JOURNAL OF PERSONALITY ASSESSMENT, *74*(2), 175–215

ARTICLES

A Replication of Rorschach and MMPI–2 Convergent Validity

Gregory J. Meyer
Department of Psychology
University of Alaska Anchorage

Robert J. Riethmiller, Regina D. Brooks,
William A. Benoit, and Leonard Handler
Department of Psychology
University of Tennessee

We replicated prior research on Rorschach and MMPI–2 convergent validity by testing 8 hypotheses in a new sample of patients. We also extended prior research by developing criteria to include more patients and by applying the same procedures to 2 self-report tests: the MMPI–2 and the MCMI–II. Results supported our hypotheses and paralleled the prior findings. Furthermore, 3 different tests for methodological artifacts could not account for the results. Thus, the convergence of Rorschach and MMPI–2 constructs seems to be partially a function of how patients interact with the tests. When patients approach each test with a similar style, conceptually aligned constructs tend to correlate. Although this result is less robust, when patients approach each test in an opposing manner, conceptually aligned constructs tend to be negatively correlated. When test interaction styles are ignored, MMPI–2 and Rorschach constructs tend to be uncorrelated, unless a sample just happens to possess a correlation between Rorschach and MMPI–2 stylistic variables. Remaining ambiguities and suggestions for further advances are discussed.

Recently, Meyer (1997b, 1999a) hypothesized that the propensity for patients to be open, spontaneous, and engaged during the administration of the Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), and the Rorschach (Exner, 1993) played a role in the convergent validity of scales derived from these two instruments. Drawing on evidence from a large sample of psychiatric patients, results supported this hypothesis and suggested that test interaction styles were important moderators of convergent validity. In the prior research, when test interaction styles were ignored, there was virtually no association between those Rorschach and MMPI–2 scales that shared similar names (mean $r = .03$). When analyses were restricted to the subset of patients who interacted with both tasks in a similar fashion, the same Rorschach and MMPI–2 scales were substantially correlated (rs ≈ .40–.60). Although more equivocal, results also suggested that patients who interacted with the Rorschach in one manner (e.g., defensive constriction) and interacted with the MMPI–2 in an opposing manner (e.g., exaggeration–dilation) produced negative correlations among these scales (rs ≈ –.20 to –.50).

To date, two different procedures have been used to define what we term *test interaction styles* or *response–character styles.* The primary procedure has been based on scales designed to measure the first principal component from each test; the other has been based on traditional indicators of test-taking style (i.e., using *F* and *K* from the MMPI–2 and *R* and *Lambda* from the Rorschach; see Meyer, 1999a, Table 1). Although these two sets of criteria select substantially different patients for analyses, similar results were obtained using both.

This study attempted to replicate these Rorschach and MMPI–2 associations with an independent sample of patients. We followed the same procedures used previously and tested eight hypotheses derived from the prior findings. First, we expected scales of response–character styles on the MMPI–2 to be uncorrelated with scales of response–character styles on the Rorschach. Clinically, this hypothesis reflects our expectation that the way patients interact with one task should have no bearing on their style of interacting with the other. Second, we expected conceptually related Rorschach and MMPI–2 scales to be uncorrelated when response styles were ignored. Clinically, this reflects our expectation that each test would generally provide distinct information that could not be obtained directly from the other. Third, we expected conceptually related Rorschach and MMPI–2 scales to be positively correlated when analyses were limited to those patients who had similar test interaction styles on both methods. Clinically, this hypothesis postulates a subset of patients who tend to obtain similarly elevated scores on the Rorschach and MMPI–2 constructs. In part, this convergence is believed to result from congruent styles of interacting with each test, although these interaction styles are viewed as emerging from two very different sources and may reflect patients' genuine characterological qualities, their deliberate efforts to manipulate the tests, or both (see Meyer, 1997b, 1999a). Fourth, although more tentative, we expected conceptually related MMPI–2 and Rorschach scales to be negatively correlated when analyses were limited to those patients who had an opposing response–character style on each method. Clinically, this hypothesis postulates a subset of patients who obtain fairly different indications of pathology and health across these two types of tests. As before, we expected that test interaction styles (emerging either from the patients' character structure or from deliberate efforts to manipulate the tests) contributed to these opposing impressions.

The remaining hypotheses were designed to challenge or solidify the core expectations outlined previously. For our fifth hypothesis, we expected conceptually unrelated MMPI–2 and Rorschach scales to remain uncorrelated across all analyses (i.e., in the full sample, in the analyses limited to patients with similar styles, and in the analyses limited to patients with opposing styles). Sixth, using a correction formula, we expected that our observed validity correlations would be larger than the correlations one could predict to see simply as a function of selecting certain patients for analysis (see Meyer, 1997b, pp. 320–321). Seventh, we expected a multitrait–multimethod matrix to indicate (a) relatively differentiated Rorschach constructs but (b) relatively undifferentiated MMPI–2 constructs that would (c) produce weak evidence for convergent validity relative to discriminant validity (Meyer, 1997b). Eighth, given the general hypothesis that conceptually related Rorschach and MMPI–2 scales should be correlated when response–character styles are correlated, we expected to see an association between construct correlations and first-factor correlations when small samples were repeatedly selected from the full population and the association was examined across samples (see Meyer, 1999a).

In addition, to address recent criticisms (Archer & Krishnamurthy, 1999) we extended prior research in two ways. First, we modified and expanded the criteria to identify test interaction styles, and second, we developed a new strategy to address concerns about selecting particular patients for analysis. Each is discussed in turn.

In prior research, two sets of criteria have been used to define test interaction styles: (a) traditional indicators of test-taking behavior that can be derived from MMPI–2 or Rorschach profiles and (b) scales for assessing the first principal component from each test. As a third set of criteria for this study, we also used actual factor scores derived from principal components analysis of the Rorschach and the MMPI–2 to measure the primary factors. Previous research (e.g., Meyer, 1999a) has been criticized because it used only a small subset of patients in the key analyses (Archer & Krishnamurthy, 1999). Indeed, about 45% of patients have been classified as having similar or opposing styles when using the factor-based scales, and only about 28% have been classified when using the profile criteria. In an effort to expand the number of patients included in the key analyses, we decided to employ a fourth, more liberal set of criteria to identify test interaction styles. Specifically, we examined all those patients who were identified by any of the three criteria sets as having a similar or an opposing test-taking style.

It is unusual to select certain patients for convergent validity analyses. Because these procedures are novel, it is appropriate to approach them with some caution (cf. Archer & Krishnamurthy, 1999). In particular, some may speculate that the selection procedures generate spurious findings. In an effort to address this question more thoroughly, the procedures used for the MMPI–Rorschach convergent validity analyses were also applied to the data from two self-report instruments, the MMPI–2 and the Millon Clinical Multiaxial Inventory–II (MCMI–II; Millon, 1987). Conducting a parallel set of analyses on two instruments from the same method family (i.e., self-report) should provide an important backdrop for considering the cross-method Rorschach and MMPI–2 findings.

METHOD

Participants

The Tennessee sample. Data were obtained from an outpatient clinic in Tennessee that serves both university students and community residents. The initial sample consisted of 472 patients who voluntarily sought treatment between 1991 and 1997. All tests were administered prior to treatment, and the MMPI–2 and Rorschach were invariably administered within 3 weeks of one another. Of the initial 472 patients, 429 completed the MMPI–2, 393 completed the Rorschach, and 350 completed both the Rorschach and the MMPI–2. The 350 patients who completed both instruments are the focus of this study. Data from 23 of these patients were excluded from the analyses; 17 appeared to complete their MMPI–2 in a random fashion (i.e., [F + Fb]/2 T score > 115 and *VRIN T* score > 80), and 6 had Rorschach protocols of questionable utility (i.e., R < 12, or R = 12 or 13 and *Lambda* > .50). For the remaining 327 patients, the average age was 29.9 (*SD* = 9.1, range = 16–73), and 56.6% were female. The sample was diverse socioeconomically and psychologically, with diagnoses that ranged from adjustment disorder to schizophrenia; however, the sample was racially homogenous—more than 95% of the patients were White.

The Chicago sample. This sample was described in detail previously (Meyer, 1997b, 1999a). Several reanalyses of this data set were undertaken as part of this study to ensure that equivalent results were being compared across both samples. Briefly, as part of a hospital-based psychological testing program, 362 patients completed a valid MMPI–2 and a valid Rorschach. For the analyses using two self-report instruments, data were obtained from 269 patients who completed a valid MCMI–II and a valid MMPI–2.

Measures for the Tennessee Sample

Rorschach protocols were administered and scored by doctoral students at the University of Tennessee. At a minimum, these students were in their 2nd year of training and had completed a one-semester course in assessment with concurrent enrollment in an advanced personality assessment course. All students received 3 hours per week of supervision in assessment by various faculty members. All Rorschach protocols were entered into a computerized program for scoring (Exner, Cohen, & McGuire, 1990), and a second computer program was used to read these files into the statistical database.

All MMPI–2s were administered at roughly the same time as the Rorschach. Some of the MMPI–2s were initially computer scored by National Computer Systems (NCS), and some were initially scored by hand. Because the NCS computerized data files were not available, all MMPI–2 items were manually entered into the final database and then scored using computer algorithms.

Reliability of Rorschach scoring. To evaluate the reliability of Rorschach scoring, three raters (Robert J. Riethmiller, Regina D. Brooks, and William A. Benoit) each scored 43 protocols (total *R* = 894). Six of these protocols were taken from Exner (1993), and the remaining 37 were randomly selected from the database. The raters scored independently, although they reviewed areas of disagreement after they completed the six protocols from Rorschach Workshops and also after completing the first 18 protocols from the data set. To assess the key variables of interest in this study, intraclass correlation coefficients (ICCs) were computed. The ICC is a chance-corrected reliability statistic that is equivalent to weighted kappa. Typically, values greater than .74 are considered to indicate excellent reliability, values from .60 to .74 are considered good, values from .40 to .59 are considered fair, and values below .40 are considered poor (Cicchetti, 1994). Following the terminology of McGraw and Wong (1996), we used the Case 2 model and calculated ICC(A,1) with $k = 3$, which means we assumed that the three raters were randomly selected from the population of student coders, and we determined the degree of absolute agreement between a single rater and any other rater. Across all 43 protocols, the ICC results were *R* = 1.0, *Lambda* = .73, *PureF%* = .93, *DEPI* = .71, *S–CON* = .77, *SCZI* = .77, and *HVI* = .72.

Two comments should be made about the preceding figures. First, all the reliability values fall at the upper end of the good classification range or in the excellent range. Thus, they are quite respectable. Nonetheless, they are slightly lower than those that have been reported in the literature (Meyer et al., 1999). This is most likely because the raters were still receiving training in the Comprehensive System.

Second, the ICC for *Lambda* is deceiving because this variable has a skewed distribution. Because *Lambda* is computed as a proportion ($F / [R - F]$), its distribution has an unstable upper tail, in which small differences in *F* can produce large differences in *Lambda*. For instance, if one rater determines that 19 of 20 responses are pure form, then *Lambda* = 19.0. If another rater determines that *F* = 18, then *Lambda* will be 9.0. Although the raters disagree on only one *F* determinant, the first rater produces a *Lambda* value that is 10 points (and many standard deviations) higher than that of the second rater. In this sample, the impact of this phenomenon can be seen by comparing the ICC for *Lambda* to the ICC for *PureF%*. Although these are conceptually equivalent variables ($F/R - F$ and F/R are simply different transformations of the same data and one can be predicted from the other with perfect accuracy), the ICC for *Lambda* was .73, whereas the ICC for *PureF%* was .93. Thus, the ICC computed for *Lambda* appears to underestimate the raters' agreement.

Next, 80 protocols were selected from the archival records and scored by one of the three raters. This served two purposes. First, it allowed us to estimate scoring accuracy in the overall database. Second, it allowed us to identify students who may have contributed poorly scored records. After identifying all of the students who contributed protocols to the archival database, we systematically sampled protocols for rescoring. Specifically, we ensured that at least two protocols were scored for every student who contributed more than five records and that at least one protocol was scored for every student who contributed between three and five records. The following agreement rates were found between the archival scoring and the final scoring: Location and Space (97%), Developmental Quality (92%), Determinants (78%), Form Quality (89%), Pairs (97%), Content (86%), Popular (96%), Organizational Activity (91%), Cognitive Special Scores (89%), and Other Special Scores (91%).[1] Although these percentage agreement figures were respectable (see Meyer, 1997a), we identified seven individuals who had less than optimal scoring, which was defined as a percentage agreement rate less than .75 for more than one segment in any protocol. In total, these individuals contributed 60 records to the database, and each protocol was rescored by Robert J. Riethmiller, Regina D. Brooks, or William A. Benoit.

