JOURNAL OF PERSONALITY ASSESSMENT, *78*(1), 104–129

Exploring Possible Ethnic Differences and Bias in the Rorschach Comprehensive System

Gregory J. Meyer
Department of Psychology
University of Alaska, Anchorage

In this study I conducted a series of analyses to explore potential ethnic bias in Rorschach Comprehensive System variables using a consecutive series of 432 patients evaluated in 1 setting. The simple association between 188 scores and ethnicity revealed no significant findings after matching on several salient demographic variables. Next, 17 analyses of convergent validity found no evidence for differential validity and no evidence for slope bias. For 13 analyses, there was also no evidence for intercept bias. However, with 4 variables predicting psychotic disorders, regression lines favored minorities and worked against European Americans. These findings are the opposite of what should be seen if the Rorschach was biased against minorities and most likely emerged for statistical reasons related to unmeasured confounds. Finally, principal components analyses revealed no evidence of ethnic bias in the Rorschach's internal structure. Although these findings need to be replicated, the available data support using the Comprehensive System across ethnic groups.

Test bias can be defined as systematic measurement error that differentially affects a certain group of individuals (Anastasi & Urbina, 1997; Jensen, 1980; Nunnally & Bernstein, 1994). As such, bias exists when irrelevant factors influence test scores and cause them to be less valid for some people than others. In the context of ethnicity, bias occurs when test results are less valid or accurate for a minority group relative to the majority group.

Recently, Wood and Lilienfeld (1999) asserted that "Blacks, Hispanics, Native Americans, and non-Americans score differently on important Rorschach variables for both the Comprehensive System and other approaches" (p. 342). Although they did not present any specific data concerning differences for

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 should not be used with U.S. minorities.

Regardless, Glass et al. (1996) also compared 47 incarcerated Alaskan Native men (from seven distinct cultural groups within Alaska) to 22 nonnative men. The samples were collected specifically for this study by three examiners using volunteers from two correctional facilities (no information was provided about ethnic variation by facility). A strength of Glass et al.'s study was their delineation of explicit hypotheses for ethnic differences in Rorschach scores. Out of 16 hypothesized differences, 3 were supported and 13 were not. The authors found other unexpected group differences, but because they were unanticipated and based on small samples, it is unclear if they will replicate in future research. In a final analysis, Glass et al. focused on just the Alaska Natives and tested an acculturation hypothesis. In contrast to expectation, there were no differences in Comprehensive System scores across levels of acculturation.

Thus, on closer examination, the studies cited by Wood and Lilienfeld (1999) do not appear to support the claim that Comprehensive System scores manifest important ethnic differences. All used a design that in other contexts Wood et al. (1999) said was flawed and inadequate. The single study that also compared minority and majority samples from within a similar setting found equivocal results.

Given the absence of sound evidence, there is certainly a need for more definitive research on possible ethnic bias with the Rorschach. Indeed, Frank (1993) recently reviewed potential Rorschach differences with Hispanic Americans.[1] His search of the literature revealed only three published studies, all of which were conducted between 1945 and 1956. Given how dated the findings were, Frank drew no conclusions about the presence or absence of differences for this ethnic group. Frank (1992) also conducted a review on African Americans. He found just seven studies published in the 60-year period from 1930 to 1990. All examined elementary, high school, or college students and only one used the Comprehensive System. Across studies, Frank found some support for one seemingly consistent difference: African Americans tended to give fewer responses than European Americans.

Based on the available research, Frank (1992) reasoned that lower *R* in the African American samples was probably not due to an intrinsic psychological factor, like lower intelligence or greater depression. Instead, he proposed that African Americans produced fewer responses because they were less likely to self-disclose in an assessment context. He saw this as part of a general response style in which African Americans were less forthcoming with strangers, similar to a pattern of limited engagement observed in research on self-disclosure during clinical interviews and psychotherapy. Unfortunately, all of the research cited by Frank was conducted many years ago. His Rorschach studies had an average publication date of 1957, whereas his 16 treatment-related studies had an average publication date of 1970. Thus, it is possible his findings reflected social dynamics that were present in an earlier era of heightened U.S. racial tension, although they may be less evident today as a result of societal changes. Nonetheless, with the exception of the isolated finding related to less engagement, Frank did not believe there were meaningful ethnic differences in personality or in Rorschach scores.

