

## On the Integration of Personality Assessment Methods: The Rorschach and MMPI

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

Despite being the most studied and used personality assessment tools, data from the Rorschach and MMPI generally disagree (Archer & Krishnamurthy, 1993a, 1993b). Independence is proposed to result from at least 3 factors: (a) the methods tap unique levels of personality, (b) personality has a complex organization, and (c) response styles generate considerable method variance that must be considered in nomothetic research. These ideas led to 5 hypotheses, each of which received support. Rorschach and MMPI response styles are uncorrelated, although response styles are quite consistent within a method family. MMPI-2 and Rorschach constructs of dysphoria, psychosis, or wariness are uncorrelated when response styles are ignored. However, robust convergent validity is evident when patients have similar response styles on each method (e.g., for dysphoria,  $M r = .59$ ) and dysphoria is expressed in opposing ways on each method when response styles are discordant (i.e.,  $M r = -.54$ ). Data from the latter analyses were correlated with genuine clinical phenomena and implications were discussed for clinical practice and research.

The Minnesota Multiphasic Personality Inventory (MMPI/MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Rorschach Inkblot Method are the two most intensively researched, most commonly taught, and most frequently used personality assessment tools (Archer & Krishnamurthy, 1993b; Lubin, Larsen, & Matarazzo, 1984; Piotrowski & Keller, 1992; Piotrowski & Zalewski, 1993). Befitting this status, Parker, Hanson, and Hunsley's (1988) meta-analytic review demonstrated both measures have substantial reliability and good convergent validity when research is conducted with a sound theoretical or empirical rationale.

Although it would seem that psychologists could be content knowing the core measures in the field are reliable and valid, this is not yet the case. After reviewing 44 studies examining MMPI and Rorschach intercorrelations, Archer and Krishna-

murthy (1993a, 1993b) demonstrated the tests generally disagree. At best, a minimal relation exists between similarly named MMPI and Rorschach variables. Why these measures give disparate views of personality has become a significant conundrum for the field of assessment. Archer and Krishnamurthy (1993a) encouraged clinicians who are confronted with Rorschach and MMPI discrepancies to "make the clinical decision to emphasize certain aspects of test findings, while suppressing results from other sources of test data" (p. 138). To guide decisions about which data should be emphasized, clinicians were encouraged to "consider the relative reliability and validity of the specific data sources" (p. 138) in order to suppress that which was least reliable and valid. Such recommendations would be most appropriate if one method was clearly superior to the other.

However, if Parker et al. (1988) are correct and each tool is equally valid and reliable, such guidelines will not be sufficient. Instead, it becomes important to articulate why Archer and Krishnamurthy's findings should emerge as they do. This article will propose several clinical and methodological reasons for Rorschach and MMPI disagreement and present data demonstrating convergent validity when these factors are considered. As a starting point, however, it is useful to define the major differences between these two measures.

### RORSCHACH AND MMPI DIFFERENCES

Historical precedent has left us with the terms *objective* to describe the MMPI and *projective* to describe the Rorschach. These terms are misnomers that obscure rather than clarify. In particular, they imply the MMPI is factual and precise, while suggesting the Rorschach is subjective or nonfactual. In actuality, both should be considered factual but imprecise measures of personality, each of which is adept at quantifying particular kinds of personality constructs (Meyer, 1996). More accurate terms with fewer untoward connotations are "self-assessment" to describe the MMPI and "performance-based" to describe the Rorschach.

Table 1 outlines some of the primary distinctions between the self-assessment/MMPI and performance-based/Rorschach methods. At the risk of being simplistic, some of the most obvious differences relate to the nature of the tasks. The MMPI lays out clear directions for how the task should be completed, and the task itself is familiar to anyone who has attended grammar school and completed a paper-and-pencil test. The MMPI also gives the patient only three response options (i.e., true, false, or no response). In contrast, the Rorschach presents stimuli that most people have never encountered before and only provides minimal structure. The expectations are ambiguous, consisting simply of the question "What might this be?," and each patient has extensive latitude for responding.

Other differences relate to what a patient must do to complete the measure. Self-assessments require patients to think about themselves, decide if certain

**TABLE 1**  
**Distinctions Between the Self-Assessment/MMPI**  
**and the Performance-Based/Rorschach Methods**  
**of Personality Assessment**

<i>Characteristics of the Self-Report Method</i>	<i>Characteristics of the Rorschach Method</i>
1. Expectations are well-defined.	1. Expectations are minimally defined.
2. Stimuli are familiar.	2. Stimuli are novel.
3. Narrow range of response options.	3. Wide range of response options.
4. Task requires patient to consider self, decide if traits are characteristic, decide how to present self, and then indicate decisions on paper.	4. Task requires patient to formulate perceptions, decide which perceptions to articulate to the examiner, and then respond to further questions.
5. Administration and scoring require minimal skill.	5. Administration and scoring require considerable skill.
6. Patients are assumed to use a similar benchmark for deciding if trait is characteristic of themselves.	6. Examiner provides stable benchmark for classifying patient characteristics.
7. Measure completed alone.	7. Measure completed with an examiner.
8. Requires an in vitro description of personal characteristics.	8. Requires an in vivo demonstration of personal characteristics.
9. At best, raw data is dependent on conscious awareness and complexity of self-representations.	9. At best, raw data is dependent on engagement with the task and ability to articulate perceptions and their determinants.
10. Dissimulation and impression management affect reported symptoms.	10. Dissimulation and impression management affect engagement with task.
11. Better tool for obtaining information about specific overt symptoms, events, and experiences.	11. Better tool for assessing personality predilections which may or may not be evident in overt behavior or consciousness.

attributes are characteristic of them, decide if they want to share this information, and then report their decisions on a piece of paper. The Rorschach requires that patients sit down with an examiner, look at 10 inked cards, formulate perceptions of what each stimulus looks like, decide if they want to share this information, articulate some responses to the examiner, and then respond to further questions during the inquiry stage.

Because the task demands are quite different, it should not be surprising if the Rorschach and MMPI obtain qualitatively different types of information. Nor should it be surprising if this information is subject to unique sources of bias and influence.

One differential source of influence is the skill required to administer and score the measures. Although considerable skill is needed to develop the items that form any self-assessment scale (as with any Rorschach scale), only minimal skill is

required to administer and score a self-assessment measure. A much greater level of sophistication is necessary to accurately perform these functions with a Rorschach. The examiner must be relatively neutral; create a cooperative working alliance; know how to encourage or constrain a patient giving too few or too many responses; know the cue words that indicate an inquiry question is necessary; know how to effectively frame inquiry questions to address specific issues; and know how to accurately classify the where, what, and why of the patient's perception.