Measures for the Chicago Sample

Meyer (1997b, 1999a) described the procedures for obtaining, scoring, and calculating reliability on the Rorschachs in this sample. All MMPI–2s or MCMI–IIs

[1]Because the raters only documented exact agreement rates for each segment, kappa values could not be computed for these records; however, if we assume that these 80 protocols form a randomly selected subset of the full sample, the score frequencies from the full sample can be used to estimate segment kappa values using the procedures outlined by Meyer (1999b). Doing so provides the following estimated kappa values: Location and Space (.96), Developmental Quality (.82), Determinants (.72), Form Quality (.82), Pairs (.92), Content (.85), Popular (.88), Organizational Activity (.76), Cognitive Special Scores (.52), and Other Special Scores (.75). Because 60 protocols from the full sample were rescored after identifying students who tended to score poorly, these figures underestimate the reliability of the final data used in the analyses.

were administered at roughly the same time as the Rorschach, and they were computer scored by NCS. Scores for these tests were obtained from the NCS files, except for four patients who only had MMPI–2 profile sheets available.

Defining Test Interaction Styles

As discussed earlier, test interaction styles (or response–character styles) have been defined in the past using two separate sets of criteria: one based on factor-derived scales and the other based on traditional scores readily available from an MMPI–2 or Rorschach summary profile. In this study, we extended previous research by including two additional criteria, one based on actual factor scores for the unrotated principal component from each test and one that combined cases identified by all three sets of criteria. These four methods of classification are described below.

Criteria using factor-based scales. The procedures for defining response–character styles using the first principal components from the Rorschach and the MMPI–2 were discussed in detail by Meyer (1997b). Because Welsh's Anxiety (*A*) scale was designed to quantify the first principal component of the MMPI–2, it has served as the key variable for measuring the first factor of the MMPI–2. Patients were selected for analysis if they fell in the upper or lower third of the *A* scale distribution. To assess its adequacy in this sample, we conducted a principal components analysis of the *A* scale along with the MMPI–2 basic, validity, and content scales. Using the 408 patients with valid MMPI–2s, the first unrotated component accounted for 48.3% of the total variance; the second, third, fourth, and fifth components accounted for 10.4, 5.7, 4.4, and 3.5% of the total variance, respectively. As expected, this dimension was defined by the *A* scale, which had a loading of .93. These findings indicate that *A* is a good measure of the huge first factor of the MMPI–2.

The first unrotated principal component from the Rorschach has been termed *Response–Engagement* (R–Engagement) or *Response–Complexity* (Meyer, 1997b). The formula to compute the R–Engagement scale was derived from a large sample of college students (Meyer, 1992), and it is calculated using *z* scores with the following weights: .436(Color Shading Blends) + .372(*FY*) + .325(*FC´*) + .3(*FC*) + .3(*CF* + *C*) + .29(Shading Blends) + .29(*m*) + .29(*R*) + .27(*S*) + .24(*FM*) + .22(*FV*) + .21(*W*) + .19(*MOR*) + .18(*M*) – .24(*Lambda*). Because the program we used to translate Rorschach scoring files into our statistical database did not calculate shading blends, this variable was omitted from the R–Engagement scale in this sample. Nonetheless, to assess the adequacy of the R–Engagement scale in this sample, we conducted a principal components analysis of this variable along with other nonredundant scores for Location, Developmental Quality, Determinants, Form Quality, and Special Scores. Using the 386 participants with valid Rorschachs, the first unrotated component accounted for 25.9% of the total variance; the second, third, fourth, and fifth components accounted for 8.8, 6.3, 5.6, and 4.7% of the total variance, respectively. As expected, this dimension was most strongly defined by the R–Engagement scale, which had a loading of .93. These findings indicate that R–Engagement is a good measure of the Rorschach's large first factor (even when it is computed without Shading Blends).

R–Engagement has proven to be an excellent marker for the Rorschach's first unrotated principal component in three independent samples; however, many of the variables that contribute to the scale are quite skewed or kurtotic. Skew for rare variables is expected (e.g., Color Shading Blends), but even relatively continuous Rorschach scores can be skewed. For instance, in this sample, not only was the Color Shading Blend distribution somewhat skewed and kurtotic (2.14 and 5.92, respectively), but the *Lambda* distribution was even more nonnormal, having a skew of 6.36 and a kurtosis of 65.11. As indicated above, the *Lambda* distribution is often nonnormal because it is computed as a proportional value where the upper limit to the distribution is constrained only by *R*. In this sample, one participant had a 14-response protocol that contained all pure-form responses, which generated an extremely deviant *Lambda* value that was 12.6 standard deviations from the median. As we described earlier, the problem of *Lambda*'s skew can be easily corrected by calculating a revised *Lambda* index, *PureF%* (i.e., *F*/*R* rather than *F*/*R* – *F*). In this sample, the distribution for *PureF%* was quite normally distributed, with $M = .39$, $Mdn = .37$, skew = .37, and kurtosis = –.14.

Because it is problematic to derive *z* scores from variables with highly skewed distributions and because highly skewed variables can create problems with correlations, which in turn can alter factor-analytic results, it is important to evaluate whether the Rorschach's first unrotated principal component is substantially different when the variables in the analysis are limited to those with minimal skew. To investigate this, all Rorschach variables that had skew greater than |1.99| were eliminated from consideration. Subsequently, the following nonredundant scores for Location, Developmental Quality, Determinants, Form Quality, and summary ratios were factor analyzed: *R, W, Dd, S, DQ+, DQO, M, FM, m, C, CF, FC, FC´, FY,* Pair, Popular, Blends, *PER, COP, Zf, Zd, Afr,* Egocentricity, *A%,* Isolation Index, *X–%, X+%,* and *PureF%*. The first unrotated principal component accounted for 23.2% of the total variance, and the next four factors accounted for 11.1, 9.1, 5.3, and 5.1% of the total variance, respectively. As expected, this factor was defined by *R* (.73), Blends (.81), and *Zf* (.84) on the positive pole and by *PureF%* (–.48) on the negative pole.

Two steps were taken to assess the similarity between this factor and the R–Engagement scale. First, regression-based factor scores for the new factor were correlated with R–Engagement (which was itself normally distributed). The magnitude of the correlation ($r = .92$) indicated a high degree of correspondence.

Next, the R–Engagement scale was entered into a factor analysis along with the normally distributed variables listed above. R–Engagement was the best marker for the underlying factor, having a loading of .94 on this dimension.

Thus, although the initial factor analysis that created the R–Engagement scale (Meyer, 1992) was conducted with skewed variables and although the R–Engagement scale continues to be calculated from skewed variables, the scale itself is normally distributed and is the best marker for the Rorschach's first unrotated principal component—even when the factor input is limited to normally distributed Rorschach variables. Given this, we proceeded with the analysis, employing the original R–Engagement scale.

Conceptually, one end of the first principal component from the Rorschach and the MMPI–2 is characterized by defensive withdrawal, cognitive–emotional simplicity, or denial. The other pole is characterized by excessive engagement, heightened sensitivity, or overreporting of problems (see Meyer, 1999a, Table 1). For the sake of simplicity, these poles are termed *constricted* and *dilated,* respectively.

Following previous conventions, constricted and dilated styles were initially defined by the upper and lower thirds of the R–Engagement scale and the *A* scale. For this sample, these criteria were operationally defined with cut-points of less than 18 and less than 25 for the MMPI–2 *A* scale. As a result, 118 patients were classified as constricted and 120 were classified as dilated (because 4.3% of the sample had scores of 17 and 4.6% had scores of 26, slightly more than one third of the patients fell in each tail). The cut points for the Rorschach R–Engagement scale were less than –.930 and greater than .64, which resulted in 109 patients being classified as constricted and 109 being classified as dilated.

Examining test interaction styles across methods, these criteria identified 80 participants (24.5%) as having similar styles on both the MMPI–2 and Rorschach (i.e., dilated on both [n = 39] or constricted on both [n = 41]), and they identified 76 participants (23.2%) as having discordant or opposing response styles across methods (dilated Rorschach and constricted MMPI–2 [n = 33] or constricted Rorschach and dilated MMPI–2 [n = 43]).

Criteria using profile scores. Following Meyer (1999a), patients were also classified by scales commonly interpreted as indicators of test-taking style. From the Rorschach, *R* and *Lambda* were used for this purpose, and *F* and *K* were used from the MMPI–2. To classify patients as dilated or constricted, we employed the same cutoff values that had been used in the prior research. These values were *R* = 21, *Lambda* = .55, *F T* score = 58, and *K T* score = 50. The median values for this sample were similar (*R* = 20, *Lambda* = .55, *F* = 61, and *K* = 45). As before, patients were classified as constricted on the Rorschach if *R* was less than 21 and *Lambda* was greater than .55. Patients were classified as dilated if *R* was greater than 21 and *Lambda* was less than .55. Parallel determinations were made using the MMPI–2 scales. Dilated patients had *F* greater than 58 and *K* less than 50, and constricted patients had *F* less than 58 and *K* greater than 50.

Examining test interaction styles across methods according to the preceding criteria, 45 patients (13.8%) adopted similar styles on the MMPI–2 and the Rorschach (i.e., dilated on both [*n* = 28] or constricted on both [*n* = 17]), and 48 patients (14.7%) adopted discordant styles across methods (dilated Rorschach and constricted MMPI–2 [*n* = 10] or constricted Rorschach and dilated MMPI–2 [*n* = 38]).

Criteria using factor scores. Although the criteria listed above are the same as those used in prior research, for a variety of reasons, these criteria are imperfect measures of test interaction styles. R–Engagement and *A* have been used to estimate patients' standing on the first unrotated principal component from each test; however, the actual first factors have not been used to identify test interaction styles, although it would be reasonable to do so. To extend the previous research (Meyer, 1997b, 1999a), the factor-analytic results described above were used to generate factor scores to quantify each patient's location on the first unrotated principal component. These scores were then used to form a third set of criteria for selecting test interaction styles. As with R–Engagement and *A*, the MMPI–2 and Rorschach factor score criteria selected patients falling in the upper and lower thirds of each distribution. For this sample, these criteria were operationally defined with MMPI–2 factor score cut points of less than or equal to –.52 and greater than or equal to .53. These cutoffs resulted in 110 patients who were classified as constricted and 110 who were classified as dilated. The cut points for the Rorschach first principal component (derived from the normally distributed variables) were less than or equal to –.48 and greater than or equal to .287. These cutoffs resulted in 109 patients who were classified as constricted and 109 who were classified as dilated.

When we examined test interaction styles across methods, the factor score criteria identified 73 patients (22.3%) as having similar styles on both the MMPI–2 and Rorschach (i.e., dilated on both [*n* = 36] or constricted on both [*n* = 37]), and they identified 71 patients (21.7%) as having discordant or opposing response styles across methods (dilated Rorschach and constricted MMPI–2 [*n* = 33] or constricted Rorschach and dilated MMPI–2 [*n* = 38]).

Criteria using the combination of factor scales, factor scores, and profile scores. To maximize the sample sizes included in our analyses, we also employed a fourth, more liberal set of criteria to identify test interaction styles. Specifically, we looked at all three criteria sets in combination. Any patient identified as having a similar style by factor scales (i.e., R–Engagement and *A*), by factor scores (i.e., regression method estimates of the first principal component), or by profile criteria was considered to have a similar style across methods. Simultaneously, any patient identified as having an opposing style by factor scale criteria, by factor score criteria, or by profile criteria was considered to have an opposing style across methods.

One complication arose in implementing these combined criteria. Two cases had been identified by the profile criteria as having opposing styles and by the factor scale criteria as having similar styles. These two cases were considered ambiguous and were excluded from the subsequent analyses. Ultimately, the combination criteria identified 109 patients (33.3%) as having similar styles on the tests and 109 (33.3%) as having opposing styles. Thus, the combination criteria employed two thirds of the initial sample, which is a substantially larger proportion than has been used in prior research.