Neither Frank (1992, 1993) nor Wood and Lilienfeld (1999) cited the largest study on ethnic differences in Comprehensive System scores. Exner (1986, p. 257) compared 498 European Americans to 102 non-European Americans. The minority sample produced more color responses ($p < .02$), which also created a difference in WSumC ($p < .05$). Exner's analysis of these 600 protocols did not support Frank's (1993) hypothesis of differences in *R*. Furthermore, the small magnitude difference[2] Exner found for color responses did not emerge as a salient difference in Frank's (1992, 1993) reviews. Thus, there does not appear to be any Rorschach score, from the Comprehensive System or other approaches, that consistently differs across ethnic background.

[1]Following Hall et al. (1999), ethnic groups will be identified by terms indicating historical origins (i.e., African American, Asian American, European American, Hispanic American).

[2]Based on Exner's *N* and reported significance levels, the association between ethnicity and color would be approximately $r = .10$, according to Rosenthal's (1991) formulas.

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 groups. For instance, sickle cell anemia is a recessive genetic disorder that is almost exclusively restricted to people of African descent. A blood test that consistently found higher "sickle cell scores" in African Americans than European Americans would not demonstrate bias. On the contrary, such a pattern of scores is required if the test is to validly measure the condition. Analogously, a scale that indicates men are heavier than women would not necessarily indicate gender bias because men *are* heavier than women (e.g., Meyer et al., 2001).

Thus, sophisticated investigations of test bias examine two more specific forms of error known as slope bias and intercept bias (Anastasi & Urbina, 1997; Jensen, 1980; Nunnally & Bernstein, 1994). These issues are investigated with regression equations that evaluate the validity of a test score against an appropriate criterion. Slope bias is present when the validity coefficient for one group differs from the validity coefficient for another group. More specifically, ethnic bias would be evident when a validity coefficient is substantially lower for a minority group than for the majority group. Intercept bias can be present with or without slope bias. Although intercept bias is less critical than slope bias, it refers to instances when a test systematically underpredicts or overpredicts criterion performance for one group. Specifically, intercept bias would be present when a fixed test score underestimates the mental health of minorities and overestimates the mental health of majority group members (or conversely, when it overestimates pathology in minorities and underestimates it in the majority group). To date, no studies have investigated slope bias or intercept bias in Rorschach scores.

Because external validation criteria are required to investigate slope and intercept bias, it is generally impossible for a single study to examine these issues for all the scores on a multi-scale measure like the Comprehensive System. Most studies do not have the resources to collect the array of criteria required to appropriately examine the validity of all test scores. As an alternative, one can investigate the internal characteristics of a test and look for differences in the test's factor structure across ethnic groups. Jensen (1980) advocated for factor analysis as a means of exploring ethnic bias and noted how it "can be used to detect predictive bias indirectly, 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 this term, rules of thumb have been developed for the Minnesota Multiphasic Personality Inventory (MMPI; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989).). Schinka, LaLone, and Greene (1998) and Schinka, Elkins, and Archer (1998) defined a clinically important ethnic difference as one that accounted for 10% of the variance in a test scale when other demographic covariates were not considered or 5% of the variance after controlling for other demographic variables. Translating these figures into alternative measures of effect size produces correlations of .3162 and .2236, respectively, and Cohen's d values of .6666 and .4588, respectively. Somewhat similarly, Greene (2000); McNulty, Graham, Ben-Porath, and Stein (1997); and Timbrook and Graham (1994) suggested that ethnicity was a clinically important predictor when there was a difference of at least 5 T-score points across ethnic groups. This criterion is equivalent to Cohen's $d = .50$ and to $r = .2425$ (when equal base rates are assumed; see Rosenthal, 1991). Thus, the guidelines that have been proposed for identifying clinically important effects of ethnicity on the MMPI are roughly $r = .25$ or $d = .50$. These guidelines will be used to frame the results of this study.