Another source of influence concerns who has greater responsibility for organizing the test data. With a self-assessment measure, each patient must accurately understand the items and all patients must use roughly equivalent mental benchmarks for deciding whether each characteristic is "mostly true" or "mostly false." With the Rorschach, responsibility for accurately decoding and consistently classifying each response is in the hands of the examiner, who again requires substantial skill to do this properly.

A third differential influence is the "interpersonalness" of each task. All self-report instruments are completed in private and do not require active interaction with an examiner. The Rorschach, on the other hand, requires consistent interaction. As such, the interpersonal dynamics of both the patient and the examiner are much more important to the completion of the Rorschach (see Schafer, 1954). Some patients may respond better to an MMPI because they prefer having control over what they disclose or because they prefer to communicate their personal attributes in a private and somewhat more anonymous fashion. Other patients may respond better to a Rorschach because they need the interpersonal structure of the task or because they are more comfortable working with someone rather than in isolation.

A fourth difference relates to the demands placed on participants. Self-assessments require patients to make personal judgments about each item in order to provide a description of what they are like. Consequently, the raw information obtained from an MMPI is very dependent on the quality of their conscious self-schema. For this information to be accurate, patients must be insightful, must not have strong emotional conflicts that result in obscuring defenses, and must have stable, well-articulated, and well-differentiated understandings of themselves.<sup>1</sup> The Rorschach does not require a patient to directly communicate personal knowledge. Rather, it requires patients to behaviorally perform. As such, it elicits a demonstration of personal characteristics, rather than a description of them (see Miller, 1987),

---

<sup>1</sup>As David Nichols, Radhika Krishnamurthy, and a helpful reviewer noted, some MMPI experts would dispute this characterization, pointing to the empirical construction of the basic scales, the presence of subtle items, the research documenting scale correlates, and the configural approach to interpretation as evidence the MMPI yield is not limited to consciously accessible information. I am not trying to suggest otherwise. Rather, my point is simply that patient responses to all items, subtle or otherwise, are dependent on their conscious understanding of themselves. This does not mean clinical inferences are limited to the consciously bound raw information provided by patients. Clinicians can certainly infer beyond what patients consciously recognize or deliberately intend to communicate.

which makes it more akin to intelligence or neuropsychological tests than to the MMPI. On the Rorschach, information is gleaned from subtle differences in what perceptions are articulated, their location, the manner in which they are described, and the qualities of ink described as giving rise to them. Although patients completing a Rorschach consistently judge the adequacy of their responses and inhibit or censor certain perceptions, because most information is obliquely obtained through nonobvious scoring categories, the information is not necessarily filtered through the patient's self-schema nor dependent on conscious awareness (see Bornstein, Rossner, Hill, & Stepanian, 1994).<sup>2</sup> Simultaneously, however, Rorschach scores are quite dependent on the ability of patients to spontaneously engage with the task and articulate responses. Cognitive and emotional difficulties that interfere with verbal articulation, behavioral spontaneity, or task engagement compromise score accuracy, and these confounds find no substantial parallel on a self-assessment measure.

Both methods are also subject to conscious efforts at impression management, although the nature of these influences differ. For self-assessments, patients must be willing to accurately share what they know to be true about themselves. Of course, some may decide not to do this and may pervasively deny unseemly traits, deny a circumscribed domain of characteristics (e.g., hearing voices), overemphasize problems in a selected area of symptomatology, or pervasively endorse pathology. Somewhat similar strategies can be taken with the Rorschach, such that patients can pervasively inhibit all kinds of responses, selectively censor certain classes of responses (e.g., those with sexual content), selectively emphasize certain kinds of responses, or pervasively respond to the task with dramatic and intense perceptions. However, because it is generally not clear how responses will be scored, *efforts at impression management are more of a gamble with the Rorschach than the MMPI.*

Finally, because of the foregoing, each method differs with respect to the information it can deliver under optimal circumstances. Self-assessments can be very effective when obtaining information about specific symptoms or experiences, whereas the Rorschach can be very effective when obtaining information about underlying personality characteristics, propensities, and mental representations. Consider questions related to hallucinations, suicidal plans, phobic fears, compulsive checking, or difficulty with early morning awakening. It does not matter how

---

<sup>2</sup>At times, even content with obvious personal relevance is articulated to the inkblots without the patient's awareness, as some patients experience their perceptions as external "facts" rather than personal productions. A particularly vivid example occurred with a patient hospitalized because of increasing homicidal impulses toward his wife. Shortly after admission he denied any homicidal ideation and wished to be discharged. However, when he was subsequently tested he reported Card II of the Rorschach looked like two men giving each other a high-five while stomping on a uterus that had blood shooting out of it. Despite his protests to the contrary, it required few inferential steps to conclude he was still preoccupied with hostile, aggressive, and domineering impulses toward women.

many Rorschach scores are considered, an examiner will never know the extent of these problems unless the patient is given a task like the MMPI that asks about them directly and then quantifies the responses. On the other hand, consider questions like the extent to which a patient perceives external events in a conventional fashion, has a psychic environment populated with nonhuman characters, has ideas that become illogical and inappropriately fused together in the absence of clear external structure, implicitly envisions interactions as destructive and conflictual, or is prone to focus on idiosyncratic details within the environment. It does not matter how many MMPI scores are considered; an examiner will never be able to quantify these characteristics unless the patient's behavior is observed in a controlled and normed setting like the Rorschach.

In summary, the strength of the MMPI is that it can quickly obtain specific symptomatic information and yield a portrait of the patient from the patient's perspective. Its limitations stem from its reliance on the patient to make sense of all the items and convey information to the examiner accurately and honestly. The strength of the Rorschach is that it requires a behavioral demonstration of personal characteristics, rather than a simple description. As such, it is able to obtain information that the patient may not be cognizant of and can do this to some extent despite a conscious effort not to reveal certain attributes. Limitations include the fact that scores are dependent on open engagement with the task and the ability to articulate perceptions.

Given all these differences, one reason MMPI and Rorschach scales are not strongly correlated may be because the methods are really measuring quite different things. In the past, psychology's propensity to apply the same term to two very different things has been termed the *jingle fallacy*, which stands in contrast to the *jangle fallacy* or our equally strong propensity to use different terms to describe what is actually the same thing (cf. Block, 1995). With the Rorschach and MMPI, a jingle fallacy emerges when we use the generic term *depression* to refer to the MMPI's Depression Content Scale and also the Rorschach Depression Index (*DEPI*). Depression on the MMPI is really depression as it is consciously understood and deliberately reported by the patient. Depression on the Rorschach is really depression as it is manifest in implicit perceptual propensities and qualities of verbal articulation. These should not be considered equivalent.