Correspondence Among the Criteria for Defining Response–Character Styles

Table 1 reports the extent of agreement, indexed by Cohen's kappa, among the four criteria sets in terms of final classifications. For each set of criteria, the overall results are formed from the conjunction of Rorschach and MMPI–2 data in which patients were classified into one of three categories: (a) similar style on both methods, (b) opposing style across methods, or (c) undefined style on at least one method. The table indicates that there is a fair amount of "play" in the various classification schemes; many patients identified by one set of criteria differ from those identified by another set of criteria. As might be expected, however, the two sets of criteria based on the first principal component from each test (i.e., the factor scales and the

TABLE 1
Overall Classification Agreement (κ) Between the Four Procedures for Classifying Test Interaction Styles on the Rorschach and MMPI–2

	Criteria			
Criteria	*Factor Scales: R–Engagement and A*	*Factor Scores*	*Profile*	*Factor or Profile*
Factor scores	—			
Factor scores	.56	—		
Profile	.27	.31	—	
Factor or profile	.72	.67	.42	—

Note. $N = 327$. MMPI–2 = Minnesota Multiphasic Personality Inventory–2. For each set of criteria, the overall results are formed from the conjunction of Rorschach and MMPI–2 data in which patients were classified as having a similar style on both methods (1), an opposing style across methods (–1), or an undefined style on at least one method (0). All coefficients are statistically significant.

factor scores) showed the highest degree of correspondence (κ = .56). Previously, Meyer (1999a) reported a kappa value of .31 between the factor scale criteria and the profile criteria, which closely matches the result in this study (κ = .27).

Target Constructs

Variables related to three psychological constructs, namely, affective distress, psychotic processes, and interpersonal suspiciousness or wariness, were used to assess convergent validity between the MMPI–2 and the Rorschach. Rorschach measures of emotional distress included the DEPI and the Suicide Constellation (*S-CON*). MMPI–2 variables included Scales *2* and *7,* Depression (*DEP*), Anxiety (*ANX*), and the Negative Emotionality/Neuroticism scale from the Personality Psychopathology Five (*PSY–5–Neg;* Harkness, McNulty, & Ben-Porath, 1995). The Rorschach measure of psychotic processes was the Schizophrenia Index (*SCZI*), and MMPI–2 measures included Scale *8,* Bizarre Mentation (*BIZ*), and the Psychoticism scale from the *PSY–5* (*PSY–5–Psy;* Harkness et al., 1995). The Rorschach measure of interpersonal wariness was the Hypervigilance Index (*HVI*), and the MMPI–2 scales included Scale *6,* Cynicism (*CYN*), Social Discomfort (*SOD*), and the Inability to Disclose component of the Negative Treatment Indicators Scale (*TRT2;* Ben-Porath & Sherwood, 1993). These variables were the same as those used in the prior research.

In addition to examining each of these variables individually, all of the scales targeting a common construct were transformed to *z* scores and then aggregated to form composite measures (for research documenting the importance of aggregation in personality research, see Cheek, 1982; Cole, Howard, & Maxwell, 1981; Cook & Campbell, 1979; Epstein, 1983; Overholser, 1992; Rushton, Brainerd, & Pressley, 1983; Tsujimoto, Hamilton, & Berger, 1990). Thus, the two Rorschach measures of negative affect were aggregated, as were the various MMPI–2 scales targeting each of the three general constructs. Because the Rorschach measured psychosis and wariness with a single scale (i.e., *SCZI* and *HVI*, respectively), only the individual scale was used in the composite analyses.

Total scores were used for Rorschach scales, and non-*K*-corrected raw scores were used for MMPI–2 scales. Rorschach distributions were approximately normal. No scales had skew or kurtosis values greater than |0.66|. Two MMPI–2 variables had values greater than |1.0| (*SOD* kurtosis = –1.06; *BIZ* skew = 1.21 and *BIZ* kurtosis = 1.54), but departures from normality were minimal and all variables were retained for the correlational analysis.

Corrections for Potential Methodological Artifacts

Conceptually unrelated variable pairs (CUVPs). As one way to evaluate the extent of bias that may be introduced into our results from the process of selecting particular patients for analysis, we examined variables that bore little conceptual relationship to each other. Following Meyer (1999a), two sets of unrelated variables were considered. In both sets, each of the Rorschach scores were expected to be uncorrelated with each of the MMPI–2 scales. For the first set, all Rorschach variables correlated positively with the Rorschach's first factor and all MMPI–2 variables correlated positively with the first factor of the MMPI–2. Thirteen Rorschach scores were used: Whole Locations (*W*), Detail Locations (*D*), Ordinary Developmental Quality (*DQo*), Ordinary Form Quality (*FQo*), Animal Movement (*FM*), Pairs, the Sum of All Human Content, Whole Human Content (*H*), Whole Animal Content (*A*), Idiographic Content (*Id*), Passive Movement (*p*), Personalized Responses (*PER*), and Cooperative Movement (*COP*). These variables had an average correlation of .36 with the R–Engagement scale and .43 with the factor scores. These results were fairly close to the average correlations found with the four Rorschach variables used in the meaningful convergent validity analyses (i.e., *DEPI*, *S–CON*, *SCZI*, and *HVI*), which had mean *r*s of .45 with the R–Engagement scale and .44 with the factor scores. These findings closely replicated prior results. Comparable correlations from Meyer (1999a) were .38 and .46 for the CUVPs with R–Engagement and with factor scores, respectively, and .49 and .48 for the meaningful variables with R–Engagement and with factor scores,[2] respectively.

Eleven MMPI–2 variables were also used: Scale *1*, Scale *3*, Somatic Complaints (*Hy4*), Ego Inflation (*Ma4*), Generalized Fearfulness (*FRS1*), Health Concerns (*HEA*), Gastrointestinal Symptoms (*HEA1*), Explosive Behavior (*ANG1*), Addiction Admission (*AAS*), Antisocial Behavior (*ASP2*), and Aggression (*PSY–5–Agg*). These variables had an average correlation of .35 with *A* and .45 with the first-factor scores of the MMPI–2. The average correlations for the 12 variables used in the meaningful convergent analyses (i.e., Scale *2*, Scale *7*, Scale *8*, etc.) were .69 and .74, respectively. These results replicated prior findings. The Chicago sample obtained correlations of .43 and .51 for the CUVPs with *A* and with factor scores, respectively, and .74 and .78 for the meaningful variables with *A* and with factor scores, respectively.[3]

---

[2]Meyer (1999a) did not calculate Rorschach first-factor scores from all normally distributed variables. To parallel this research, this factor was created in his data set using the same normally distributed scores listed earlier. Results using this new scale are referred to throughout this article.

[3]These results differ slightly from those reported by Meyer (1999a). In the process of double checking Meyer's analyses, an error was discovered with an *A* scale score. Although all MMPI–2s for the Chicago patients had been computer scored through NCS, only the summary profiles were available for four patients. As such, the data for these patients could not be read by a computer program and had to be entered by hand. During manual entry of scores, one patient's *T* score on Scale *A* had been erroneously inserted as his raw score. This error was corrected for analyses reported in this article with the Chicago data. Note that this data entry error did not affect the classification of any patient, and thus it did not affect any of Meyer's primary analyses.

The 13 Rorschach scores and 11 MMPI–2 scores produced 143 variable pairs (i.e., *W* with Scale *1*, *W* with Scale *3*, etc.), which were labeled *CUVPs selected for high first-factor correlations* (CUVPs–HighFF). As expected, these variables were essentially uncorrelated in the full sample of 327 patients (mean r = .0165; range = –.11 to .12). Meyer (1999a) had reported very similar values (mean r = –.03; range = –.13 to .11).

The second set of CUVPs had been selected without regard to first-factor loadings. As before, there were 13 Rorschach scores and 11 MMPI–2 scales. The 13 Rorschach scores consisted of *W*, *D*, pure form (*F*), *FQo*, *FM*, Pairs, the adjusted *D* score (*AdjD*), Popular, *A*, *Id*, the affective ratio (*Afr*), *PER*, and good form quality percent (*X+%*).[4] These variables had an average correlation of .17 with the R–Engagement scale (range = –.40 to .56) and .22 with the Rorschach first factor (range = –.50 to .63). In the Chicago sample, the average correlation with R–Engagement was .21 (range = –.32 to .54) and the average correlation with factor scores was .28 (range = –.39 to .54). Thus, both samples found similar results. The second set of 11 MMPI–2 variables consisted of Scale *3*, Need for Affection (*Hy2*), *Ma4*, Gender Masculinity (*GM*), Gender Femininity (*GF*), Social Responsibility (*Re*), *FRS1*, *HEA1*, *ASP2*, *PSY–5–Agg*, and Positive Emotionality (*PSY–5–Pos*). These variables had an average correlation of –.05 with *A* (range = –.64 to .42) and an average correlation of –.03 with the MMPI–2 first factor (range = –.67 to .52). Results were similar in the Chicago sample; the corresponding correlations were –.05 with the *A* scale (range = –.73 to .55) and –.02 with the MMPI–2 first factor (range = –.76 to .65).

As with the first set of variables, these 13 Rorschach scores and 11 MMPI–2 scales produced 143 variable pairs that were essentially uncorrelated in the full sample of patients (*M r* = .01; range = –.14 to .11). These results mirrored those reported by Meyer (1999a; *M r* = –.01; range = –.19 to .15). The variables in this set were labeled *CUVPs not selected for first-factor correlations* (CUVPs–NotFF).

Applying the factor score selection procedures to MMPI–2 and MCMI–II scales. To understand the impact of selecting groups of patients for analysis, the procedures used for the MMPI–Rorschach convergent validity analyses were applied to the data from two self-report instruments, the MMPI–2 and the MCMI–II. These analyses were conducted on 269 patients from the Chicago sample (a second self-report instrument was not available in the Tennessee sample). As Meyer (1997b) reported, the first factors from the MMPI–2 and MCMI–II are highly correlated. Consequently, it is virtually impossible to find patients who would be classified as having opposing test interaction styles on the instruments, so the MMPI–2

[4]Meyer's (1999a, p. 11) text inadvertently listed 14 variables for this portion of his analysis. In contrast to what was written, the variable *DQo* had not been included in his analyses and was not included in our analyses.

and MCMI–II analyses were limited to selecting patients with similar styles on the tests. In addition, because profile criteria have not been developed for the MCMI–II and because we wished to simplify the presentation of this data, we only examined patients selected on the basis of factor scores from the first unrotated principal component of each test.

For the MCMI–II, a factor analysis was conducted of the clinical and validity scales (using raw scores). The first unrotated principal component accounted for 54.4% of the total variance; the second, third, fourth, and fifth components accounted for 19.4, 8.1, 6.9, and 3.1% of the total variance, respectively. Thus, as with the MMPI–2, the first factor of the MCMI–II is about twice as large as that of the Rorschach. Regression-based factor scores were used to indicate each patient's placement on the MCMI–II first factor. The MCMI–II and MMPI–2 factor scores had a correlation of .83 ($N = 269$, $p < .001$). The MCMI–II validity scale Disclosure was by far the best marker for the first factor of the MCMI–II, with a loading of .97 on this dimension. The correlation between Disclosure and the MMPI–2 *A* scale was .73 ($N = 269$, $p < .001$).

Using factor scores from the MMPI–2 and MCMI–II, 139 patients (51.7%) were classified as having a similar test interaction style across tests. More specifically, 72 patients fell in the upper third of the first-factor distributions on both tests, and 67 patients fell in the lower third of the first-factor distributions on both tests.

The same three constructs targeted in the Rorschach–MMPI analyses were used again for the MCMI–MMPI analyses. For the MMPI–2, negative affect, psychotic processes, and interpersonal wariness were measured by the variables described above. For the MCMI–II, negative affect was measured by the Dysthymia, Major Depression, and Anxiety scales; psychotic processes were measured by the Thought Disorder, Schizotypal, and Borderline scales, and interpersonal wariness was measured by the Schizoid, Avoidant, and Paranoid personality disorder scales. As in the prior analyses, these individual scales were aggregated to form composite measures of each construct (by summing *z* scores for each scale).

Estimating a lower boundary for construct convergence. As another strategy for assessing potential methodological artifacts, a formula for estimating a lower boundary of construct convergence was used. Even though the first factor from the MMPI–2 and the first factor from the Rorschach are expected to be uncorrelated, selecting the upper and lower thirds on these dimensions forces the primary factors to be correlated in the selected sample. It also forces any scales that are correlated with these dimensions to be correlated in the selected sample. In an effort to estimate the degree of correlation that resulted from selecting patients on the upper and lower thirds of each first factor, a formula from the factor-analytic literature was recommended by James M. Wood for this purpose (see Meyer, 1997b, pp. 320–321). The formula provides the expected correlation for two variables, each of which loads separately on two correlated factors. The formula relies on three correlations: (a) the correlation between an MMPI–2 scale and the MMPI–2 first factor (e.g., Scale *2* with MMPI–2 first-factor scores), (b) the correlation between a Rorschach scale and the Rorschach first factor (e.g., *DEPI* with Rorschach first-factor scores), and (c) the correlation between the MMPI–2 and Rorschach first factors in the selected sample (e.g., in those patients with similar styles as defined by factor scores). Once these correlations are obtained, the three values are multiplied. The product indicates the expected correlation for the MMPI–2 and Rorschach scales in the selected sample (e.g., the expected correlation of Scale *2* with *DEPI* in the group of patients with similar styles as defined by factor scores).