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 factors. With respect to the latter, R and Form% are strongly associated with the richness or complexity of a Rorschach protocol. Indeed, research has shown that approximately 25% of the variance in Rorschach scores is related to the Rorschach's first principal component (Meyer et al., 2000). Because the Rorschach's first factor is effectively defined by R and Form%, if Frank (1992) is correct and minority patients are less engaged with the Rorschach, it will be important to consider whether ethnic differences persist after controlling for R and Form%. The other moderator variables that will be considered include age, education, gender, and psychiatric inpatient status.

The second goal of this study is to use more sophisticated procedures to search for direct evidence of slope or intercept bias across a range of Rorschach scores. This will be accomplished by the regression method outlined in Nunnally and Bernstein (1994, pp. 357–364). Given that there is no reason to believe Rorschach constructs should have differential validity across ethnic groups, I hypothesized there would be no evidence of slope or intercept bias.

The final goal of this study is to examine the internal structure of Comprehensive System scores for evidence of ethnic bias. This will be accomplished by comparing the factor structure of the variables that are considered to be of primary importance when interpreting Comprehensive System results. The analysis follows Jensen's (1980, pp. 446–450) procedures and tests the null hypothesis of no ethnic differences using his most stringent method (Equation 9.19) that simultaneously tests for differences in the pattern and size of factor loadings.

METHOD

Participants in the Full and Partially Matched Samples

This sample has been used previously, although ethnicity has never been examined before. Details concerning the sample and the reliability of Rorschach scoring can be found in Meyer (1999). Briefly, as part of a hospital-based psychological testing program, 432 patients were evaluated with the Rorschach and had information available on the following demographic variables: age, gender, marital status (coded as married vs. not), education level (coded as 1 = less than 11 years, 2 = 12 years, 3 = 13 to 15 years, 4 = 16 years, and 5 = more than 16 years), setting (coded as psychiatric inpatient vs. not), and self-reported ethnicity.

The ethnic composition of the sample was European American = 242, African American = 157, Hispanic American = 14, Asian American = 16, Native American = 2, and other/mixed = 1. Because most patients were either European American or African American, this ethnic contrast forms a basis for the analyses. However, as in other studies (Schinka, Elkins, et al., 1998; Schinka, LaLone, et al., 1998), European Americans will also be contrasted with a broader group consisting of all non-European Americans.

To control for potential moderators, two options are available: statistical partialling or demographic matching. Partial correlations are advantageous because they use all the members of a data set, which maximizes statistical power. A disadvantage comes from the fact that many Rorschach scores are not normally distributed. Nonnormal distributions do not present a problem if the correlational results are simply considered descriptive of the present sample (see Cohen & Cohen, 1983, pp. 51–52; Hays, 1981, pp. 459–461). However, they can create problems when generating inferences about other samples that may have scores with different distributional properties. Demographic matching is advantageous because once it is completed Rorschach scores can be analyzed with nonparametric measures of association and the results can be readily generalized to other samples. The disadvantage is that matching uses just a subset of the available patients. To contend with these competing strengths and limitations, this study utilized both approaches.