Just as MRI and PET scans provide distinct information about brain functioning, and optical and infrared telescopes provide distinct information about galaxies, MMPI and Rorschach scores ought to capture unique information about personality. It is curious, however, that personality research has developed such strong expectations for two very different tools to yield quite similar information. Rather than appreciating Rorschach and MMPI differences, the field often seems stuck on the goal of demonstrating both methods provide redundant information. This expectation probably has roots in Campbell and Fiske's (1959) pioneering work on test validity.

## THE MULTITRAIT–MULTIMETHOD MATRIX

Campbell and Fiske's (1959) article on convergent and discriminant validity has had a broad impact on psychology, being the most frequently cited publication in the history of the *Psychological Bulletin* (Sternberg, 1992). In formulating what they termed the multitrait–multimethod (MTMM) matrix, Campbell and Fiske laid out a technique that would assist in determining the validity of psychological tests. They believed this test-focused procedure would be a necessary element in the process of explicating the nomological network that supports the construct validity for a test (Cronbach & Meehl, 1955). Campbell and Fiske's reasoning was quite straightforward: If a test is to be considered valid, it should display convergent and discriminant validity. It should be highly correlated with scales of the same construct obtained from an independent method of measurement and it should be uncorrelated with theoretically independent constructs—especially when those independent constructs are quantified by the same method of measurement.

Although they did not explicate the terminology, Campbell and Fiske's MTMM matrix presupposed three components that sum to form any observed test score: (a) actual measurement of the intended construct; (b) systematic measurement error, or the consistent but erroneous measurement of unintended factors; and (c) random measurement error, or the irregularities arising from specific content, settings, occasions, or subject states that cause scores to fluctuate over theoretically independent measurements.

To the extent observed scores are determined by actual measurement of the intended construct rather than systematic or random error, Campbell and Fiske reasoned that two independent methods of measuring the same construct should display strong convergent validity. It is this reasoning that has apparently led to the expectation that unique methods of assessment should provide redundant information. However, Campbell and Fiske (1959) also discussed caveats regarding convergent validity coefficients. Two issues they mentioned are quite relevant for understanding the lack of association between MMPI and Rorschach scores. The first concerns the nature of the constructs we try to measure in personality assessment, whereas the second was a major thrust of their article and concerns the pervasive influence of method variance.

### Personality Constructs

A critical requirement for the MTMM matrix is that constructs must be uniformly expressed across levels of personality. Uniformity is essential because it allows the construct to be equally measurable by all methods of assessment. Campbell and Fiske (1959) cautioned researchers about assuming this requirement has been met. In fact, they noted how a failure to demonstrate convergence across methods could

lead to a refined understanding of the personality construct, rather than to the criticism or abandonment of a test.

Unfortunately, when a test does not display convergent validity, the test is generally considered suspect, not the construct. For instance, if depression scales from the MMPI and another self-report test have strong convergent validity but neither set of measures correlate with the *DEPI*, then the Rorschach scale is considered dubious rather than our notions regarding depression. The underlying assumption is that people who appear depressed on one method should also appear depressed on all other methods. This assumption, however, is grounded in a one-dimensional view of human nature that holds it is impossible for people to be "depressed" and "not depressed" at the same time. Superficially, this belief is logical and it is true in a strict sense if depression is defined as the specific behavioral symptoms contained in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM-IV*]; American Psychiatric Association, 1994). However, this belief does not always conform to clinical reality.

Not only can patients appear depressed on one test and not another because of motivated distortions, but more important, a variety of clinical conditions can be identified that exemplify how affective states and psychiatric symptoms are not unitary across levels of personality. Identifying such conditions necessarily requires a more psychodynamic view of personality and psychopathology (e.g., Gabbard, 1990). For instance, consider patients with somatization, conversion, or somatoform pain disorders. Most of these people are not consistently aware of the distress, dysphoria, or psychic conflict that accompanies their condition and their underlying emotional discomfort is rarely expressed through direct and overt dysphoric symptomatology. Thus, patients with these difficulties will generally not report depression on the MMPI (nor a structured interview). However, depression may well emerge on Rorschach-derived scores if the method can truly assess underlying personality dynamics and propensities.

Many other kinds of patients (and nonpatients) protect their self-image and conscious experience from emotional distress by using defenses of denial, repression, intellectualization, splitting, and/or projection. Some have conditions where their overt symptoms can be understood as compensatory defenses for opposing unconscious needs or feelings. People with these difficulties may not consciously recognize or "own" their troubling affects, even though these emotional states are still part of their underlying psychological experience. For example, the grandiosity and excessive need for admiration displayed by the narcissistic patient can be understood as defensive counter-reactions to the underlying feelings of inadequacy, shame, and inferiority that are intolerable for the patient to bear consciously (Kohut, 1971, 1984). The grandiosity implicit in paranoid ideas of reference can serve as compensation for underlying feelings of inadequacy, whereas paranoid beliefs that harm and danger are ever-present in the environment can reflect underlying anger that has been disavowed and projected externally (Akhtar, 1990). The schizoid



individual who overtly expresses minimal affect and little or no desire for social engagement can often be understood as having a defensive counter-reaction to acutely sensitive emotional reactions and powerful desires for warm, close, fusing interpersonal relationships (Akhtar, 1987; Gabbard, 1990; Guntrip, 1968).

Although not exhaustive, the foregoing illustrates several conditions that should create disparities when data are obtained from different methods of personality assessment. If the MMPI can provide valid data drawn from conscious self-awareness and verbally mediated self-representations and if the Rorschach can provide valid data about tacit self- and other-representations; unreflected upon modes of perceiving the world; and underlying feelings, conflicts, and dynamics, then these methods should disagree when they are assessing the kinds of clinical conditions described earlier. Therefore, a second reason why the MMPI and Rorschach disagree is because certain clinical conditions can force the scores to go in opposite directions when each test is accurately measuring what it should.