Once the predicted degree of correlation between each MMPI–2 variable and each Rorschach variable has been calculated, the predicted value can be subtracted from the observed correlation to obtain a residual. This residual correlation reflects the extent of construct overlap that remains after first-factor variance is removed.

There would not be a complication with this correction procedure if one could confidently attribute all first-factor variance to nonmeaningful forms of bias rather than to actual trait variance. In other words, if patients only obtained low scores on the Rorschach and MMPI–2 first factors because they were deliberately defensive (i.e., Style 1 in Meyer's, 1999a [Table 1], schematic representation) and only obtained high scores on the first factors because they were deliberately exaggerating disturbance (i.e., Style 4), then this correction formula would work accurately; however, patients are not high and low on the first factors simply because of deliberate efforts to manipulate the test data. For instance, many patients can obtain high scores on the first factor of each test because they have genuine and severe forms of psychopathology, particularly in the areas of affective distress and psychotic processes. Thus, to some extent, the formula overcorrects the observed correlations in the selected samples because it assumes that all the first-factor variance among the selected patients has to do with meaningless sources of influence rather than with genuine elements of affective distress, psychotic processes, and interpersonal wariness. Because this assumption is not true, the residual correlations provide an underestimate of Rorschach and MMPI–2 construct overlap. Nonetheless, if the residual correlations are interpreted as an estimated lower boundary or floor value, they provide a better coefficient for bracketing the true extent of construct overlap than do the coefficients obtained when response styles are ignored.

In the analyses presented here, formula-based corrections were applied to the Rorschach and MMPI–2 results derived from the Tennessee and Chicago samples. In addition, the corrections were applied to the MCMI–II and MMPI–2 convergent validity results. For the sake of uniformity across data sets, corrections were only applied to findings derived from factor score criteria (i.e., individual scales were correlated with factor scores, and corrections were applied to samples selected by factor scores).

Repeatedly drawing randomly selected samples from the full population. As a final way to check for sources of bias, convergent validity was assessed without selecting any patients on the basis of their test interaction styles. Instead, because small samples produce fluctuating or imprecise results (Meyer, 1999a), sample-to-sample fluctuations were used to evaluate the general relationship between test interaction styles and the convergence of Rorschach and MMPI–2 constructs. Specifically, by repeatedly drawing smaller samples from the population of 327 patients, the association between Rorschach and MMPI–2 first factors should fluctuate from one sample to the next. In turn, the correlations between Rorschach and MMPI–2 constructs should fluctuate across samples in tandem with the first factors.

To test this hypothesis, 300 patient samples were created by taking 300 random draws from the full population of 327 patients. On each draw, two variables were created. Variable A consisted of the correlation between the Rorschach and MMPI–2 first factors, and Variable B was created by correlating each of the 17 conceptually meaningful Rorschach and MMPI–2 construct pairs (i.e., Scale *2* with *DEPI,* Scale *2* with *S–CON,* Scale *7* with *DEPI,* etc.) and then averaging the 17 correlations to obtain a summary index of convergent validity among constructs. The correlation between Variable A and Variable B was then calculated across the 300 subsamples, treating each subsample as a single observation.

Following prior research, these analyses were conducted twice. On the first occasion, samples of 30 patients were randomly selected on each of the 300 draws. These samples were small enough to ensure substantial variability in the extent to which the first factors were (i.e., appeared to be) correlated in the 300 samples. On the second occasion, samples of 75 patients were randomly selected on each of the 300 draws. Although of moderate size, samples this large should show less random variation in first-factor correlations.

Finally, it should be noted that this study only examined the selection criteria and variables described previously. Findings were not culled from a larger array of results that may have led to different conclusions.

RESULTS

Convergence of Scales to Assess Test Interaction Styles

Table 2 reports the associations between Rorschach and MMPI–2 indicators of test interaction styles. Although there are a few statistically significant correlations, they are generally of small magnitude. The results in Table 2 closely mirror those reported by Meyer (1997b). In the Chicago sample, the same Rorschach scales had correlations from .02 to .09 with the MMPI–2 first factor, .05 to .10 with *A,* –.09 to .06 with *F,* .01 to .08 with *Fb,* –.10 to .09 with *L,* and –.11 to .04 with *K.*

TABLE 2
Correlations Between Scales of Test Interaction Styles From the Rorschach and MMPI–2: Tennessee Sample

	MMPI–2 Scale					
Rorschach Scale	*1st Factor*	*A*	*F*	*Fb*	*L*	*K*
Norm. 1st Factor	.10	.15**	.08	–.01	.17**	–.05
R–Engagement	.07	.13*	.04	–.04	–.18**	–.03
R	.08	.10	.06	.04	–.15**	–.03
Lambda	–.04	–.09	.00	.06	.23***	.03

Note. $N = 327$. MMPI–2 = Minnesota Multiphasic Personality Inventory–2. R–Engagement refers to the first principal component of the Rorschach using factor scores derived from a sample of college students (Meyer, 1992). Norm. 1st Factor refers to the same dimension quantified by factor scores derived from this sample with only variables that were normally distributed (see text). 1st Factor refers to factor scores for the first principal component of the MMPI–2 derived from this sample.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

With respect to the key variables used in our subsequent analyses, Table 2 indicates that R–Engagement had a correlation of .13 with the *A* scale, and the first principal component derived from the normally distributed Rorschach scores had a correlation of .10 with the first unrotated principal component derived from the MMPI–2 basic, content, and validity scales. For comparable variables, Meyer (1997b) reported similar coefficients of .10 and .04, respectively.[5]

Convergence of Construct Scales

Table 3 reports the convergent validity correlations obtained when all patients are considered and test interaction styles are ignored. Several variable pairs were statistically significant, although the correlations were again small in magnitude. The average correlation among the 17 variable pairs was .055. These results parallel those reported by Meyer (1997b), who found correlations ranging from –.09 to .15 across the same 17 variable pairs, with an average of .03.

Table 4 presents convergent validity results when the analyses are limited to patients who have similar test interaction styles on the two methods. For each of the three construct domains (emotional distress, psychotic processes, and interpersonal wariness), results are presented using the four methods for defining test interaction styles. Overall, regardless of which criteria were used, the results indicate that substantial correlations emerged between similarly named Rorschach and MMPI–2 scales. The average correlation across all the results listed in Table 4 is .42.

[5]Using the error-corrected *A* scale in Meyer's data, the correlation between R–Engagement and *A* is .11 rather than .10; using the Rorschach first factor derived from normally distributed variables in Meyer's data, the first factors had a correlation of .05 rather than .04.

TABLE 3
Correlations Between Rorschach and MMPI–2 Scales in Three Content Areas When Response Styles Are Ignored: Tennessee Sample

Affective Distress			*Psychotic Processes*		*Interpersonal Wariness*	
MMPI–2 Scale	*Rorschach DEPI*	*Rorschach S–CON*	*MMPI–2 Scale*	*Rorschach SCZI*	*MMPI–2 Scale*	*Rorschach HVI*
Scale *2*	.07	–.01	Scale *8*	.10	Scale *6*	–.04
Scale *7*	.13*	.08	*BIZ*	.06	*CYN*	–.02
DEP	.13*	.05	*PSY–5–Psy*	.01	*SOD*	–.03
ANX	.17**	.06			*TRT2*	–.02
PSY–5–Neg	.15**	.05				

Note. N = 327. MMPI–2 = Minnesota Multiphasic Personality Inventory–2; *DEPI* = Depression Index; *S–CON* = Suicide Constellation; *SCZI* = Schizophrenia Index; *HVI* = Hypervigilance Index; *DEP* = Depression; *ANX* = Anxiety, *PSY–5–Neg* = Negative Emotionality/Neuroticism; *BIZ* = Bizarre Mentation; *PSY–5–Psy* = Psychoticism; *CYN* = Cynicism; *SOD* = Social Discomfort; *TRT2* = Inability to Disclose.
*$p < .05$. **$p < .01$.

TABLE 4
MMPI–2 and Rorschach Convergent Validity for Three Constructs Using the Four Criteria Sets to Define Patients With Similar Response–Character Styles: Tennessee Sample

	Response–Character Styles Defined By:							
Construct	*Factor Scales (R–Engagement and A)*[a]		*Factor Scores*[b]		*Profile*[c]		*Factor or Profile*[d]	
Affective Distress	*DEPI*	*S–CON*	*DEPI*	*S–CON*	*DEPI*	*S–CON*	*DEPI*	*S–CON*
Scale *2*	.40***	.23*	.25*	.11	.52***	.34*	.29**	.17
Scale *7*	.52***	.45***	.36**	.33**	.61***	.51***	.43***	.35***
Depression	.58***	.48***	.38**	.29*	.59***	.41**	.43***	.33**
Anxiety	.51***	.44***	.41***	.29*	.65***	.39**	.46***	.33***
PSY–5–Neg	.48***	.36**	.34**	.26*	.56***	.27	.40***	.26**
Psychotic Processes	*SCZI*		*SCZI*		*SCZI*		*SCZI*	
Scale *8*	.65***		.66***		.60***		.55***	
Bizarre Mentation	.49***		.53***		.48**		.43***	
PSY–5–Psy	.47***		.53***		.45**		.40***	
Interpersonal Wariness	*HVI*		*HVI*		*HVI*		*HVI*	
Scale *6*	.44***		.46***		.34*		.40***	
Cynicism	.37**		.53***		.36*		.36***	
Social Discomfort	.48***		.49***		.31*		.44***	
Inability to Disclose	.45***		.38**		.01		.33**	

Note. See Table 3 for an explanation of abbreviations.
[a]n = 80. [b]n = 73. [c]n = 45. [d]n = 109.
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 5
MMPI–2 and Rorschach Convergent Validity for Three Constructs Using the Four Criteria Sets to Define Patients With Opposing Response–Character Styles: Tennessee Sample

	Response–Character Styles Defined By:							
Construct	*Factor Scales (R–Engagement and A)*[a]		*Factor Scores*[b]		*Profile*[c]		*Factor or Profile*[d]	
Affective Distress	*DEPI*	*S–CON*	*DEPI*	*S–CON*	*DEPI*	*S–CON*	*DEPI*	*S–CON*
Scale *2*	–.24*	–.33**	–.07	–.28*	–.05	–.24	–.13	–.24*
Scale *7*	–.29*	–.42***	–.03	–.23	–.00	–.06	–.15	–.27**
Depression	–.23*	–.44***	–.01	–.37**	.04	–.23	–.11	–.36***
Anxiety	–.28*	–.41***	–.00	–.24*	.04	–.08	–.10	–.26**
PSY–5–Neg	–.20	–.33**	.08	–.19	.17	.12	–.03	–.18
Psychotic Processes	*SCZI*		*SCZI*		*SCZI*		*SCZI*	
Scale *8*	–.42***		–.48***		–.33*		–.40***	
Bizarre Mentation	–.35**		–.35**		–.13		–.32**	
PSY–5–Psy	–.43***		–.43***		–.31*		–.39***	
Interpersonal Wariness	*HVI*		*HVI*		*HVI*		*HVI*	
Scale *6*	–.41***		–.56***		–.33*		–.42***	
Cynicism	–.38**		–.45***		–.48**		–.41***	
Social Discomfort	–.41***		–.44***		–.39**		–.38***	
Inability to Disclose	–.34**		–.37**		–.44**		–.32**	

Note. See Table 3 for an explanation of abbreviations.
[a]$n = 76$. [b]$n = 71$. [c]$n = 48$. [d]$n = 104$.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5 presents construct correlations for those analyses that were limited to patients who had an opposing style of interacting on the tests. The expectation that scales would be negatively correlated under these circumstances was generally supported. The average correlation across the results presented in Table 5 is –.26; however, the scales for negative affect were, on average, less highly correlated ($M = –.17$) than were the scales for psychosis ($M = –.36$) or for wariness ($M = –.41$). These results are in contrast to those found by Meyer (1997b, 1999a), who reported that the strongest negative correlations emerged for the construct of negative affect.