To the extent that ethnicity may be associated with diagnosis in this sample (e.g., as a function of particular referral patterns), psychopathology could influence Rorschach scores under the guise of ethnicity. To evaluate this possibility, patients were dichotomously classified as suffering from a psychotic disorder (i.e., schizophrenia, affective disorder with psychotic features, delusional disorder, brief psychotic disorder, psychotic disorder not otherwise specified, schizotypal personality disorder, or borderline personality disorder) and separately classified as suffering from a depressive disorder (i.e., major depression, bipolar disorder with mixed or depressed features, dysthymia, or depressive disorder not otherwise specified). Patients could meet criteria for both classifications. Diagnoses were obtained from hospital billing records and had been assigned independent of the testing by clinicians before the patient was referred for evaluation (see Meyer, 2000). In addition to psychotic and depressive disorder diagnoses, patients were also classified on a gross scale indicating severity of disturbance. All diagnoses were reliably rated on a five-point continuum of impairment, with higher scores indicating greater severity (see Dawes, 1999; Meyer & Resnick, 1996). For patients who received multiple diagnoses, the final score was determined by the maximum severity assigned to any of their diagnoses. Unfortunately, independent diagnoses were not available for all patients, so analyses using diagnostic information were based on a reduced *n*.

The contextual and demographic information described earlier controlled for a number of potentially salient influences on Rorschach scores. However, socioeconomic 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]

Assessing the Simple Association Between Ethnicity and Comprehensive System Scores

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.

Assessing Slope and Intercept Bias

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 Impairment Index (EII; Perry & Viglione, 1991), and the Conceptual Ego Strength Index (CESI; Meyer & Resnick, 1996). The CESI, which has not been described in the published literature before, uses a range of information from each response. Specifically, it generates four scores for each response that are then divided by *R* and averaged. The four scores (and weighting scheme) are form quality (*o*/+ = 0, *u* = 1, *–/none* = 2), form-dominance (any form dominated determinant = 0, any nonform dominated determinant = 1, any formless determinant = 2; for blends, average the individual elements; for pure form, consider the score missing), primary process content (i.e., An, Bl, Ex, Fd, Fi, Sx, Xy, MOR, AG; none = 0, else = 1 point for each), and cognitive Special Scores (none = 0; each Level 1 score = 1; each Level 2 score, ALOG, or CONTAM = 2).

For the MMPI–2 criterion measures, only those patients who displayed a similar test interaction style on both the MMPI–2 and Rorschach were used (Meyer, 1999). Following previous studies (Meyer, 1999; Meyer et al., 2000), the focus was on composite predictor and criterion scales. A MMPI–2 composite measure of negative affectivity combined Scale 2, Scale 7, Depression, Anxiety, and the Negative Emotionality/Neuroticism scale from the Personality Psychopathology Five (PSY–5). This was used as the criterion for a Rorschach negative affect composite formed from the DEPI and S – CON. An MMPI–2 composite measure of psychotic processes combined Scale 8, Bizarre Mentation, and the Psychoticism scale from the PSY–5. This was used as the criterion for the SCZI. An MMPI–2 composite of interpersonal wariness combined Scale 6, Cynicism, Social Discomfort, and the Inability to Disclose component of the Negative Treatment Indicators Scale. This was used as the criterion for the HVI.

Assessing Bias in Internal Test Structure

To examine ethnic bias in internal structure, Jensen's (1980) method begins by separating the minority and majority samples. Subsequently, principal components analysis is applied to each data set and the relevant number of factors is extracted from each correlation matrix. Coefficients of similarity are then produced by correlating the factor loadings for all variables on each factor in the majority sample with the corresponding loadings for each factor in the minority sample. Although informative, these coefficients do not test for significant differences in factor structure. Consequently, one then obtains the factor loadings for each variable on every factor (e.g., for 15 variables and 3 factors, there would be 45 factor loadings in each sample). After transforming the loadings to Fisher's *Z* coefficients (Z_r), the difference between every Z_r loading from the majority and minority samples are statistically evaluated by means of a chi-square (Jensen's Equation 9.19) that tests the null hypothesis of equal factor loadings.

