diagnostic variables[b]														
Psychotic disorder			28.41			54.74	–3.72**			31.78			56.03	–3.74**
Depressive disorder			73.86			62.11	1.70			74.77			64.66	1.64
Diagnostic severity	3.34	1.04		3.54	1.16		–1.20	3.36	1.05		3.56	1.15		–1.36

[illegible]

	Unmatched Sample						
	Uncorrected Coefficients		*Partialled Coefficients*			*Partially Matched Subsample*	
Comprehensive System Score	r_s	*r*	*1*	*2*	*3*	r_s	*r*
N of variables	188	188	188	188	188	185	185
M	–0.090	–0.090	–0.014	–.009	–.025	–0.046	–0.045
Minimum	–0.308	–0.290	–0.211	–.153	–.226	–0.220	–0.205
Maximum	0.198	0.198	0.178	.177	.156	0.185	0.185
SD	0.110	0.106	0.082	.071	.070	0.091	0.088
Kurtosis	–0.414	–0.325	–0.386	–.502	–.315	–0.588	–0.580
Skewness	0.418	0.465	0.281	.316	.129	0.369	0.406
% of variables with							
r > .25	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r < –.25	5.3	4.3	0.0	0.0	0.0	0.0	0.0
r > .20	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r < –.20	15.4	14.9	0.5	0.0	0.5	2.7	1.1
r > .15	3.2	2.1	3.2	1.1	0.5	1.6	1.1
r < –.15	33.0	29.3	3.7	1.1	2.7	10.8	10.3
% of variables with a statistically significant association with ethnicity[c]	22.3	22.3	0.5	0.0	0.5	0.0	0.0

Note. Based on scale where European American = 0, Non-European American = 1; Negative correlations indicate European Americans obtained higher scores; blank cells indicate the Rorschach score was not assigned in the subsample; r_s = Spearman rank order correlation; *r* = Pearson correlation; Partialled 1 = coefficients after controlling for R and Form% (except when those variables were the predictors); Partialled 2 = coefficients after controlling for the prior variables and education level, inpatient status, gender, and marital status; Partialled 3 = coefficients after controlling for the prior variables and psychotic diagnosis (*n* = 300).

[a]*n* = 399. [b]*n* = 240. [c]Bonferroni corrections were applied to account for the number of statistical tests in each column, not the full table.

The first section of Table 3 reports summary measures that characterize the distribution of ethnicity effect sizes. On average, there was a slight tendency for European Americans to obtain higher scores across the 188 Comprehensive System variables. However, these effects tended to be small (M r between –.009 and –.090). As can be seen from the second section of the table, when statistical controls were applied to the full sample or when patients were partially matched on demographic factors, no ethnicity effects were larger than *r* = |.25|, the criterion that has been used to define clinically important demographic influences in MMPI research.

The final section of Table 3 indicates the percentage of effect sizes that were statistically significant. To protect against inflated alpha levels and to account for

the large number of statistical tests that were conducted for this table, alpha levels were Bonferroni-adjusted (i.e., alpha/number of tests). These adjustments accounted for the 185 to 188 statistical tests reported in each column of Table 3. They did not account for the total number of statistical tests computed across all columns and thus are conservatively more likely to call a result "statistically significant" than would be warranted given the full experiment-wise error rate. As the table indicates, there were no statistically significant associations between ethnicity and the 188 Comprehensive System scores in the partially matched subsample of patients. In the unmatched sample, once the primary moderators were controlled (i.e., Partial 2), there were no significant associations. When a psychotic disorder was also controlled (i.e., Partial 3), only R had a statistically significant level of association, with European Americans producing more responses than African Americans.

Table 4 presents a summary of the findings that emerged from comparing European Americans to non-European Americans. In general, ethnicity had less of an impact in this set of analyses. As before, when statistical controls were applied to the full sample or when patients were partially matched on demographic factors, no effects were larger than $r = |.25|$. The bottom section of Table 4 indicates that there were no statistically significant associations between ethnicity and the 188 scores in the full sample of patients once moderators were controlled. In the partially matched subsample, there were no statistically significant associations in the parametric analysis, although two variables (Ay and 2AB + Art + Ay) reached a statistically significant level of association in the nonparametric analysis. For both of the latter, European Americans obtained higher scores, suggesting a higher propensity for intellectualization.