Table 6 presents summary information derived from Tables 3, 4, and 5. The average correlation for each construct under each of the selection conditions is presented, along with the average results in the full sample. Table 6 also lists the results for the composite variables. As expected, the composite variables demonstrate how aggregation reduces error. The composite scales produced validity co-

TABLE 6

Summary Coefficients (Means and Composite Variables) for Construct Pairs Hypothesized to Be Positively or Negatively Correlated and for Two Sets of 143 Construct Pairs Hypothesized to Be Uncorrelated: Tennessee Sample

	Criteria Used for Selection									
	Factor Scale (R–Engagement and A)		*Factor Score*		*Profile*		*Factor or Profile*		*No Selection (Full Sample)*	
Style and Variable	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*
Similar styles[a]										
Affective distress	.45***	.55***	.30**	.38**	.49***	.61***	.35***	.44***	.09	.11*
Psychotic processes	.54***	.58***	.57***	.61***	.51***	.54***	.46***	.50***	.06	.06
Interpersonal wariness	.44***	.59***	.47***	.62***	.26	.33*	.38***	.53***	–.03	–.04
CUVPs–NotFF	.00		.00		–.03		–.03		.02	
CUVPs–HighFF	.16		.18		.13		.14		.01	
Discordant styles[b]										
Affective distress	–.32**	–.39**	–.13	–.17	–.03	–.05	–.18	–.23*	.09	.11*
Psychotic processes	–.40***	–.43***	–.42***	–.45***	–.26	–.29*	–.37***	–.40***	.06	.06
Interpersonal wariness	–.39**	–.54***	–.46***	–.62***	–.41**	–.63***	–.38***	–.54***	–.03	–.04
CUVPs–NotFF	.04		.03		.04		.03		.02	
CUVPs–HighFF	–.10		–.16		–.16		–.11		.01	

Note. CUVP = conceptually unrelated variable pair; NotFF = not selected for first-factor correlations; HighFF = selected for high first-factor correlations. Mean coefficients are based on 10 construct pairs for affective distress, 3 construct pairs for psychotic processes, 4 construct pairs for interpersonal wariness, and 143 construct pairs for each set of CUVPs. The results from the full sample are reported in both the similar and discordant styles sections of the table for illustrative purposes.

[a]*ns* = 80, 73, 45, 109, and 327 for the factor scale, factor score, profile, combination criteria, and full sample, respectively. [b]*ns* = 76, 71, 48, 104, and 327 for the factor scale, factor score, profile, combination criteria, and full sample, respectively.

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

efficients that were higher than the average of the individual coefficients when associations were clearly evident in the data (e.g., for negative affect using factor scales to define similar styles, *r*s = .55 vs. .45), although the composite scales also indicated a lack of association when this conclusion was appropriate (e.g., in the full sample). Paralleling the results reported by Meyer (1999a), Table 6 suggests that the profile criteria are somewhat less effective than the factor criteria for producing the theoretically expected convergent associations.

Table 6 also presents results for the two sets of 143 variable pairs expected to remain uncorrelated after selecting patients on the basis of response–character styles. As anticipated, for each of the eight selection conditions, the average coefficients for both sets of 143 CUVPs were not large enough to be statistically significant. Also as expected, the CUVPs that were selected because they had high loadings on each method's first factor tended to track the first-factor correlations. When patients with similar styles were examined, the average of the CUVPs–HighFF variables had a positive sign (.13–.18). When patients with opposing styles were examined, the average of the CUVPs–HighFF variables had a negative sign (–.10 to –.16). Finally, as anticipated, the average correlations among the CUVPs that were selected without regard to first-factor loadings (CUVPs–NotFF) tended to hover near zero (–.03 to .04), and they often had a sign that was opposite to that which was expected for the meaningful variable pairs.

For patients with similar test interaction styles, all four sets of selection criteria produced convergent correlations for the meaningful variable pairs that were substantially larger than those produced by either set of CUVPs. All the meaningful constructs had mean *r*s greater than .30 and composite scale *r*s greater than .38, whereas all the CUVPs had mean *r*s less than .19. Thus, when various criteria are used to identify similar test interaction styles across methods, Rorschach and MMPI–2 constructs will tend to converge. This construct convergence does not appear to be caused by the simple alignment of methodological artifacts, because neither set of 143 CUVPs produced similarly large correlations.

For patients with opposing test interaction styles, all four criteria sets produced negative convergent validity coefficients for psychosis and wariness that were substantially larger than either set of CUVPs (i.e., the minimum composite *r* for the meaningful variables was –.29, and the maximum average for the meaningless variables was –.16). However, the results for negative affect were different. As expected, these scales were negatively correlated when using the factor scales of R–Engagement and *A* to define response–character styles (*M* = –.32, composite = –.39), but when using the other selection criteria, the correlations were no longer statistically significant and no longer capable of being differentiated from the CUVPs–HighFF variables.

The structure of Table 7 duplicates that of Table 6; however, the results in Table 7 are for the Chicago sample. Previously, Meyer (1999a) had calculated and presented results for two of the four sets of selection criteria listed in Table 7 (i.e., factor scales

TABLE 7

Summary Coefficients (Means and Composite Variables) for Construct Pairs Hypothesized to Be Positively or Negatively Correlated and for Two Sets of 143 Construct Pairs Hypothesized to Be Uncorrelated: Chicago Sample

	Criteria Used for Selection									
	Factor Scale[a] *(R–Engagement and A)*		*Factor Score*		*Profile*[a]		*Factor or Profile*		*No Selection (Full Sample)*	
Style and Variable	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*	*M*	*Composite*
Similar styles[b]										
Affective distress	.59***	.71***	.48***	.57***	.44***	.51***	.44***	.53***	–.01	–.01
Psychotic processes	.49***	.53***	.50***	.53***	.45***	.49***	.46***	.49***	.11*	.11
Interpersonal wariness	.37***	.50***	.45***	.59***	.38**	.50***	.36***	.50***	.05	.07
CUVPs–NotFF	–.03		–.02		.02		–.03		–.03	
CUVPs–HighFF	.03		.10		.18		.07		–.01	
Discordant styles[c]										
Affective distress	–.54***	–.63***	–.38**	–.45***	–.26	–.32*	–.41***	–.49***	–.01	–.01
Psychotic processes	–.19	–.20	–.42***	–.44***	–.16	–.18	–.15	–.16	.11*	.11
Interpersonal wariness	–.28*	–.34**	–.38**	–.47***	–.24	–.32*	–.29**	–.37***	.05	.07
CUVPs–NotFF	.01		.01		.02		.01		–.03	
CUVPs–HighFF	–.18		–.26		–.25		–.17		–.01	

Note. CUVP = conceptually unrelated variable pair; NotFF = not selected for first-factor correlations; HighFF = selected for high first-factor correlations. Mean coefficients are based on 10 construct pairs for affective distress, 3 construct pairs for psychotic processes, 4 construct pairs for interpersonal wariness, and 143 construct pairs for each set of CUVPs. The results from the full sample are reported in both the similar and discordant styles sections of the table for illustrative purposes.

[a]The results in these two sets of columns were previously reported by Meyer (1999a) and are presented here for completeness. [b]*ns* = 87, 85, 54, 111, and 362 for the factor scale, factor score, profile, combination criteria, and full sample, respectively. [c]*ns* = 78, 76, 51, 113, and 362 for the factor scale, factor score, profile, combination criteria, and full sample, respectively.

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

and profile criteria); however, he had not calculated results using the factor score criteria or the combined criteria. These analyses were generated[6] for this article to present exactly matching analyses across the two independent data sets.[7]

In general, it can be seen that the results are quite similar in both samples. The most notable disparities emerge with the discordant test interaction styles. The Chicago patients tended to produce larger negative convergent correlations for the negative affect construct than did the Tennessee patients. In contrast, the Chicago patients tended to produce lower correlations for the constructs of psychosis and wariness than did the Tennessee patients (except when the factor score criteria were applied to the Chicago sample).

Multitrait–Multimethod Matrices

Following Meyer (1997b), Table 8 presents a multitrait–multimethod matrix for the three Rorschach and MMPI–2 constructs. Composite scales were used in these analyses. Rather than presenting a separate table for each of the four selection criteria, to facilitate generalization only the results from the combined factor and profile criteria are presented (these criteria selected the largest number of patients with similar and opposing styles). In Table 8, results for the Tennessee sample are given above the diagonals and results for the Chicago sample are given below the diagonals. These specific findings for the Chicago sample had not been generated or reported previously. The first section of Table 8 reports results for all patients, whereas the second and third sections report results for patients selected to have similar and opposing styles, respectively.

Given that Tables 6 and 7 already presented the relevant data on cross-method convergent validity, the primary purpose of Table 8 is to consider cross-method convergent validity relative to cross-method discriminant validity. Before turning attention to this issue, however, it is important to consider the evidence on within-method discriminant validity. As expected, in each set of analyses (i.e., all patients, similar styles, and opposing styles) and in both samples, the MMPI–2 scales for the three different constructs are very highly correlated. These heterotrait–monomethod correlations had an average value of .79 across analyses and samples. In contrast, but also as expected, the same heterotrait–monomethod correlations among the three Rorschach constructs were much lower, averaging just .32

[6]For four of the Chicago patients, MMPI–2 first-factor scores were estimated using a regression equation based on the 13 basic and validity scales. This equation had a near-perfect ability to predict first-factor scores (R = .9948).

[7]When rechecking Meyer’s (1999a) analyses, a typographical error was found in the statistical syntax for computing the MMPI–2 negative affect composite variable. When this error was corrected, three of the four composite correlations increased by .01 in the theoretically expected direction. The corrected results are reported in Table 7.

TABLE 8

Summary Multitrait–Multimethod Matrix of Associations Between Rorschach and MMPI–2 Composite Scales Using the Combined Factor or Profile Criteria to Select Patients: Chicago Sample (Below the Diagonals) and Tennessee Sample (Above the Diagonals)

	Rorschach Scale			*MMPI–2 Scale*		
	Distress	*Psychosis*	*Wariness*	*Distress*	*Psychosis*	*Wariness*
All patients[a]						
Rorschach						
1. Distress	—	.24[b,***]	.21[b,***]	.11[c,*]	.11[d,*]	.10[d]
2. Psychosis	.27[b,***]	—	.34[b,***]	.08[d]	.06[c]	.07[d]
3. Wariness	.38[b,***]	.43[b,***]	—	.10[d]	.04[d]	–.04[c]
MMPI–2						
4. Distress	–.01[c]	.12[d,*]	.11[d,*]	—	.72[b,***]	.78[b,***]
5. Psychosis	–.03[d]	.11[c,*]	.02[d]	.74[b,***]	—	.77[b,***]
6. Wariness	.01[d]	.12[d,*]	.07[c]	.82[b,***]	.80[b,***]	—
Similar styles[e]						
Rorschach						
1. Distress	—	.37[b,***]	.33[b,***]	.44[c,***]	.39[d,***]	.42[d,***]
2. Psychosis	.43[b,***]	—	.41[b,***]	.54[d,***]	.50[c,***]	.50[d,***]
3. Wariness	.47[b,***]	.45[b,***]	—	.65[d,***]	.48[d,***]	.53[c,***]
MMPI–2						
4. Distress	.53[c,***]	.50[d,***]	.57[d,***]	—	.72[b,***]	.82[b,***]
5. Psychosis	.46[d,***]	.49[c,***]	.37[d,***]	.77[b,***]	—	.81[b,***]
6. Wariness	.59[d,***]	.52[d,***]	.50[c,***]	.83[b,***]	.84[b,***]	—
Opposing styles[f]						
Rorschach						
1. Distress	—	.16[b]	.11[b]	–.23[c,*]	–.11[d]	–.17[d]
2. Psychosis	.17[b]	—	.32[b,**]	–.42***	–.40[c,***]	–.33[d,**]
3. Wariness	.30[b,**]	.37[b,***]	—	–.49[d,***]	–.41[d,***]	–.54[c,***]
MMPI–2						
4. Distress	–.49[c,***]	–.17[d]	–.43***	—	.78[b,***]	.80[b,***]
5. Psychosis	–.42[d,***]	–.16[c]	–.34[d,***]	.78[b,***]	—	.78[b,***]
6. Wariness	–.47[d,***]	–.18[d]	–.37[c,***]	.88[b,***]	.84[b,***]	—

Note. MMPI–2 = Minnesota Multiphasic Personality Inventory–2. Coefficients are from the composite variables.

[a]*N*s = 362 and 327 for the Chicago and Tennessee samples, respectively. [b]Heterotrait–monomethod discriminant validity correlation. [c]Convergent validity correlation (monotrait–heteromethod). [d]Heterotrait–heteromethod discriminant validity correlation. [e]*n*s = 111 and 109 for the Chicago and Tennessee samples, respectively. [f]*n*s = 113 and 104 for the Chicago and Tennessee samples, respectively.

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

across analyses and samples. Thus, the MMPI–2 constructs were less differentiated than the Rorschach constructs. When patients were elevated on one type of scale from the MMPI–2 (e.g., negative affect), they were quite likely to be elevated on the other two types of scales as well (e.g., psychosis and wariness). With the Rorschach, this was much less likely to happen.