In an effort to ensure that the correlation matrix in the minority and majority samples was sufficiently large to achieve stable findings (see Gorsuch, 1997; Guadagnoli & Velicer, 1988), I used data from all European Americans and all Non-European Americans. To provide results that would be directly relevant to clinical practice, I initially considered the 67 variables from the lower portion of the Comprehensive System Structural Summary because these scores are given the greatest weight in interpretation (Exner, 1991). Furthermore, the variables in the upper portion of the structural summary form the basis for the scores in the lower section. Because of part–whole relationships, including them in the analysis would have introduced redundancy and colinearity. The only variable from the lower portion of the Structural Summary that was not initially considered was EB Pervasive, which is undefined for many patients. In addition to these 67 variables, the CESI was included as an overall measure of disturbance.

To ensure all variables were suitable for principal components analysis, I examined the distribution for each score and eliminated or revised those that had excessive skew or kurtosis. Curran, West, and Finch (1996) indicated a normal distribution has $sk = 0$ and $Ku = 0$, a moderately nonnormal distribution has $sk = 2.0$ and $Ku = 7.0$, and a severely nonnormal distribution has $sk = 3.0$ and $Ku = 21.0$. Based on these guidelines, two variables (CP and MQnone) with severely nonnormal distributions (i.e., $sk > 5$ and $Ku > 28$) were eliminated. Efforts to substitute a percentage or dichotomous version of the scores were not successful. However, substitutions were effective for five other variables. *PureF%* was used instead of Lambda, Blend% was used instead of the raw number of Blends, and a dichotomous score was used instead of the raw value for the Intellectualization Index (> 4 vs. not), MQ– (> 1 vs. not), and Fr + rF (> 0 vs. not).

Next, the correlation matrix was used as a guide to eliminate redundant variables. WSumC, WSum6, Ma, Mp, FM + m, Adjusted D, es, Adjusted es, and the sum of all shading had high correlations with other scores because they shared part–whole relationships. Consequently, they were eliminated. In addition, Sum6 had a strong correlation with Level 2 Special Scores, so it was replaced by two more differentiated variables that do not suffer from a part–whole relationship: Level 1 Special Scores and a score that combined CONTAM and ALOG. Finally, because R is a linear function of W, D, and Dd, the variable D was eliminated from the matrix to prevent problems of colinearity.

Of the remaining 59 variables, 6 had $Ku > 7$, although no values exceeded 10.8. Nine variables had $sk > 2$, although no values exceeded 2.7. Using Curran et al.'s (1996) guidelines, 11 variables (18.6%) in the final set were affected by moderate departures from nonnormality and no variables suffered from severe nonnormality.

Next, the adequacy of the correlation matrix for factoring was examined in the full sample using the Bartlett Test of Sphericity, $\chi^2 = 26142.39$, $p < .00001$, and the Kaiser–Meyer–Olkin Measure of Sampling Adequacy (= .7666), both of which indicated 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.

RESULTS

Preliminary Analyses

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.

TABLE 1
Demographic, Rorschach, and Diagnostic Differences Between European American, African American, and All Non-European American Patients

	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 as closely as possible on education, inpatient status, gender, marital status, and age. There were 120 European Americans who could be closely matched with 120 African Americans and 147 European Americans who could be closely matched with 147 non-European Americans. Even though Rorschach variables were not considered during the matching process, the matched subsamples no longer produced significant differences on R or Form%. However, differences persisted in the frequency with which minority patients received a psychotic disorder diagnosis. Because patients were not matched on all relevant variables (e.g., intelligence, socioeconomic status) and because diagnostic differences persisted after matching, these samples will be referred to as "partially matched" to prevent the erroneous inference that the groups were equated.

The Simple Association Between Ethnicity and Comprehensive System Scores

Spearman, Pearson, and partial correlations were computed between the 188 Comprehensive System scores and ethnicity. Table 3 summarizes the results contrasting European Americans to African Americans.[6] Findings for the unmatched patients are in the first five data columns, and results for the partially matched subsamples are presented in the last two columns. Negative values indicate that European Americans scored higher on the Comprehensive System variable and positive values indicate that African Americans scored higher.