### Method Variance

In their original work, Campbell and Fiske (1959) noted how intrinsic features of any measurement tool interfered with the ability to accurately quantify a targeted construct. These method factors were understood as pervasive confounds affecting all tests, including the halo effect in observer ratings, apparatus factors in studies with laboratory animals, and so on. Method variance is a major component of systematic measurement error and is evident in a MTMM matrix when there is poor convergent validity across methods and poor discriminant validity within a method. Of importance, Campbell and Fiske (1959) observed that method factors typically account for a greater proportion of variance in test scores than the actual traits the tests are designed to measure. When observed scores are determined more by the systematic error of method variance than by the intended construct, personality research is obviously compromised.

All of the factors discussed in Table 1 can contribute to method variance. For instance, unskilled examiners can generate considerable systematic error in Rorschach scores, as can poor reading comprehension in MMPI scores. However, it is useful to expand on some common individual differences that generate method variance because these response style factors are ubiquitous in clinical settings.

When completing the MMPI, some patients are situationally guarded, nondisclosing, and consciously responding to items in a socially desirable fashion (Style 1-M). Others are characterologically defended, lacking insight and self-awareness, and unable to accurately describe themselves. These patients are prone to portray themselves as overly virtuous or ideal and to have minimal awareness of affective distress (Style 2-M). Others are quite open, insightful, and forthcoming about their problems or lack thereof (Style 3-M). Still others are situationally dramatic in their

presentation and strive to portray themselves in an unrealistically pathological fashion (Style 4-M). Finally, other patients are characterologically fragile, self-critical, or hypersensitive to distress. They will persistently—but honestly—think of their personality glass as half-empty rather than half-full, and experience themselves in an overly pathological and symptomatic light (Style 5-M).

A similar range of response styles are readily observed with the Rorschach. Some patients are situationally quite constricted and leery about what information will be obtained from them. They offer few responses, minimal investment in the task, resistive inquiry, and minimal determinant articulation (Style 1-R). Others have similar overt behaviors, although the root cause is very different, emanating from genuinely limited or impoverished internal resources and bewilderment with the abstractness or complexity of the task (Style 2-R). Other patients are articulate, strive to work cooperatively with the examiner, maintain active engagement with the task and spontaneously give responses, articulate determinants, and synthesize blot locations in accordance with their internal predilections (Style 3-R). Still others consciously strive to be dramatic, exhaustive, creative, or shocking in their responses to the test. Consequently, they appear very engaged and generate many structurally complex, synthesized responses (Style 4-R). Finally, some other patients also appear highly engaged with the task, producing lengthy and complex protocols. However, their engagement is driven by an unconscious propensity to unload chaotic internal states, as if the task opens their fragile psychological boundaries and unleashes a stream of emotionally laden perceptions (Style 5-R).

Even though MMPI and Rorschach scores can demonstrate validity in nomothetic research, four of these five response styles infuse observed scale scores with an extensive degree of systematic error—positive error when scores are inflated (Styles 4 and 5) and negative error when scores are deflated (Styles 1 and 2). Furthermore, even though these types of error are not mutually exclusive, they can emerge from two very different sources: either from genuine aspects of character or from manipulated forms of communication (also see Paulhus, 1984, 1986). That is, positive or negative systematic error can be due to authentic features of a patient's intrinsic states and traits (i.e., Styles 2 and 5) or to conscious efforts that are designed to convey a particular message or achieve a desired end (i.e., Styles 1 and 4).

In clinical practice, it is clear these different response styles affect the validity of obtained scores. A built-in strength of the MMPI is that it contains scales to assess these forms of systematic error. Although a *T* score of 40 on Scale 2 generally suggests energetic optimism and cheerful self-confidence, the meaning of this score can be quite different depending on other features of the profile. When *L*, *F*, and *K* and Scales 1, 2, and 3 each have a "V" rather than caret shape, the *T* score of 40 on Scale 2 really means the patient has significant struggles with underlying depressive issues, even though these are consciously disavowed and contributing to a somatoform condition (e.g., Greene, 1991).

Less attention has been paid to response styles on the Rorschach. However, in their classic work on *Psychodiagnostic Testing*, Rappaport, Gill, and Schafer (1968) labeled the disengaged or constricted approach *coarctated* and labeled the engaged approach *dilated*. They saw this dimension of responding to the test as one of the most central dimensions to consider when interpreting a protocol. Their reasoning was similar to the conventional wisdom found with an MMPI; they believed the meaning of a score elevation was altered depending on whether the protocol was coarctated or dilated.

In applied practice, clinicians can recognize validity does not reside solely within a particular test scale or index, but rather is also a function of how patients interact with the task. Thus, clinicians can derive some accurate inferential knowledge about a patient when considering overall MMPI or Rorschach profiles, although this same knowledge cannot be gained by considering single scale elevations that disregard response styles. Nomothetic research, on the other hand, must typically disregard the configural and contextual information that alerts clinicians to response style influences and modifies the meaning of scale scores. This is because statistics are generally calculated from scales considered in isolation. Such an approach implicitly assumes validity is solely a function of test scales. These nonveridical data points must then bear the burden of demonstrating validity for the scale, for the test as a whole, and for the conceptual reasoning that would accompany skilled use of the instrument. Such an approach is part of the reason why it is generally impossible to separate method variance from trait variance when conducting nomothetic research (also see Cronbach, 1995).

To what extent do response styles affect observed scores on the MMPI and Rorschach? Consistent with Campbell and Fiske's (1959) observation, the first and largest dimension within each test can be considered a response style dimension (Edwards & Edwards, 1991; Meyer, 1992a, 1992b). The first MMPI dimension generally accounts for about 50% of the total test variance, and about 75% to 80% of the common variance (Edwards & Edwards, 1991).<sup>3</sup> The first Rorschach dimension typically accounts for about 30% of the total variance in test scores, and about 50% of the common variance (Meyer, 1992a, 1992b). Although it is not possible to specify what proportion of each dimension may be due exclusively to method

---

<sup>3</sup>The first dimension of the MMPI is more typically given a trait interpretation rather than a method interpretation. As a trait it is considered a dimension of negative affectivity or distressed emotionality (e.g., Johnson, Butcher, Null, & Johnson, 1984). Although it would be clinically and theoretically important to try to disentangle where response styles end and negative affectivity begins, it may be impossible to accomplish and is certainly beyond the scope of this article. Suffice it to say that this dimension responds to three very different but deeply intertwined qualities. The first concerns the authentic but objectively inaccurate experience that some people have regarding their own level of disturbance (Styles 2 and 5). The second concerns conscious efforts to under- or over-report psychic problems (Styles 1 and 4). The third concerns the theoretically "accurate" trait information that could be quantified by this dimension (approximated by Style 3).

variance rather than intended trait variance, it is clear response styles have a pervasive influence on each test.