These within-method discriminant validity correlations also are consistent with the factor-analytic results reported in the Methods section. Recall that the first principal component of the MMPI–2 explained 48.3% of the total test variance, whereas the first principal component of the Rorschach explained a more modest 25.9% of the total variance. Both the factor-analytic findings and the heterotrait–monomethod findings in Table 8 indicate that test interaction styles of guarded defensiveness versus openness or exaggerated dilation have a stronger impact on MMPI–2 results than they do on Rorschach results. Consequently, the Rorschach has a better potential to measure unique or differentiated elements of personality than does the MMPI–2 (though whether the Rorschach actually does this is a separate question).

Returning to Table 8, a basic pattern of convergent and discriminant validity would be indicated when the cross-method convergent correlation for a construct exceeds the cross-method discriminant correlations from the column and row that intersect at that construct (Campbell & Fiske, 1959). For instance, when considering the Tennessee patients with similar styles, the Rorschach and MMPI–2 negative affect scales should correlate more highly with each other (.44) than with any alternative cross-method construct (i.e., Rorschach negative affect with MMPI–2 psychosis [.39] and with MMPI–2 wariness [.42]; MMPI–2 negative affect with Rorschach psychosis [.54] and with Rorschach wariness[.65]). Across all three sections of Table 8, the results do not demonstrate such a pattern of convergent and discriminant validity. This is true for both the Chicago sample and the Tennessee sample.

In addition to the pattern just described, convergent and discriminant validity can also be demonstrated when a variable's convergent correlation exceeds its own within-method discriminant correlations (see Campbell & Fiske, 1959). For instance, when considering the Tennessee patients with similar styles, the cross-method psychosis correlation (.50) should exceed the psychosis–negative affect and psychosis–wariness correlations that emerge within the Rorschach method (.37 and .41, respectively) and within the MMPI–2 method (.72 and .81, respectively). This pattern is not evident in any of the sections of Table 8 for either sample. However, Table 8 does provide some evidence for a weaker form of convergent and discriminant validity. Specifically, for patients identified as having similar test interaction styles, all the Rorschach scales demonstrated cross-method convergent validity coefficients that were higher than the Rorschach's corresponding within-method discriminant validity coefficients. Thus, the Rorschach composite scale of negative affect correlated more highly with the MMPI–

2 scale of negative affect than it did with the Rorschach scales of psychotic processes and wariness. Similarly, the Rorschach scale of psychosis correlated more highly with the MMPI–2 scale of psychosis than it did with the Rorschach scales of negative affect and wariness. In turn, the Rorschach scale of wariness correlated more highly with the MMPI–2 scale of wariness than it did with the Rorschach scales of negative affect or psychosis. This pattern was evident in both the Chicago and Tennessee samples.

For patients with opposing styles, the previous pattern was less consistent across samples. For the Tennessee patients, all three constructs displayed this pattern when considering the absolute value of the convergent correlation. Specifically, the Rorschach scale for distress had a convergent correlation of .23, and this cross-method correlation was larger than the corresponding within-method discriminant correlations (i.e., .16 with psychosis and .11 with wariness). For the Rorschach scale of psychosis, the cross-method convergent correlation was .40, and this was larger than both within-method discriminant correlations (.16 and .32). The Rorschach wariness scale had a convergent correlation of .54 with the MMPI–2, which exceeded both of the corresponding within-method Rorschach discriminant correlations (.11 and .32). In the Chicago sample, the Rorschach scale of negative affect correlated more highly with the MMPI–2 scale of negative affect (.49) than it did with the Rorschach scales of psychosis (.17) and wariness (.30). The Rorschach scale of wariness correlated more highly with the MMPI–2 scale of wariness (.37) than it did with the Rorschach scales of negative affect (.30), although it was equally correlated with the Rorschach scale for psychosis (.37). The Chicago results were least supportive of the Rorschach scale of psychosis. For this scale, the convergent correlation (.16) was actually smaller than both of its discriminant correlations (.17 and .37).

The MMPI–2 scales never showed a pattern of weak convergent–discriminant validity like that described for the Rorschach. In all analyses, the MMPI–2 scale for negative affect was more highly correlated with the MMPI–2 scales for psychosis and wariness than it was with the Rorschach scale for negative affect. In like fashion, the MMPI–2 scale for psychosis was always more highly correlated with the MMPI–2 scales for negative affect and wariness than it was with the Rorschach scale for psychosis, and the MMPI–2 wariness scale was always more highly correlated with the MMPI–2 scales for negative affect and psychosis than it was with the Rorschach wariness scale.

MCMI–II and MMPI–2 Correlations Using Patients Selected by Test Interaction Styles

Table 9 presents a multitrait–monomethod matrix that summarizes the results of the analyses conducted on the MCMI–II and MMPI–2. Only results for the com-

TABLE 9
Summary Multitrait, Single-Method Matrix of Associations Between MCMI–II and MMPI–2 Composite Scales When Using Factor Score Criteria to Select Patients

	MCMI–II Scale			*MMPI–2 Scale*		
	Distress	*Psychosis*	*Wariness*	*Distress*	*Psychosis*	*Wariness*
All patients[a]						
MCMI–II						
1. Distress	—					
2. Psychosis	.85[b,*]	—				
3. Wariness	.74[b,*]	.91[b,*]	—			
MMPI–2						
4. Distress	.87[c,*]	.78[d,*]	.69[d,*]	—		
5. Psychosis	.67[d,*]	.77[c,*]	.68[d,*]	.73[b,*]	—	
6. Wariness	.69[d,*]	.77[d,*]	.78[c,*]	.80[b,*]	.77[b,*]	—
Similar styles[e]						
MCMI–II						
1. Distress	—					
2. Psychosis	.91[b,*]	—				
3. Wariness	.83[b,*]	.94[b,*]	—			
MMPI–2						
4. Distress	.94[c,*]	.91[d,*]	.85*	—		
5. Psychosis	.77[d,*]	.89[c,*]	.84*	.80[b,*]	—	
6. Wariness	.82[d,*]	.90[d,*]	.90[c,*]	.88[b,*]	.86[b,*]	—

Note. MCMI–II = Milton Clinical Multiaxial Inventory–II; MMPI–2 = Minnesota Multiphasic Personality Inventory. Coefficients are from the composite variables.
[a]N = 269. [b]Heterotrait/within-instrument discriminant validity correlation. [c]Cross-instrument convergent correlation (monotrait–monomethod). [d]Heterotrait/cross-instrument (but monomethod) discriminant validity correlation. [e]n = 139.
*$p < .001$.

posite scales are given, first in the full sample and then in the sample selected by the factor score criteria. Because the results in Table 9 are drawn from within a single method, the most striking feature of Table 9 when compared with Table 8 is the magnitude of all the correlations. Two instruments from within the same method family produce larger correlations than two instruments from distinct method families. However, these larger correlations do not imply genuine validity. With the cross-instrument single-method coefficients in Table 9, the person-to-person variability in scores that results from methodological artifacts is thoroughly confounded with the person-to-person variability in scores that may result from genuine trait differences. Thus, these coefficients are artificially inflated by many factors (see Campbell & Fiske, 1959).

The second notable feature of Table 9 is the within-instrument discriminant correlations. Both the MMPI–2 and the MCMI–II produce very high correlations among these three presumably distinct constructs. Thus, both instruments are relatively hampered in their ability to discriminate among distinct constructs.

Table 9 also indicates that all correlations tend to increase when moving from the full sample to the sample of patients selected on the basis of similar test interaction styles. However, the changes are not nearly as dramatic as those found in Table 8.

When considering convergent and discriminant validity, Table 9 indicates that the MCMI–II and MMPI–2 results do not differ from the Rorschach and MMPI–2 results. Specifically, there is not a pattern of basic convergent and discriminant validity. The cross-instrument convergent correlations do not clearly or uniformly exceed the cross-instrument discriminant correlations from within the intersecting column and row. When considering all patients, neither the MMPI–2 nor the MCMI–II showed evidence of the weaker form of convergent and discriminant validity that had been displayed by the Rorschach in Table 8. However, this weak form of convergence and discrimination was evident for the MMPI–2 when patients with similar styles were examined. Under these selection conditions, the MMPI–2 scales correlated more highly with their corresponding MCMI–II scales than with MMPI–2 scales measuring alternative constructs.

Estimating a Lower Boundary for Construct Convergence

Table 10 presents the results that were obtained when calculating lower bound estimates of convergent validity. Six sets of analyses are presented. In the top section of the table, the Rorschach and MMPI–2 findings are presented for the Tennessee samples selected on the basis of similar and opposing styles. The middle section of the table presents the corresponding data using the Chicago sample. The final section of the table presents the MCMI–II and MMPI–2 analyses for all patients and for those identified as having similar styles. As indicated previously, to ensure uniformity across analyses, these findings were limited to stylistic patterns identified by the factor score criteria. For each set of analyses, two separate correction results are presented, one addressing the mean results for individual scales and the other addressing the composite scales. For each correction analysis, three values are reported: (a) the observed correlation, (b) the correlation predicted as a result of first-factor correlations, and (c) the residual correlation, calculated as the observed correlation minus the predicted correlation.

With one exception, the meaningfully related Rorschach and MMPI–2 variables had residual correlations that are in the expected direction (i.e., positive in the patients with similar styles and negative in the patients with opposing styles). The exception occurred with the construct of affective distress in the Tennessee patients with opposing styles. Here, the correlations were predicted to be larger (in a negative direction) than they actually were. In general, the data also indicate that the residual correlations were more substantial for patients with similar test inter-

TABLE 10
Lower Boundary Estimates of Construct Convergence: Observed Correlations, Correlations Predicted to Be Present as a Result of Selecting Patients by Factor Score Criteria, and Residual Correlations (Observed – Predicted)

	Individual Scale Correlation (M)			*Composite Correlation*		
Sample, Style, and Variable	*Observed*	*Predicted*	*Residual*	*Observed*	*Predicted*	*Residual*
Tennessee sample: Rorschach and MMPI–2 convergence						
Similar styles[a]						
Affective distress	.30	.20	.11	.38	.25	.13
Psychotic processes	.57	.29	.28	.61	.32	.29
Interpersonal wariness	.47	.29	.18	.62	.42	.20
CUVPs–HighFF	.18	.15	.03			
Opposing styles[b]						
Affective distress	–.13	–.22	.08	–.17	–.28	.11
Psychotic processes	–.42	–.33	–.09	–.45	–.35	–.10
Interpersonal wariness	–.46	–.32	–.14	–.62	–.46	–.15
CUVPs–HighFF	–.16	–.16	.01			
Chicago sample: Rorschach and MMPI–2 convergence						
Similar styles[c]						
Affective distress	.48	.26	.22	.57	.32	.25
Psychotic processes	.50	.27	.22	.53	.29	.24
Interpersonal wariness	.45	.29	.16	.59	.41	.18
CUVPs–HighFF	.10	.17	–.07			
Opposing styles[d]						
Affective distress	–.38	–.28	–.10	–.45	–.34	–.11
Psychotic processes	–.42	–.30	–.13	–.44	–.32	–.12
Interpersonal wariness	–.38	–.32	–.06	–.47	–.44	–.03
CUVPs–HighFF	–.26	–.18	–.07			
Chicago sample: MCMI–II and MMPI–2 convergence						
Unselected[e]						
Affective distress	.77	.60	.17	.87	.67	.19
Psychotic processes	.66	.56	.09	.77	.66	.11
Interpersonal wariness	.45	.35	.10	.78	.61	.17
Similar styles[f]						
Affective distress	.87	.68	.19	.94	.76	.18
Psychotic processes	.79	.64	.16	.89	.75	.15
Interpersonal wariness	.62	.40	.22	.90	.69	.21

Note. MMPI–2 = Minnesota Multiphasic Personality Inventory–2; CUVPs–HighFF = conceptually unrelated variable pairs selected for high first-factor correlations; MCMI–II = Millon Multiaxial Clinical Inventory–II. Mean coefficients are based on 10 correlations for affective distress, 3 correlations for psychotic processes, 4 correlations for interpersonal wariness, and 143 correlations for CUVPs–HighFF. The table entries contain slight imprecision because values were rounded to two decimal places.

[a]n = 73. [b]n = 71. [c]n = 85. [d]n = 76. [e]n = 269. [f]n = 139.

action styles than for patients with opposing styles. Across both the Chicago and Tennessee samples, the average residual correlations were .21 for similar styles and –.07 for opposing styles.

Although the corrected Rorschach and MMPI–2 coefficients are substantially lower than those initially observed, what is most revealing is the comparable impact that the correction equation had on the MCMI–II and MMPI–2 coefficients. Across all results, the residual MCMI–MMPI correlations averaged .16. For those patients who were selected on the basis of similar styles, the residual MCMI–MMPI correlations averaged .19. Both of these values were less than the average residual that was found for Rorschach and MMPI–2 convergent validity when patients were selected for similar styles (.21). Thus, despite the limitations that may be possessed by these residual lower bound estimates of construct convergence, the data indicate that for patients with a similar test interaction style, Rorschach and MMPI–2 construct convergence is at least as strong as MCMI–II and MMPI–2 construct convergence.