Three sets of partialled correlations were investigated. In the first column of partialled coefficients (i.e., Partial 1), the effects of R and Form% were statistically controlled. Next, the additional effects of education, inpatient status, gender, and marital status were removed (Partial 2). Finally, the additional impact of a psychotic disorder diagnosis was controlled (Partial 3), using the reduced sample of patients who had diagnostic information available. The primary result of partialling psychotic disorders was to reduce the impact of ethnicity on cognitive special scores, form quality, and the SCZI (e.g., the association between ethnicity and the WSum6 dropped from .10 to –.00 when moving from Partialled 2 to Partialled 3).

The results from the partially matched sample differ from the partialled findings in the full sample in that there were no controls for R, Form%, or psychotic diagnoses applied to the partially matched patients. In addition, only 185 scores were considered because three scores (MQ+, CFB, and OBS Criterion 5) never occurred in this subsample.

[6]A table containing specific results for each score is available from Gregory J. Meyer.

TABLE 2
Demographic, Rorschach, and Diagnostic Differences for the Subsamples Closely Matched on Education, Inpatient Status, Gender, Marriage, and Age

	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

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.

[a]n = 120 for EA versus AA; n = 147 for EA versus NEA. [b]n = 88 for EA versus 95 for AA; n = 107 for EA versus 116 for NEA.

**$p < .01$.

TABLE 3
Summary of Effect Sizes Indicating the Association of European American Versus African American Ethnicity With 188 Comprehensive System Variables in the Unmatched Sample[a] and the Subsample Partially Matched on Demographic Variables[b]

	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.

Differential Convergent Validity and Assessment of Potential Slope and Intercept Bias

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-

TABLE 4
Summary of Effect Sizes Indicating the Association of European American Versus Non-European American Ethnicity With 188 Comprehensive System Variables in the Full Sample[a] and the Subsample Partially Matched on Demographic Variables[b]

	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
$r < -.15$	21.3	23.9	0.5	0.5	2.7	5.3	4.8
% of variables with a statistically significant association with ethnicity[c]	15.4	14.9	1.1	0.0	0.0	1.1	0.0

Note. Based on scale where European American = 0, Non-European American = 1; Negative correlations indicate European Americans obtained higher scores; r_s = Spearman rank order correlation; r = Pearson correlation; Partialed 1 = coefficients after controlling for R and Form% (except when those variables were the predictors); Partialed 2 = coefficients after controlling for the prior variables and education level, inpatient status, gender, and marital status; Partialed 3 = coefficients after controlling for the prior variables and psychotic diagnosis ($n = 327$).

[a]$n = 432$. [b]$n = 294$. [c]Bonferroni corrections accounted for the number of statistical tests in each column, not the full table.

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

TABLE 5
Simple Tests for Differential Convergent Validity in the Partially Matched Samples Using Education, Diagnostic, and MMPI–2 Information As Criterion Measures

	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; z_{diff} = *z* value testing for ethnic differences in the magnitude of validity correlations (positive values indicate higher validity in the majority sample; negative values indicate higher validity in the minority sample). No z_{diff} values are statistically significant ($p < .05$).
*$p < .05$. **$p < .01$.

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 diagnosis, whereas minorities are more likely to have been assigned the diagnosis. These results suggest that the Rorschach *underpathologizes* minorities. Again, this is the opposite of what one would expect to see with a culturally or ethnically biased test. However, these findings are likely to be artifactual. As indicated in Table 2, the demographic matching procedure did not equate the European and minority samples. Instead, after matching on age, gender, marital status, and inpatient status, the minority samples continued to have a significantly higher proportion of patients with psychotic disorder diagnoses.