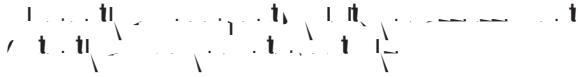


Table 5 presents convergent validity coefficients for the 17 targeted Rorschach predictor–criterion pairs. The left half of the table presents results for European Americans partially matched with African Americans, whereas the right side presents results for European Americans partially matched with non-European Americans. On each side of the table validity coefficients are reported for three groups: the combined majority and minority samples, just the European Americans, and then just the minority sample. Following these coefficients is a *z* statistic (z_{diff}) that evaluates whether the magnitude of the validity coefficients are different in the majority and minority samples. For some predictor–criterion pairs, it can be seen that validity coefficients are slightly larger in the minority sample (i.e., negative z_{diff} values), whereas for other predictor–criterion pairs the validity coefficient is slightly larger in the majority sample (i.e., positive z_{diff} values). However, these differences are due to sampling error. Across all the entries in Table 5 there are no statistically sig-

Unmatched Sample

	Uncorrected Coefficients		*Partialled Coefficients*			*Partially Matched Subsample*	
Comprehensive System Score	r_s	*r*	*1*	*2*	*3*	r_s	*r*
N of variables	188	188	188	188	188	187	187
M	–0.071	–0.071	–0.006	–0.004	–0.018	–0.032	–0.030
Minimum	–0.270	–0.269	–0.182	–0.155	–0.202	–0.221	–0.193
Maximum	0.191	0.196	0.184	0.149	0.139	0.178	0.189
SD	0.097	0.095	0.079	0.068	0.067	0.084	0.082
Kurtosis	–0.448	–0.337	–0.519	–0.605	–0.450	–0.348	–0.418
Skewness	0.412	0.438	0.309	0.270	–0.033	0.344	0.355
% of variables with							
r > .25	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r < –.25	1.1	1.1	0.0	0.0	0.0	0.0	0.0
r > .20	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r < –.20	6.9	7.4	0.0	0.0	0.5	1.6	0.0
r > .15	1.1	1.1	3.7	0.0	0.0	2.1	2.1
r							

nificant differences ($p < .05$) in Rorschach validity across ethnic groups. In fact, the mean z_{diff} was –.16 across the 17 analyses with African Americans and it was .00 across the 17 analyses with non-European Americans.

Table 6 extends these analyses by presenting findings for the regression equations that specifically tested for slope and intercept bias. Results for the partially matched sample of European and African Americans are presented on the left and results for the European and non-European Americans are presented on the right. Across all analyses, the tests for slope bias (i.e., differential validity) are not statistically significant after adjusting for multiple exploratory tests. Slope bias coefficients for Blend% with Education level reach a traditional ($p < .05$) level of significance in both samples. If the Blend% findings are not considered to be the

[illegible]2

	Partially Matched EA Versus AA				*Partially Matched EA Versus NEA*			
Criterion and Predictor	*All*	*EA*	*AA*	z_{diff}	*All*	*EA*	*NEA*	z_{diff}
Education level with								
Zf%	.16*	.05	.26**	–1.70	.15*	.07	.23**	–1.45
Blend%	.23**	.13	.36**	–1.84	.23**	.15	.33**	–1.66
DQ + %	.18**	.10	.25**	–1.15	.17**	.10	.23**	–1.17
EA%	.25**	.22*	.27**	–0.46	.27**	.27**	.26**	0.14
M%	.19**	.16	.21*	–0.34	.22**	.24**	.19*	0.42
Intell%	.24**	.23*	.27**	–0.34	.24**	.22**	.26**	–0.38
N	240	120	120		294	147	147	
Psychotic disorder with								
SCZI	.35**	.35**	.33**	0.15	.37**	.34**	.37**	–0.23
X + %	–.25**	–.21	–.27**	–0.48	–.25**	–.19*	–.30**	–0.86
X – %	.27**	.24*	.24*	–0.01	.28**	.23*	.28**	–0.43
WSum6	.37**	.42**	.36**	0.48	.37**	.37**	.38**	–0.06
Depressive disorder with								
DEPI	.05	–.06	.12	–1.23	.05	.01	.07	0.41
Diagnostic severity with								
HEV	.25**	.20	.32**	–0.87	.23**	.19*	.26**	–0.51
EII	.39**	.44**	.36**	0.61	.36**	.40**	.32**	0.74
CESI	.44**	.54**	.35**	1.61	.42**	.50**	.35**	1.36
N	183	88	95		223	107	116	
Parallel MMPI–2 composite criteria with								
DEPI/S – Con	.68**	.76**	.53*	1.41	.67**	.74**	.54**	1.32
SCZI	.55**	.61**	.46*	0.74	.54**	.59**	.45*	0.74
HVI total	.49**	.46**	.57**	–0.56	.47**	.42**	.54**	–0.58
N	57	34	23		67	40	27	

Note. EA = European American; AA = African American; NEA = Non-European American; diffz

spurious result of sampling error, they would indicate that the Rorschach is more valid for minorities than for European Americans (see Table 5), which is the opposite of what one would expect to see with a culturally or ethnically biased test.