Response–Character Styles as General Moderators of Construct Convergence

The final set of analyses did not deliberately select any particular groups of patients. Instead, they capitalized on the random variation that emerges when taking samples from a population as a way to test the basic postulate that test interaction styles promote construct convergence. Samples were randomly drawn from the full population 300 times, and each time, two variables were computed: Variable A consisted of the correlation between Rorschach and MMPI–2 first factors, and Variable B consisted of the average correlation among the relevant Rorschach and MMPI–2 construct pairs. The correlation between Variable A and Variable B was then calculated across the 300 subsamples. This procedure was done twice, first with 300 random samples of 30 patients each and then with 300 random samples of 75 patients each.

When 30 patients were selected on each draw, the correlation between factor score correlations and construct correlations was .62 (N = 300, p < .001). When factor scales (i.e., R–Engagement and the MMPI–2 *A* scale) were used instead of factor scores for Variable A, the relevant correlation was .67 (N = 300, p < .001). The Chicago sample produced very similar correlations of .58 and .66, respectively (when using the first factor from the normally distributed Rorschach variables and the corrected *A* scale; see Footnotes 2 and 3). Finally, when the three pairs of composite scales were used for Variable B in the Tennessee sample, the correlations were .70 and .64 (N = 300, p < .001) for factor scores and factor scales, respectively.

When 75 patients were selected on each draw, the correlation between Variable A and Variable B was .58 when we used factor scores for the first principal component of each method and .60 when we used R–Engagement and the *A* scale (for both, $N = 300$, $p < .001$). The Chicago sample (Meyer, 1999a) produced very similar correlations of .60 and .65, respectively. Finally, when the three pairs of composite scales were used for Variable B in the Tennessee sample, the correlations were .69 and .59 ($N = 300$, $p < .001$) for factor scores and factor scales, respectively.

These results indicate that construct convergence tends to be a function of test interaction styles. In general, the stronger the positive or negative correlation between Rorschach and MMPI–2 first factors, the stronger the positive or negative correlation between Rorschach and MMPI–2 trait constructs.

The relation between these variables is graphically presented in Figure 1 for random samples of $n = 30$ and in Figure 2 for random samples of $n = 75$. These fig-

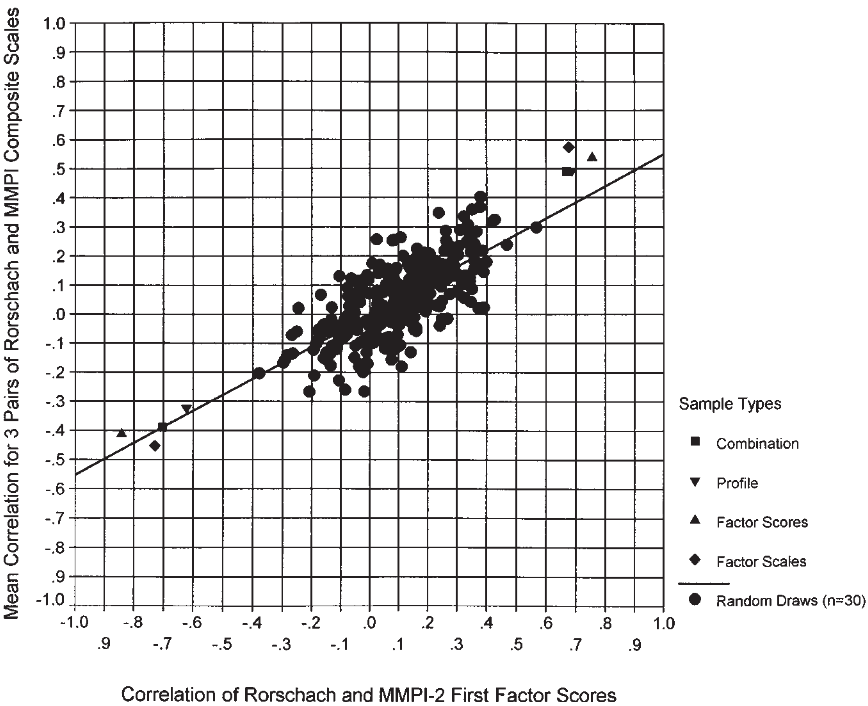

FIGURE 1 The convergence of Rorschach and Minnesota Multiphasic Personality Inventory–2 (MMPI–2) constructs as a function of test interaction styles across 300 randomly selected subsamples of 30 patients each. Also shown are the samples deliberately selected on the basis of factor score, factor scale, profile, and combined criteria.

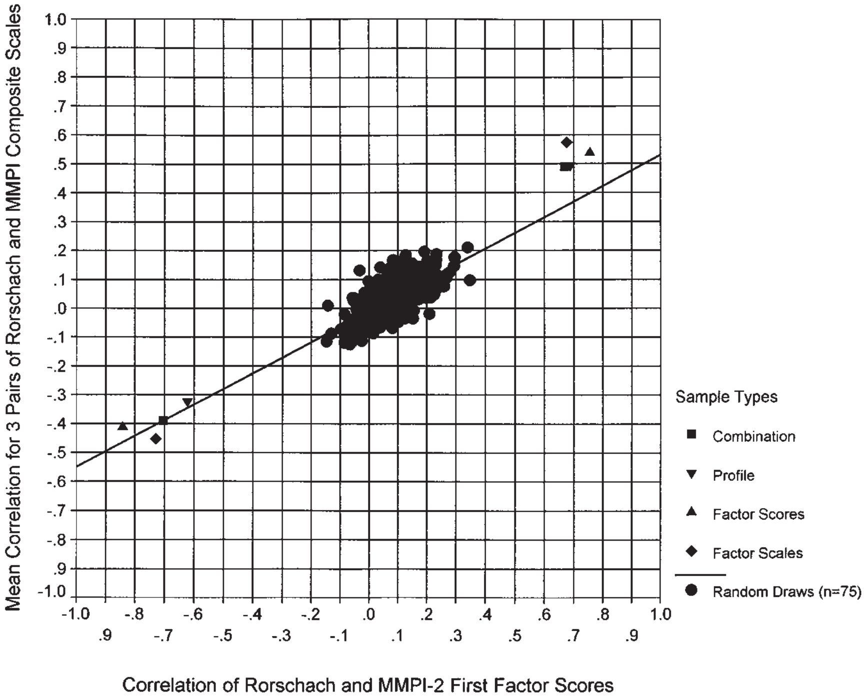

FIGURE 2 The convergence of Rorschach and Minnesota Multiphasic Personality Inventory–2 (MMPI–2) constructs as a function of test interaction styles using 300 randomly selected subsamples of 75 patients each. Also shown are the samples deliberately selected on the basis of factor score, factor scale, profile, and combined criteria.

ures use factor scores to quantify each first factor and the average of the three pairs of composite scales to quantify construct convergence. In both figures, the results from each of the 300 randomly drawn samples are represented by a circle. For reference purposes, results are also presented for the samples deliberately selected by each of the four selection criteria.[8] Samples selected by the factor scale criteria are indicated by diamonds, the factor score criteria are indicated by upward pointing triangles, the profile criteria are indicated by downward pointing triangles, and the combined criteria are indicated by squares.

When considering these figures, it should be recalled that the population consisted of all 327 patients with MMPI–2 and Rorschach data. In the population, Variable A, the correlation between Rorschach and MMPI–2 first factor scores, had a value of $r = .10$ (see Table 2), and Variable B, the average correlation among

[8]When these other data points are taken into account, the correlation across samples becomes .78 for $n = 30$ and .86 for $n = 75$.

the three pairs of composite scales, had a value of $r = .04$ (see Table 6). Thus, theoretically, for every sample obtained on one of the 300 random draws, Variable A should have a value of .10 and Variable B should have a value of .04. This does not happen, however, because random sampling error affects each draw, such that the results in each sample only approximate the population parameters. Smaller samples produce more sampling error than do larger samples. Consequently, as seen by the degree of dispersion in Figure 1 when compared with Figure 2, 300 draws of 30 patients at a time produce more widely disparate values for Variable A and Variable B than do 300 draws of 75 patients.

Statistical theory predicts that the average of the values observed across randomly selected samples should converge on the population parameter (Hays, 1981). Across the 300 samples, the average value for Variable A should be .10, and the average value for Variable B should be .04. This was the pattern obtained. In Figure 1, the average values are .09 and .05 for Variables A and B, respectively. In Figure 2, the values are .09 and .04, respectively.

In addition to supporting the hypotheses behind this line of research, Figure 1 also underscores how relatively small samples can produce results that differ dramatically from one investigator to another. Due to natural fluctuations in sampling error, some samples will find Rorschach and MMPI–2 first factors that are positively correlated, whereas others will find that they are negatively correlated or uncorrelated. Similarly, some samples will find Rorschach and MMPI–2 construct pairs that happen to be positively correlated, whereas other samples will find the same variables to be uncorrelated or negatively correlated.

Figures 1 and 2 also demonstrate that the samples deliberately selected on the basis of test interaction styles anchor the tails of each scattergram. Because these deliberately selected samples each contain a relatively large number of patients, Figure 2 provides the best indication of how they differ from randomly selected samples. Selecting patients on the basis of these styles is one way to produce positive (or negative) convergent validity coefficients between Rorschach and MMPI–2 constructs.

At the same time, however, Figures 1 and 2 demonstrate how the relation between first-factor correlations and construct correlations is far from perfect. For instance, Figure 1 illustrates how two relatively small samples can find the same moderately high degree of first-factor correlations, yet still observe very different results when it comes to the average degree of construct convergence. This can be seen by looking at the section of the figure where first-factor correlations fall in the range between .30 and .40. Within this column of data in the graph, it can be seen that the average degree of construct convergence can range anywhere from .00 to .40. Thus, variables other than test interaction styles also contribute to the convergence and divergence of Rorschach and MMPI data.

As found previously (Meyer, 1999a), the extent to which Rorschach and MMPI–2 first factors are correlated also helps to determine the extent to which there is a correlation among the CUVPs selected for high first-factor loadings (CUVPs–HighFF). Table 6 reveals that these meaningless variable pairs also tended to track first-factor alignment. As a result, when the average of the CUVPs–HighFF was used as Variable B and the random draw analyses were repeated, there continued to be a correlation between first-factor alignment and the revised Variable B. This was true for samples of 30 patients ($r = .38$ with factor scales; $r = .57$ with factor scores; for both, $N = 300$, $p < .001$) and for samples of 75 ($r = .38$ with factor scales; $r = .62$ with factor scores; for both, $N = 300$, $p < .001$). The key distinction between these analyses and those reported in Figures 1 and 2 was in the magnitude of the Variable B correlations; genuine construct pairs produced larger coefficients than did meaningless construct pairs. Replicating Meyer (1999a), a similar process was not evident for the CUVPs–NotFF set. When these variable pairs were used as Variable B and the random draws were repeated, there tended to be a negative correlation between first-factor alignment and the revised Variable B (for 30 patients, $r = –.05$ with factor scales, $r = –.04$ with factor scores; for 75 patients, $r = –.10$ with factor scales, $r = –.13$ with factor scores). Taken together, these results indicate that when Rorschach and MMPI–2 first factors are aligned in a sample, there will tend to be a correlation between any variable pairs so long as both members of the pair are highly correlated with their first factors. However, as seen in Table 6, the magnitude of the convergent correlations is most pronounced with conceptually meaningful variable pairs.

DISCUSSION

This article has presented a long and detailed series of analyses related to Rorschach and MMPI–2 convergent validity. Overall, results with a large sample of outpatients from Tennessee closely paralleled previous findings that relied on a large sample of inpatient and outpatients from Chicago. In addition, analyses using new and more liberal procedures to identify test interaction styles extended the previous results. In combination, the replicated and extended findings add credence to the idea that the convergence of Rorschach and MMPI–2 constructs is partially a function of how patients interact with the tests. When patients approach each test with a similar style, conceptually aligned constructs tend to correlate. Although a less robust finding, when patients approach each test in an opposing manner, conceptually aligned constructs tend to be negatively correlated. When test interaction styles are ignored, MMPI–2 and Rorschach constructs tend to be uncorrelated. An exception to the last conclusion occurs when a sample just happens to have a correlation between Rorschach and MMPI first factors. Under these conditions, there will also tend to be a parallel correlation among Rorschach and MMPI constructs.