TABLE 6
Tests for Differential Slope and Intercept Bias in the Partially Matched Samples Using Education, Diagnostic, and MMPI–2 Information As Criterion Measures

	Partially Matched EA Versus AA[a]				*Partially Matched EA Versus NEA*[b]			
	Slope Bias		*Intercept Bias*		*Slope Bias*		*Intercept Bias*	
Criterion With Predictor	*r(part)*	*p*	*r(part)*	*p*	*r(part)*	*p*	*r(part)*	*p*
Education level with								
Zf%	.107	.0945	.012	.8543	.086	.1367	.010	.8646
Blend%	.132	.0367	.035	.5776	.117	.0406	.036	.5299
DQ + %	.055	.3932	.000	.9738	.061	.2943	.000	.9964
EA%	.022	.7259	.009	.8831	.010	.8622	.012	.8307
M%	.000	.9895	.003	.9675	.038	.5021	.003	.9659
Intell%	.061	.3364	.037	.5582	.035	.5432	.032	.5729
Psychotic disorder with								
SCZI	.008	.9003	.234*	.0006	.006	.9149	.209*	.0007
X + %	.064	.3615	.250*	.0004	.080	.2093	.229*	.0004
X – %	.006	.9273	.228*	.0013	.015	.8151	.202	.0016
WSum6	.036	.5886	.264*	.0001	.044	.4744	.231*	.0002
Depressive disorder with								
DEPI	.094	.2064	.122	.1015	.031	.6447	.105	.1189
Diagnostic severity with								
HEV	.109	.1306	.101	.1628	.060	.3626	.089	.1782
EII	.038	.5758	.094	.1713	.000	.9775	.081	.2004
CESI	.051	.4455	.078	.2405	.039	.5229	.078	.2005
Parallel MMPI–2 composite criteria with								
DEPI/S – Con	.173	.0802	.119	.2357	.162	.0843	.137	.1498
SCZI	.082	.4714	.069	.5429	.058	.5924	.060	.5816
HVI Total	.045	.7081	.104	.3797	.078	.4987	.089	.4343

Note. EA = European American; AA = African American; NEA = Non-European American.
[a]$n = 240$ for education criterion; $n = 183$ for diagnostic criteria; $n = 57$ for MMPI–2 criteria. [b]$n = 294$ for education criterion; $n = 223$ for diagnostic criteria; $n = 67$ for MMPI–2 criteria.
*Statistically significant following Bonferroni correction to adjust for the 34 statistical tests in each comparison (there are a total of 68 statistical tests in this table).

Assessment of Potential Bias in Internal Test Structure

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 differ across ethnic groups at a level beyond chance. As can be seen, there are no statistically significant differences for any of the seven overall factor solutions. If each of the 19 factor comparisons are tested individually, the third factor in the three-factor varimax rotated solution does differ using a traditional significance level. However, this comparison is not significant when alpha is adjusted to account for the 19 statistical tests. More important, there was no evidence of factor differences for any of the individual factors in the two-factor solution or in the more differentiated four-factor solution. Thus, considering all seven of the overall factor solutions or the individual factors within the differentiated four-factor solution, there is no evidence to suggest that the internal structure of the Rorschach is ethnically biased.

DISCUSSION

This study conducted a series of analyses to explore potential ethnic bias in Rorschach Comprehensive System variables using a consecutive series of 432 patients

TABLE 7
Examining Potential Bias in the Rorschach's Internal Structure Across European American and Non-European American Samples

	Correspondence Across Majority and Minority Samples					
Factor Solution	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	χ^2	*p*
1 factor no rotation	.89				3.909	.096
2 factors no rotation	.89	.88			3.291	.139
Varimax	.91	.89			3.049	.162
3 factors no rotation	.89	.88	.75		3.581	.117
Varimax	.94	.87	.74[a]		4.268	.078
4 factors no rotation	.89	.88	.81	.75	3.443	.127
Varimax	.95	.93	.86	.87	2.878	.180

[a] If this factor is tested individually, it shows a statistically significant difference ($p < .05$) across samples.

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 $r = |.23|$.

Next, a series of analyses compared the validity of Rorschach scores within samples of European Americans, African Americans, and non-European Americans. Across 17 predictor–criterion pairs, there was no evidence to indicate the Rorschach was more valid for one group than another.