For 13 of the 17 predictors, there is no evidence of intercept bias. However, all four of the psychotic disorder predictors show a moderate level of intercept bias. In every instance, the Rorschach underestimates the likelihood of psychosis in the minority sample. Thus, for a given test score (e.g., X – % = .30; SCZI = 5), European Americans are less likely to have been assigned a psychotic disorder diagno-

Table 7 presents the coefficients of factor similarity across the European American and non-European American samples. As can be seen, there is substantial correspondence across the majority and minority samples. The varimax rotated solutions all produce coefficients above .85, except for the third factor in the three-factor solution, where the association is a less-than-desirable .74. These associations do not provide a direct test of factor comparability. However, the chi-square analyses reported in the last two columns statistically evaluate whether the factor loadings dif-

evaluated in one setting. The simple association between 188 Comprehensive System scores and ethnicity initially seemed to support Frank's (1992) hypothesis that minorities would be less engaged with the Rorschach because they produced fewer responses and a higher proportion of pure form responses. However, once patients were matched on education, impatient status, gender, marital status, and age, these differences disappeared. The latter is consistent with a larger body of research conducted on MMPI ethnic differences (Greene, 2000). Seeming racial differences evaporate when researchers control for relevant demographic factors. In this study, after matching patients on key variables or statistically controlling for them, ethnicity was not associated with 188 Comprehensive System scores at a level beyond chance. The average ethnicity–test score association was near zero and no associations exceeded $r = |.23|$.

Scale 6, Bizarre Mentation, Psychoticism) showed the same pattern of findings in the regression equations as the Rorschach scores. There was no evidence of slope bias or differential validity, but the intercepts differed significantly for the majority and minority samples. As with the Rorschach, the direction of difference was such that the MMPI–2 regression lines derived from the European Americans underestimated psychotic disorders for the minorities. Furthermore, the amount of underprediction was quite similar (i.e., partial *r*s for MMPI–2 scales ranged from .185 to .279; for Rorschach scales in Table 6 they ranged from .202 to .264).

Assuming that the psychotic disorder diagnoses accurately reflect patient symptomatology and not simply bias on the part of the clinicians who assigned the diagnoses, the intercept differences would disappear if appropriate covariates were entered into the regression equations. For instance, if a particular route of referral or particular funding source (e.g., Medicaid) was associated both with psychotic disorders and with the Comprehensive System predictor scores, these factors could be covaried to reduce or remove the intercept differences (see Anastasi & Urbina, 1997; Linn & Werts, 1971; Schmidt et al., 1980).

Because the convergent validity and regression analyses could only examine a limited number of reasonable hypotheses in this data set, the final analyses indirectly tested for bias across a broader range of scores. The Comprehensive System's internal structure was evaluated using 59 variables that were both psychometrically suitable for analysis and central to interpretation. Across ethnic groups, the factor solutions did not differ in the pattern or size of the variable loadings. These results are consistent with the regression analyses and they strongly argue against the prospect of slope bias or differential validity for this broader collection of 59 variables central to interpretation (Humphreys & Taber, 1973; Jensen, 1980).

There are several potential limitations to this study. First, the ethnic groups that formed the central contrast in these analyses likely differed on salient variables that were not measured and could not be controlled (e.g., occupation, referral source). Second, theo4.4pationJmo4.4pyon on broation,o4atiooobeenation,whoon sy

duce the pattern of intercept bias that was observed for both the Rorschach and the MMPI scales. It would also suggest that both tests may have been more accurate than the diagnoses assigned by clinicians.[7] Furthermore, although this study used an ecologically valid clinical sample, the findings may differ in other groups. For instance, the clinical symptomatology in these patients may have been sufficiently pronounced to "wash out" more subtle ethnic differences. This could be evaluated through additional research using nonpatient samples.

Although the findings from this study need to be replicated in other samples and settings, researchers and clinicians should be clear that the available data clearly support the cross-ethnic use of the Comprehensive System. The evidence does not support the argument that it is questionable to use the Comprehensive System with minorities (e.g., Lilienfeld, Wood, & Garb, 2000; Wood & Lilienfeld, 1999), much less the stronger and unqualified assertions that "the Comprehensive System should not be used to evaluate members of American minority groups or individuals from outside of the United States" (Garb, Wood, Nezworski, Grove, & Stejskal, 2001, p. 437).