Disregarding response styles is not so problematic when research is conducted within a single assessment method. In fact, a rosier picture is obtained under these circumstances because response styles should be consistent within a method family and falsely magnify convergent validity coefficients. That is, the systematic error generated by response styles will consistently sweep scores for a construct higher (or lower) than they should be, causing the validity coefficients for two tests to be artificially inflated.

It seems unlikely, however, that response styles will operate in a consistent fashion across different methods of assessment. The factors that make someone report unrealistically virtuous ideas about him or herself on a self-report questionnaire are probably not the same factors that make someone constricted when sitting down with an examiner to articulate what ambiguous inkblots look like. Therefore, a third reason why we have not found convergent validity between the MMPI and Rorschach is because response styles operate independently on each test and tend to obscure the true extent of construct overlap.

### Implications

To summarize, several reasons have been proffered as to why seemingly similar MMPI and Rorschach scales do not correlate with each other. A primary reason may be the scales actually measure different constructs, despite similar names. A second reason may be a function of personality complexity. For many clinical conditions, there should be a disparity between those characteristics that are consciously recognized or deliberately reported and those characteristics that are present on a more implicit level of personality. Finally, because response styles have a significant impact on the scores obtained from both methods, these factors may obscure the true extent of construct convergence when methods are as different as the MMPI and Rorschach.

Taking these ideas into account led to the following hypotheses: (a) the MMPI and Rorschach are sufficiently distinct so that response styles should be uncorrelated across methods; (b) response styles should be highly correlated when different tests come from the same method family; (c) consistent with previous research, no substantial correlations should be evident between theoretically similar Rorschach and MMPI-2 scales when response styles are ignored; (d) rather than fighting the confounding effects of method variance, if the analysis is limited to patients with similar response styles on each method, there should be strong positive correlations between Rorschach indices and corresponding scales from the MMPI-2; and (e) there should be strong negative correlations between Rorschach indices and corresponding MMPI-2 scales when patients display discordant response styles on each method.

## METHOD

### Participants

As part of a hospital-based psychological testing program, 370 patients completed both an MMPI-2 and Rorschach. However, 8 participants were excluded from the analyses; 2 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 single analysis within the self-report method family, data were obtained from 283 patients who had completed an Millon Clinical Multiaxial Inventory (MCMI-II; Millon, 1987) and an MMPI-2. For the primary sample of 362 patients, 52% were psychiatric inpatients, 30% were psychiatric outpatients, 15% were general medical patients, and 3% were drawn from other settings. The average age of the sample was 34.9 ( $SD = 11.5$ , range = 17 to 72); 55% were women; 60% were White, 32% African American, 4% Asian, and 4% other.

### Measures

All Rorschachs were either administered by myself or by a student in training with me. I scored all protocols according to Comprehensive System guidelines (Exner, 1993), and most were also scored by a trainee, with discrepancies resolved through discussion. Reliability for this database was calculated across 63 protocols and found to be satisfactory. Percentage of exact agreement between myself and an independent coder was determined across the following categories: location and space (94%), developmental quality (88%), movement (90%), shading and achromatic color (94%), color (95%), form-dominance of determinants (89%), color shading or shading blends (97%), all determinants exact (74%), form quality (82%), pairs (95%), content (85%), popular (97%), organizational activity (88%), special scores that form the weighted sum of six special scores (82%), and other special scores (88%). All MMPI-2s or MCMI-IIs were administered at roughly the same time as the Rorschach and computer-scored by National Computer Systems (NCS). Scores were obtained from the NCS files except for 4 patients who had only MMPI-2 profile sheets available.

Dependent variables were considered in three content clusters: affective distress, psychotic processes, and wary interpersonal suspiciousness. Rorschach measures of emotional distress included the Depression Index ( $DEPI$ ) and the Suicide Constellation ( $S-CON$ ). MMPI-2 variables included Scale 2, Scale 7, Depression ( $DEP$ ), Anxiety ( $ANX$ ), and the Negative Emotionality/Neuroticism scale from the Personality Psychopathology Five ( $PSY-5-Neg$ ; Harkness, McNulty, & Ben-Porath, 1995). Despite real differences in the constructs being measured by these two methods, I believed that fairly strong construct overlap was still present.

The Rorschach measure of psychotic processes was the Schizophrenia Index (*SCZI*), whereas MMPI-2 measures included Scale 8, Bizarre Mentation (*BIZ*), and the Psychoticism scale from the *PSY-5* (*PSY-5-Psy*; Harkness et al., 1995). I believed these cross-method constructs had moderate correspondence.

The Rorschach measure of interpersonal wariness was the Hypervigilance Index (*HVI*). The MMPI-2 does not have any scales that measure interpersonal wariness in a way that is parallel to the *HVI*'s construct of emotional detachment combined with heightened alertness and extensive processing of environmental cues. Nonetheless, a range of MMPI-2 scales were selected, each of which address an oblique manifestation of the Rorschach construct. The MMPI-2 measures 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). I believed these constructs would have mild cross-method correspondence.

Total scores were used for Rorschach scales and non-*K*-corrected raw scores were used for MMPI-2 scales. Rorschach distributions were approximately normal, with no scale having skew or kurtosis values exceeding |0.86|. Several MMPI-2 variables (i.e., *PSY-5-Psy*, *PSY-5-Neg*, *ANX*, and *BIZ*) had skew or kurtosis values exceeding |1.0|, but no values exceeded |4.0| and all variables were retained for the correlational analysis.

## Procedures

Welsh's Anxiety Scale (*A*) was designed to quantify the first MMPI factor. To assess its adequacy, I conducted a principal components analysis of this scale along with the MMPI-2 basic, validity, and content scales. Using 470 participants from the full database, the first unrotated component accounted for 51.3% of the total variance (the second, third, fourth, and fifth components accounted for 11.7, 5.9, 4.1, and 3.7% of the total variance, respectively). As expected, this dimension was defined by the *A* scale, which had a loading of .95. These findings indicate *A* is a good measure of the MMPI-2's huge first factor.

The first factor from the Rorschach has been termed *Response Articulation* in the past (Meyer, 1992b), although I now consider *Response-Engagement* (*R-Engagement*) or response complexity to be more accurate descriptors. Previously, a formula for calculating *R-Engagement* was derived from a large sample of college students (Meyer, 1992b) 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*). To assess the adequacy of this scale, I conducted a principal components analysis of this variable along with other nonredundant scores for location, developmental quality, determinants, form quality, and special scores. Using 430 participants from the full database, the first unrotated component accounted for 23.4% of the total variance (the second, third, fourth, and fifth components

accounted for 8.9, 4.8, 4.4, and 3.5% of the total variance, respectively). As expected, this dimension was most strongly defined by the *R-Engagement* scale, which had a loading of .96. These findings indicate this scale is a good measure of the Rorschach's large first factor.