Our general findings nicely replicate and extend the results of previous research. Nonetheless, it is still possible that unrecognized methodological artifacts have influenced the results from both samples of data. This prospect is minimized by the fact that we used three separate strategies to test for methodological artifacts, and none of them could account for the findings. First, as demonstrated in Figures 1 and 2, the pattern of Rorschach and MMPI–2 convergent validity holds when no selection criteria are imposed on the patient samples. Second, when selection criteria are imposed, convergent validity correlations for meaningful pairs of Rorschach and MMPI–2 scales consistently exceed the correlations found for conceptually meaningless variable pairs (see Tables 6 and 7). This was particularly true for those patients with similar test interaction styles. Finally, when we calculated lower bound estimates of construct convergence that corrected for our selection procedures, we saw that our observed correlations were consistently larger than those predicted on the basis of our selection methodology. In fact, as the data in Table 10 indicate, when the same selection procedures were applied to scores from two self-report tests, we found that Rorschach and MMPI–2 convergent validity correlations were slightly larger than MCMI–II and MMPI–2 convergent validity correlations.

Although our results could not be explained by the methodological challenges we tested, it is still possible that we have misunderstood some of the factors that appear to govern Rorschach and MMPI–2 convergent validity. As such, it would be optimal if other thoughtful methodological challenges were put forward and tested against the evidence.

In addition, although very similar findings have emerged from two large and independent data sets, further efforts at replication are warranted. To gain certainty about the generalizability of the findings and the processes that produce them, it would be optimal for other researchers to repeat the analyses on large data sets using the same procedures and variables reported here. If the general findings are replicated, it would then seem warranted to extend the analyses in new ways (e.g., to additional pairs of conceptually aligned MMPI–2 and Rorschach variables and to self-report tests other than the MMPI–2). One way in which these analyses could be extended would be to apply the same procedures to an adolescent population. Although Krishnamurthy, Archer, and House (1996) reported no evidence for the convergence of Rorschach and MMPI–2 constructs in an adolescent sample and Archer and Krishnamurthy (1999) more recently criticized elements in this general line of research, a direct attempt at replication has not yet been tried with their adolescent data (see Meyer, 1999a).

Because the analyses described previously are complex and involve numerous steps, they are prone to errors (e.g., Footnotes 3 and 7; cf. Meyer, 1998). Also, researchers attempting to replicate the design may have questions about explicit procedures that go beyond the information described in the Method section of this article. To facilitate the work of other researchers, several SPSS syntax files have been prepared that detail the steps and commands used for the analyses. These syntax files can be obtained by contacting Gregory J. Meyer.

Despite the generally remarkable degree of replication across samples, some ambiguities remain. In particular, it is unclear how one should understand some of the discrepancies that have emerged for patients with opposing test interaction styles. Meyer's (1997b, 1999a) original analyses found strong convergence for negative affect scales but only mild convergence for scales of psychotic processes and wariness. In a new analysis of his data (see Table 7), strong convergence was found for all three trait constructs when factor scores were used to identify test interaction styles. In contrast, the Tennessee sample found strong convergence for psychotic processes and wariness but relatively little evidence of convergence for negative affect,[9] except when factor scales were used to identify stylistic patterns.

In combination, these results indicate that there is a greater degree of instability associated with discordant test interaction styles than with similar test interaction styles. This instability may be a function of general sampling error, or it may indicate that the results are quite sensitive to initial sampling conditions in which relatively small differences in the patients selected for analysis have a large impact on the final results. Alternatively, given the (limited) data that has been presented on patients with opposing test interaction styles (see Meyer, 1997b, pp. 323–326), the somewhat inconsistent findings may point to particular dynamics and defensive operations in the kinds of patients classified by the various criteria or in the kinds of patients seen in the Chicago setting versus the Tennessee setting.

Another ambiguity associated with this line of research has to do with the criteria for defining styles. The profile criteria were fixed across both samples, with the same cutoff scores being employed for each. Because normative data are available for each of the profile scores, a fixed set of criteria has seemed appropriate. However, the factor criteria have been relative rather than fixed. As such, patients have been selected for analysis using the upper and lower thirds of the scale distributions, rather than using fixed values to define high and low scores. Because normative data are not available for the factor score criteria, reliance on relativistic criteria has seemed appropriate. However, it would be worthwhile to explore the extent to which these findings can be replicated using fixed criteria for the factors.

A final ambiguity relates to the multitrait–multimethod matrices. These matrices did not provide clear evidence of convergent validity relative to discriminant validity. This was true when Rorschach and MMPI–2 scales were evaluated in a traditional multitrait–multimethod matrix, but it was also true when MCMI–II and MMPI–2 scales were evaluated in a multitrait–monomethod matrix. At best, there was evidence for a weak form of convergence and discrimination. However, this pattern was only present for the Rorschach when considering the Rorschach–MMPI–2 correlations, and it was only present for the MMPI–2 when considering

[9]In the Tennessee sample, scatter plots suggested a more curvilinear relation to the data, such that Rorschach indicators of affective distress were lower when MMPI–2 indicators of affective distress were either very low or very high.

the MCMI–MMPI correlations. Although these results are discouraging, the matrices in Tables 8 and 9 are fairly typical of what is found in the literature. For instance, in their classic work, Campbell and Fiske (1959) reported 11 tables of multitrait–multimethod correlations. Because some tables relied on three or four distinct methods of assessment, there was a total of 24 cross-method comparison cells. Only one (4.2%) displayed the basic pattern of expected convergent and discriminant validity. In the same article, there were 48 instances when cross-method convergent correlations could be compared to within-method discriminant correlations. Only three examples (6.3%) showed the weak pattern of convergence and discrimination we observed. Findings similar to those in Tables 8 and 9 can be found in recent large-scale studies (e.g., Cole, Truglio, & Peeke, 1997) and in Achenbach, McConaughy, and Howell's (1987) meta-analytic review (but see Costa & McCrae, 1992, for more promising results).

With respect to clinical practice, one implication that should be obvious is that the Rorschach and the MMPI–2 are distinct assessment methods. Globally, the results of one test are independent of the results of the other. Because of this, clinicians should use both measures when they are interested in obtaining a broad understanding of their patients' personalities. Each method alone is incomplete and may provide a limited degree of information.

With the global independence of Rorschach and MMPI–2 scores as our launching point, the goal of the analyses detailed here was simply to demonstrate that this independence can be altered to produce cross-method correlations if researchers take into account the manner in which patients interact with these two testing tasks. Clinicians can anticipate a greater degree of correspondence between Rorschach and MMPI–2 findings related to affective distress, psychosis, and interpersonal wariness when patients display a similar style of interacting with each task. Conversely, clinicians can anticipate that these Rorschach and MMPI–2 constructs will show a greater degree of disagreement when patients approach each task in a very different manner (i.e., a constricted style on one test and a dilated style on the other).

Unfortunately, knowing about these cross-method correlations does not translate into direct or easy implications for clinical practice. In part, this is because the effects under consideration, although of substantial magnitude, do not apply to each and every patient. Some patients with similar test interaction styles still produce rather different looking Rorschach and MMPI–2 scores, whereas some patients with opposing test interaction styles still produce fairly similar test scores. A second complicating factor is that test interaction styles do not emerge from uniform patient characteristics. Rather, very different factors may give rise to a constricted style or a dilated style on these tests. For instance, an MMPI protocol and a Rorschach protocol may substantially agree with each other, showing either relatively few or many signs of disturbance; however, this agreement may be the result of a patient's deliberate efforts to distort the data obtained from each test. Alternatively, the same pattern of agreement may reflect the patient's genuine limitations and difficulties. Obviously, given these alternatives, a clinician could interpret the very same pattern of test results in vastly different ways. Unfortunately, test scores alone do not seem capable of providing the clinician with sufficient clues about how the results should be interpreted. Rather, test scores have to be considered in conjunction with other information from referral sources, significant others, history, and observed behavior, as well as the general context of the evaluation, to determine which of these very different alternatives (distorted vs. genuine results) should be given credence. These are interpretive complications that readily arise when the Rorschach and MMPI generally agree with each other. Different factors come to the forefront when considering Rorschach and MMPI scores that disagree with each other. Under these circumstances, the clinician not only needs to consider the preceding factors but must also take into account the strengths and limitations of both tests, the sensitivity of each method to different elements of personality, and the legitimate personality dynamics that may produce a pattern of opposing test scores.

Currently, there is more theorizing than hard data available to understand the types of patients and processes that produce discrepant MMPI–2 and Rorschach results. However, clinically informed reasoning has provided useful suggestions for conceptualizing these test patterns (e.g., Finn, 1996; Meyer, 1996, 1997b, 1999a), and preliminary research has documented strong, expected associations between discrepant Rorschach–MMPI findings and criterion diagnoses that indicated that conscious awareness was being protected from underlying emotional distress (Meyer, 1997b). It will be useful for future research to explore and document in more detail the characteristics of the patients who produce each of the four basic cross-method patterns (i.e., constricted on both methods, dilated on both methods, constricted on the Rorschach and dilated on the MMPI–2, and constricted on the MMPI–2 and dilated on the Rorschach). Given that each pattern can emerge from quite different personality characteristics and motivational states, it may be best to begin the analyses by identifying groups of patients who have the personality traits or motivational states that should lead to a particular cross-method pattern and then examine the test data to see if the expected pattern emerges. All in all, there is much more to be learned and documented about the combined use of the Rorschach and MMPI–2 in clinical practice. Clinically informed qualitative methods and fine-grained, multimethod criterion measures of personality should serve us well as we pursue these goals.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Archer, R. P., & Krishnamurthy, R. (1999). Reply to Meyer on the convergent validity of the MMPI and Rorschach. *Journal of Personality Assessment, 73,* 319–321.

Ben-Porath, Y. S., & Sherwood, N. E. (1993). *The MMPI–2 content component scales: Development, psychometric characteristics, and clinical application.* Minneapolis: University of Minnesota Press.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology, 43,* 1254–1269.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

Cole, D. A., Howard, G. S., & Maxwell, S. E. (1981). Effects of mono- versus multiple-operationalization in construct validation efforts. *Journal of Consulting and Clinical Psychology, 49,* 395–405.

Cole, D. A., Truglio, R., & Peeke, L. (1997). Relation between symptoms of anxiety and depression in children: A multitrait–multimethod–multigroup assessment. *Journal of Consulting and Clinical Psychology, 65,* 110–119.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51,* 360–392.

Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E., Jr., Cohen, J. B., & McGuire, H. (1990). *RIAP Version 2: Rorschach Interpretation Assistance Program.* Asheville, NC: Rorschach Workshops.

Finn, S. E. (1996). Assessment feedback integrating MMPI–2 and Rorschach findings. *Journal of Personality Assessment, 67,* 543–557.

Harkness, A., McNulty, J., & Ben-Porath, Y. (1995). The Personality Psychopathology Five (*PSY–5*): Constructs and MMPI–2 scales. *Psychological Assessment, 7,* 104–114.

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

Krishnamurthy, R., Archer, R. P., & House, J. J. (1996). The MMPI–A and Rorschach: A failure to establish convergent validity. *Assessment, 3,* 179–191.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46.

Meyer, G. J. (1992). The Rorschach's factor structure: A contemporary investigation and historical review. *Journal of Personality Assessment, 59,* 117–136.

Meyer, G. J. (1996). The Rorschach and MMPI: Toward a more scientifically differentiated understanding of cross-method assessment. *Journal of Personality Assessment, 67,* 558–578.

Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment, 9,* 480–489.

Meyer, G. J. (1997b). On the integration of personality assessment methods: The Rorschach and MMPI–2. *Journal of Personality Assessment, 68,* 297–330.

Meyer, G. J. (1998). Error in research and assessment data with an erratum for Meyer (1993). *Journal of Personality Assessment, 71,* 195–211.

Meyer, G. J. (1999a). 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. (1999b). Simple procedures to estimate chance agreement and kappa for the interrater reliability of response segments using the Rorschach Comprehensive System. *Journal of Personality Assessment, 72,* 230–255.

Meyer, G. J., Baxter, D., Exner, J. E., Jr., Fowler, J. C., Hilsenroth, M. J., Piers, C. C., & Resnick, J. (1999). *An examination of interrater reliability for scoring the Rorschach Comprehensive System in five data sets.* Manuscript submitted for publication.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory–II: Manual for the MCMI–II* (2nd ed.). Minneapolis, MN: National Computer Systems.

Overholser, J. C. (1992). Aggregation of personality measures: Implications for personality disorder research. *Journal of Personality Disorders, 6,* 267–277.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94,* 18–38.

Tsujimoto, R. N., Hamilton, M., & Berger, D. E. (1990). Averaging multiple judges to improve validity: Aid to planning cost-effective clinical research. *Psychological Assessment, 2,* 432–437.

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

Received June 22, 1999
Revised November 15, 1999