Given that an absence of bias is the typical finding across well designed studies in the personality and cognitive testing literature (e.g., Greene, 2000; Kline & Lachar, 1992; McNulty et al., 1997; Neisser et al., 1996; Timbrook & Graham, 1994), clinicians and researchers should expect the findings from this study to generalize to other samples. In other words, an absence of ethnic bias should be the default expectation. Researchers and clinicians should modify this presumption only when strong data from well-designed studies indicate otherwise.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

Aposhian, M. A. (1995). Iranians and the Rorschach: A question of cross-cultural validity (Doctoral dissertation, The Wright Institute, 1994). *Dissertation Abstracts International, 55, 3633B.*

Baca, B. J. (1994). A comparison of selected Rorschach variables of Hispanic pedophiles and Hispanic normals with each other, and with nonpatient normals and character disorders: Some issues of acculturation (Doctoral dissertation, The Union Institute, 1993). *Dissertation Abstracts International, 54, 5934B.*

Burns, B., & Viglione, D. J. (1996). The Rorschach Human Experience variable, interpersonal relatedness, and object representation in nonpatients. *Psychological Assessment, 8,* 92–99.

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

[7] I thank the two reviewers who suggested this possibility and the next.

Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment, 11,* 297–302.

Exner, J. E., Jr. (1986). *The Rorschach: A comprehensive system* (Vol. 1, 2nd ed.). New York: Wiley.

Exner, J. E., Jr. (1991). *The Rorschach: A Comprehensive System* (Vol. 2, 2nd ed.). New York: Wiley.

Exner, J. E., Jr. (1993). *The Rorschach: A Comprehensive System* (Vol. 1, 3rd ed.). New York: Wiley.

Frank, G. (1992). The response of African Americans to the Rorschach: A review of the literature. *Journal of Personality Assessment, 59,* 317–325

Frank, G. (1993). The use of the Rorschach with Hispanic Americans. *Psychological Reports, 72,* 276–278.

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Towards a resolution of the Rorschach controversy. *Psychological Assessment, 13,* 433–448.

Glass, M. H., Bieber, S. L., & Tkachuk, M. J. (1996). Personality styles and dynamics of Alaska native and nonnative incarcerated men. *Journal of Personality Assessment, 66,* 583–603.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377–393.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68,* 532–560.

Greene, R. L. (2000) *The MMPI–2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265–275.

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,* 186–197.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart & Winston.

Humphreys, L. G., & Taber, T. (1973). Ability factors as a function of advantaged and disadvantaged groups. *Journal of Educational Measurement, 10,* 107–115.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86,* 721–735.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jørgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment, 7,* 259–280.

Kline, R. B., & Lachar, D. (1992). Evaluation of age, sex, and race bias in the Personality Inventory for Children (PIC). *Psychological Assessment, 4,* 333–339.

Krall, V., Sachs, H., Lazar, B., Rayson, B., Crowe, G., Novar, L., & O'Connell, L. (1983). Rorschach norms for inner city children. *Journal of Personality Assessment, 47,* 155–157.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1,* 27–66.

Linn, R. L., & Werts, C. E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement, 8,* 1–4.

McNulty, J. L., Graham, J. R., Ben-Porath, Y., & 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.

Meyer, G. J. (1999). The convergent validity of MMPI and Rorschach scales: An extension using profile scores to define response-character styles on both methods and a re-examination of simple Rorschach response frequency. *Journal of Personality Assessment, 72,* 1–35.

Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment, 75,* 46–81.

Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165.

Meyer, G. J., & Resnick, J. (1996, July). *Assessing ego impairment: Do scoring procedures make a difference?* Paper presented at the Fifteenth International Congress of Rorschach and Projective Methods, Boston.

Meyer, G. J., Riethmiller, R. J., Brooks, G. D., Benoit, W. A., & Handler, L. (2000). A replication of Rorschach and MMPI–2 convergent validity. *Journal of Personality Assessment, 74*, 175–215.

Meyer, G. J., Viglione, D. J., & Exner, J. E., Jr. (2001). On the superiority of *Form%* over Lambda for research on the Rorschach Comprehensive System. *Journal of Personality Assessment, 76*, 68–75.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

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

Perry, W., & Viglione, D. J. (1991). The Ego Impairment Index as a predictor of outcome in melancholic depressed patients treated with tricyclic antidepressants. *Journal of Personality Assessment, 56*, 487–501.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*