Using 316 participants, the first MCMI-II factor accounted for 57% of the total test variance (the second, third, fourth, and fifth components accounted for 17.9, 7.6, 6.2, and 3.2% of the total variance, respectively). This factor was most strongly defined by the *Disclosure* validity scale, which had a factor loading of .98.<sup>4</sup>

For the MMPI-2, participants were considered openly responsive (Style 4-M or 5-M) if they scored in the upper third of the *A* distribution (i.e., > 21), and considered defensively constricted (Style 1-M or 2-M) if they scored in the lower third (i.e., < 11). Using these criteria, 126 participants were considered expressive, whereas 124 were considered constricted. For the Rorschach, participants were considered openly responsive (Style 4-R or 5-R) if they scored in the upper third of the *R-Engagement* distribution (i.e., > .728), and were considered guardedly constricted (Style 1-R or 2-R) if they fell in the lower third of this distribution (i.e., < -1.103). Using these criteria, there were 121 participants classified as openly responsive on the Rorschach and 121 participants classified as constricted. Examining response styles across methods, 87 participants adopted the same style on both the MMPI-2 and Rorschach (i.e., openly responsive on both,  $n = 46$ , or defensive on both,  $n = 41$ ), whereas 78 participants adopted discordant response styles across methods (openly responsive Rorschach and constricted MMPI-2,  $n = 35$ , or constricted Rorschach and expressive MMPI-2,  $n = 43$ ).

## RESULTS

The first hypothesis proposed response styles on the Rorschach and MMPI would be independent. Table 2 demonstrates this hypothesis was supported. Whether operationally defined by the *A* and *R-Engagement* scales, or by factor scores derived in this sample, the MMPI's first factor is uncorrelated with the Rorschach's first factor ( $r_s = .10$  and  $.04$ , respectively). Secondarily, it was expected that other response style indicators from the MMPI-2 (i.e., *L*, *F*, *K*, and *Fb*) and Rorschach (i.e., *R* and *Lambda*) would be unrelated. The data in Table 2 support this hypothesis as well, as these scales have an average correlation of .01.

The second hypothesis stated response styles should be strongly correlated within the same method family. This hypothesis was tested by correlating the first factor from the MMPI-2 with the first factor from the MCMI-II and also by correlating the *L*, *F*, *Fb*, and *K* validity scales from the MMPI-2 with the *Disclosure* and *Debasement* validity scales from the MCMI-II. These correlations are presented in Table 3. In contrast to the previous analysis, the first dimension of the

<sup>4</sup>Factor solutions for each of these analyses are available on request.

TABLE 2  
Correlation Between Rorschach Scales of Response Style  
and MMPI-2 Scales of Response Style

Rorschach Scale	MMPI-2 Scale					
	A	Ist Fact	F	Fb	L	K
<i>R-Engagement</i>	.10	.05	.05	.02	-.10	-.02
<i>Ist Factor</i>	.08	.04	-.09	.02	-.01	.04
<i>R</i>	.05	.02	.02	.01	-.07	.00
<i>Lambda</i>	.05	.09	.06	.08	.09	-.11*

Note. *R-Engagement* refers to the first principal component of the Rorschach using factor scores derived from a sample of college students (Meyer, 1992b). *Ist Factor* refers to the same dimension quantified by factor scores derived from the present sample. *Ist Fact* refers to factor scores for the first principal component of the MMPI-2 derived from the present sample.  $N = 362$  for all correlations except those with *Fb*, where  $N = 359$ .

\* $p < .05$ .

TABLE 3  
Correlation Between MMPI-2 Scales of Response Style  
and MCMI-II Scales of Response Style

MCMI-II Scale	MMPI-2 Scale					
	Ist Factor	A	F	Fb	L	K
<i>Ist Factor</i>	.85*	.78*	.68*	.73*	-.44*	-.76*
<i>Disclosure</i>	.80*	.74*	.64*	.67*	-.44*	-.74*
<i>Debasement</i>	.87*	.85*	.69*	.80*	-.39*	-.68*

Note. *Ist Factor* refers to factor scores on the first unrotated principal component derived from the present sample.  $N = 283$  for all correlations.

\* $p < .001$ .

MMPI-2 is strongly correlated with the first dimension of the MCMI-II ( $r = .85$ ). Furthermore, the clinically derived indicators of response style from each test are also strongly related, having an average correlation of .63. Table 3 also indicates the clinical markers of response style from one test are strongly correlated with the first component from the other test. For instance, the MMPI-2's first factor is correlated with the MCMI-II *Debasement* and *Disclosure* scales at magnitudes of .87 and .80, respectively. These findings offer additional support for considering each test's first factor to be largely determined by response styles.

The third hypothesis postulated that there would be no association between Rorschach and MMPI scales of similar constructs when response styles were disregarded. This hypothesis was also supported. As can be seen in Table 4, despite



considerable statistical power, there are no non-zero correlations between MMPI-2 and Rorschach scales of emotional distress. The average convergent validity coefficient for this construct is  $-.006$ . Turning to scales of psychosis and wariness, there are again no salient correlations, although one low magnitude coefficient is statistically significant in each case. Nonetheless, the general pattern is clear. For all practical purposes, when one disregards response styles in a heterogeneous sample, MMPI and Rorschach scales of affective distress, psychotic processes, and interpersonal wariness are unrelated.

The fourth hypothesis stated Rorschach and MMPI scales of similar constructs should be positively correlated when the analysis is limited to those participants who display the same type of response style on both assessment methods. Table 5 presents the results of this analysis. It can be seen there is substantial overlap between scores of emotional discomfort when patients have similar response styles on both methods. In general, the Rorschach's *DEPI* and *S-CON* correlate with MMPI-2 scales in the range of  $.55$  to  $.65$  ( $M = .594$ ). A similar pattern of strong positive correlations is apparent when the constructs of psychosis and interpersonal wariness are examined. The Rorschach Schizophrenia Index now correlates in the range between  $.45$  and  $.55$  with corresponding MMPI-2 scales. The correlations for wariness are a bit more variable, although still substantial ( $M r = .37$ ). Overall, these data indicate that under some circumstances, strong convergent validity can be demonstrated for these very different methods of assessment.