For the same 17 predictor–criterion pairs, regression analyses tested for slope or intercept bias. There was again no evidence of differential validity (i.e., slope bias) in these analyses. However, for the four Comprehensive System variables that predicted psychotic disorder diagnoses (i.e., SCZI, X + %, X – %, and WSum6), the majority and minority samples had distinct intercepts. This pattern of intercept bias but no slope bias has been observed fairly often with cognitive tests (e.g., Anastasi & Urbina, 1997; Hunter, Schmidt, & Hunter, 1979; Neisser et al., 1996; Schmidt, Pearlman, & Hunter, 1980). More important, as with other tests, the intercept differences observed in this study created a form of bias that *favored* minorities and worked against the European Americans. Specifically, if the majority regression equations were used to predict psychotic disorders for minority patients, they would consistently underestimate this type of problem. As such, the findings were directly opposite the pattern that should be seen if the Rorschach was biased against minorities.

Why did the intercept differences emerge for psychotic disorders in this study? The most important factor that would explain these findings is the presence of an unmeasured and uncontrolled third variable (see Anastasi & Urbina, 1997; Linn & Werts, 1971; Schmidt et al., 1980). Mathematically, intercept bias that favors a minority group, like that seen in this study, occurs if the majority and minority groups differ on one or more additional variables that are themselves correlated with both the predictor and criterion. As demonstrated in Table 2, despite matching on some salient demographic factors, the minority samples still had a higher rate of psychotic disorder diagnoses. In the absence of slope bias, this fosters the pattern seen in Table 6. For every level of the Rorschach predictors, the rate of psychotic disorders is higher in the minority than the majority groups. More important, a supplemental analysis of MMPI–2 scales assessing psychotic symptoms (i.e., Scale 8, 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, the diagnoses that were used in the bias analyses had been assigned by treating clinicians prior to any testing. Although this ensured they were not confounded by Rorschach results, it also means that they were assigned by clinicians who had sufficient lingering questions about their patients to refer them for an evaluation. Relatedly, no information is available concerning the reliability of these diagnoses and it is possible that clinicians assigned diagnoses in an ethnically biased manner. This factor could help explain why psychotic diagnoses continued to be more prevalent for minorities after matching on several demographic parameters (Tables 1 and 2) and could be related to the intercept bias observed for psychosis (Table 6). For instance, clinicians may have been less willing to record a psychotic disorder diagnosis for European Americans, regardless of their presenting symptoms. Alternatively, because the diagnoses used in this study had been obtained from billing records, it is possible reimbursement concerns resulted in psychotic diagnoses being assigned more liberally to minority patients. Either of these factors would produce 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.

REFERENCES

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* (Rev.). Newbury Park, CA: Sage.
Sanchez, E. M. (1993). Affective scoring of native American female Rorschach protocols (Doctoral dissertation, Miami Institute of Psychology of the Caribbean Center for Advanced Studies, 1992). *Dissertation Abstracts International, 53, 6027B*.
Sangro, F. M. (1997). Location tables, form quality, and popular responses in a Spanish sample of 470 subjects. *Rorschachiana, 22*, 38–66.
Schinka, J. A., Elkins, D. E., & Archer, R. P. (1998). Effects of psychopathology and demographic characteristics on MMPI–A scale scores. *Journal of Personality Assessment, 71*, 295–305.
Schinka, J. A., LaLone, L., & Greene, R. L. (1998). Effects of psychopathology and demographic characteristics on MMPI–2 scale scores. *Journal of Personality Assessment, 70*, 197–211.
Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Personnel Psychology, 33*, 705–724.
Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI? *Psychological Assessment, 6*, 212–217.
Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment, 6*, 341–349.
Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). *Assessment, 6*, 115–129.
Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432–442.

Gregory J. Meyer
Department of Psychology
University of Alaska Anchorage
3211 Providence Drive
Anchorage, AK 99508
E-mail: afgjm@uaa.alaska.edu

Received January 22, 2001
Revised May 30, 2001