The fifth hypothesis stated scales of similar constructs should be negatively correlated when patients displayed different response styles on each method. The correlations for this analysis are presented in Table 6. The data clearly support the hypothesis for the construct of emotional distress, as these correlations cluster around a magnitude of  $-.55$ . Similar, but less pronounced disparities were evident for the constructs of psychosis and interpersonal wariness. The Rorschach Schizophrenia Index was negatively correlated with each of its corresponding MMPI-2 scales, although the findings were not statistically significant using the more conservative two-tailed rejection region. Negative correlations of slightly larger magnitude ( $M = -.28$ ) were observed when the *HVI* was examined in relation to MMPI-2 scales of interpersonal wariness. The findings are certainly most robust for the construct of depression, although the general pattern supports the hypothesis that constructs will be negatively correlated across methods when patients display opposing response styles on each method.

### Additional Analyses

MTMM matrices were constructed in order to place the previous monotrait-heteromethod convergent validity findings within a broader context. Tables 7, 8, and 9 present a condensed matrix of the correlations observed when response styles are

**TABLE 4**  
**Disregarding Response Styles: Correlation Between Rorschach and MMPI-2 Scales in Three Content Areas When Response Styles Are Ignored**

<i>Affective Distress</i>			<i>Psychosis</i>		<i>Wariness</i>	
<i>MMPI-2 Scales</i>	<i>Rorschach DEPI</i>	<i>Rorschach S-CON</i>	<i>MMPI-2 Scales</i>	<i>Rorschach SCZI</i>	<i>MMPI-2 Scales</i>	<i>Rorschach HVI</i>
Scale 2	-.09	-.05	Scale 8	.15*	Scale 6	.15*
Scale 7	.00	-.01	<i>BIZ</i>	.09	<i>CYN</i>	-.07
<i>DEP</i>	.01	-.00	<i>PSY-5-Psy</i>	.09	<i>SOD</i>	.04
<i>ANX</i>	.02	.04			<i>TRT2</i>	.08
<i>PSY-5-Neg</i>	-.00	.02				

*Note.* *DEPI* = Depression Index; *S-CON* = Suicide Constellation; *SCZI* = Schizophrenia Index; *HVI* = Hypervigilance Index; *DEP* = Depression Content Scale; *ANX* = Anxiety Content Scale; *PSY-5-Neg* = Personality Psychopathology Five Negative Emotionality/Neuroticism Scale; *BIZ* = Bizarre Mentation Content Scale; *PSY-5-Psy* = Personality Psychopathology Five Psychoticism Scale; *CYN* = Cynicism Content Scale; *SOD* = Social Discomfort Content Scale; *TRT2* = Inability to Disclose component of the Negative Treatment Indicators Content Scale. *N* = 362 for Scales 2, 6, 7, and 8; *N* = 359 for Content Scales; *N* = 358 for *PSY-5* scales.

\**p* < .01.

**TABLE 5**  
**Similar Response Styles: Correlation Between Rorschach and MMPI-2 Scales in Three Content Areas When Response Styles Are Similar on Each Assessment Method**

<i>Affective Distress</i>			<i>Psychosis</i>		<i>Wariness</i>	
<i>MMPI-2 Scales</i>	<i>Rorschach DEPI</i>	<i>Rorschach S-CON</i>	<i>MMPI-2 Scales</i>	<i>Rorschach SCZI</i>	<i>MMPI-2 Scales</i>	<i>Rorschach HVI</i>
Scale 2	.42**	.50**	Scale 8	.54**	Scale 6	.45**
Scale 7	.62**	.67**	<i>BIZ</i>	.46**	<i>CYN</i>	.22*
<i>DEP</i>	.66**	.63**	<i>PSY-5-Psy</i>	.48**	<i>SOD</i>	.38**
<i>ANX</i>	.61**	.62**			<i>TRT2</i>	.44**
<i>PSY-5-Neg</i>	.59**	.62**				

*Note.* *DEPI* = Depression Index; *S-CON* = Suicide Constellation; *SCZI* = Schizophrenia Index; *HVI* = Hypervigilance Index; *DEP* = Depression Content Scale; *ANX* = Anxiety Content Scale; *PSY-5-Neg* = Personality Psychopathology Five Negative Emotionality/Neuroticism Scale; *BIZ* = Bizarre Mentation Content Scale; *PSY-5-Psy* = Personality Psychopathology Five Psychoticism Scale; *CYN* = Cynicism Content Scale; *SOD* = Social Discomfort Content Scale; *TRT2* = Inability to Disclose component of the Negative Treatment Indicators Content Scale. *n* = 87.

\**p* < .05. \*\**p* < .001.

TABLE 6  
 Different Response Styles: Correlation Between Rorschach  
 and MMPI-2 Scales in Three Content Areas When Response Styles  
 Are Discordant Across Both Assessment Methods

<i>Affective Distress</i>			<i>Psychosis</i>		<i>Wariness</i>	
<i>MMPI-2 Scales</i>	<i>Rorschach DEPI</i>	<i>Rorschach S-CON</i>	<i>MMPI-2 Scales</i>	<i>Rorschach SCZI</i>	<i>MMPI-2 Scales</i>	<i>Rorschach HVI</i>
Scale 2	-.51**	-.43**	Scale 8	-.22	Scale 6	-.22
Scale 7	-.59**	-.57**	<i>BIZ</i>	-.16	<i>CYN</i>	-.38*
<i>DEP</i>	-.55**	-.52**	<i>PSY-5-Psy</i>	-.20	<i>SOD</i>	-.23*
<i>ANX</i>	-.56**	-.57**			<i>TRT2</i>	-.27*
<i>PSY-5-Neg</i>	-.56**	-.57**				

Note. *DEPI* = Depression Index; *S-CON* = Suicide Constellation; *SCZI* = Schizophrenia Index; *HVI* = Hypervigilance Index; *DEP* = Depression Content Scale; *ANX* = Anxiety Content Scale; *PSY-5-Neg* = Personality Psychopathology Five Negative Emotionality/Neuroticism Scale; *BIZ* = Bizarre Mentation Content Scale; *PSY-5-Psy* = Personality Psychopathology Five Psychoticism Scale; *CYN* = Cynicism Content Scale; *SOD* = Social Discomfort Content Scale; *TRT2* = Inability to Disclose component of the Negative Treatment Indicators Content Scale. *n* = 78.

\**p* < .05. \*\**p* < .001.

ignored, similar, and discordant, respectively. Stronger evidence for convergent validity is indicated when test scales also display discriminant validity, or no relation with theoretically independent constructs. It is easiest to examine discriminant validity when constructs should be fully independent (i.e., *r* = 0), because then any empirical overlap can be attributed to systematic error. However, when the constructs should overlap—when affective distress should be associated with interpersonal cautiousness, for instance—it is more difficult to determine the extent to which method variance is confounding genuine trait variance.

Within a MTMM matrix, evidence for discriminant validity is obtained from three sources. The first is discriminant validity within a single method. This information is derived from the italicized heterotrait-monomethod coefficients given in each table. Although the values “wobble” a bit moving from table to table, there is consistently much more overlap among the three constructs when they are measured by the MMPI. For instance, in Table 7 the weighted average correlation among the different constructs is .57 for the MMPI-2 and .32 for the Rorschach. In Table 9 the values are .68 for the MMPI-2 and .26 for the Rorschach. The stronger correlations among the MMPI scales probably reflect the larger and more pervasive impact of response styles on the MMPI, which cause almost all scales to vary together and make it more difficult to discriminate appropriately among constructs.

Evidence for discriminant validity is also obtained when the bolded convergent validity coefficients are larger than the italicized heterotrait-monomethod coefficients. Examining Table 8, it is clear this criterion is not met if the bolded convergent

TABLE 7  
 Disregarding Response Styles: An Averaged Multitrait-Multimethod  
 Summary Matrix on the Relation Between Rorschach  
 and MMPI-2 Scales of Affective Distress, Psychotic Processes,  
 and Interpersonal Wariness

	<i>Rorschach Constructs</i>			<i>MMPI-2 Constructs</i>		
	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>
R-Distress	(.58)					
R-Psychosis	.24**	—				
R-Wariness	.34**	.43**	—			
M-Distress	<b>-.01</b>	<u>.11*</u>	<u>.10</u>	(.81)		
M-Psychosis	<u>-.02</u>	<b>.11*</b>	<u>.02</u>	.63**	(.79)	
M-Wariness	<u>.01</u>	<u>.09</u>	<b>.02</b>	.54**	.54**	(.36)

*Note.* R-Distress = Rorschach emotional distress, reflecting the average correlates of the Depression Index and the Suicide Constellation; R-Psychosis = the Rorschach Schizophrenia Index; R-Wariness = the Rorschach Hypervigilance Index; M-Distress = MMPI-2 emotional distress, reflecting the average correlates of Scale 2, Scale 7, the Depression and Anxiety Content Scales, and the Negative Emotionality/Neuroticism Scale; M-Psychosis = MMPI-2 psychosis, derived from the average correlates of Scale 8, the Bizarre Mentation Content Scale, and the Psychoticism Scale; M-Wariness = MMPI-2 interpersonal wariness, reflecting the average correlates of Scale 6, the Cynicism and Social Discomfort Content Scales, and the Inability to Disclose scale. Values within parentheses are "reliability" coefficients reflecting the average of the monotrait-monomethod correlations, coefficients in bold are convergent validity coefficients (monotrait-heteromethod), italicized coefficients are heterotrait-monomethod discriminant validity correlations, and underlined coefficients are heterotrait-heteromethod discriminant validity correlations.  $N = 362$  for Scales 2 and 7;  $N = 359$  for Content Scales,  $N = 358$  for PSY-5 scales.

\* $p < .05$ . \*\* $p < .001$ .

validity coefficients are compared to the MMPI-2's italicized discriminant validity coefficients. Because response styles make it difficult for the MMPI to discriminate among constructs, they also make it difficult for the MMPI to demonstrate differentiated convergent validity. However, the picture is better when the contrast is with the Rorschach's discriminant validity correlations. Table 8 indicates affective distress can be differentiated from other Rorschach constructs, although psychosis has difficulty in this regard and wariness cannot. The final source of evidence for discriminant validity is obtained when the bolded convergent validity coefficients are larger than the underlined heterotrait-heteromethod coefficients found in the same row and column. Here again the analysis in Table 8 provides evidence supporting the discriminant validity of affective distress, but marginal evidence for psychosis, and nonsupporting evidence for wariness. Undoubtedly, the validity coefficients for the wariness construct are attenuated by the diverse MMPI-2 scales used to operationally define this construct.

**TABLE 8**  
**Similar Response Styles: An Averaged Multitrait–Multimethod**  
**Summary Matrix on the Relation Between Rorschach**  
**and MMPI–2 Scales of Affective Distress, Psychotic Processes,**  
**and Interpersonal Wariness**

	<i>Rorschach Constructs</i>			<i>MMPI–2 Constructs</i>		
	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>
R–Distress	(.60)					
R–Psychosis	.42*	—				
R–Wariness	.46*	.45*	—			
M–Distress	<b>.59*</b>	<b>.48*</b>	<b>.57*</b>	(.86)		
M–Psychosis	<b>.47*</b>	<b>.49*</b>	<b>.36*</b>	.68*	(.82)	
M–Wariness	<b>.47*</b>	<b>.41*</b>	<b>.37*</b>	.60*	.60*	(.42)

*Note.* R–Distress = Rorschach emotional distress, reflecting the average correlates of the Depression Index and the Suicide Constellation; R–Psychosis = the Rorschach Schizophrenia Index; R–Wariness = the Rorschach Hypervigilance Index; M–Distress = MMPI–2 emotional distress, reflecting the average correlates of Scale 2, Scale 7, the Depression and Anxiety Content Scales, and the Negative Emotionality/Neuroticism Scale; M–Psychosis = MMPI–2 psychosis, derived from the average correlates of Scale 8, the Bizarre Mentation Content Scale, and the Psychoticism Scale; M–Wariness = MMPI–2 interpersonal wariness, reflecting the average correlates of Scale 6, the Cynicism and Social Discomfort Content Scales, and the Inability to Disclose scale. Values within parentheses are “reliability” coefficients reflecting the average of the monotrait–monomethod correlations, coefficients in bold are convergent validity coefficients (monotrait–heteromethod), italicized coefficients are heterotrait–monomethod discriminant validity correlations, and underlined coefficients are heterotrait–heteromethod discriminant validity correlations. *N* = 87.

\**p* < .001.

Overall, although convergent validity can be demonstrated for MMPI and Rorschach constructs under certain conditions, the findings indicate discriminant validity is more difficult to attain. This appears to be a particular problem for the MMPI because method variance seems to cause poor discrimination among constructs and also hamper cross-method validation.

I also examined whether response styles could be treated as a single uniform dimension. If so, one could statistically partial response style effects from the scale scores and anticipate positive cross-method validity coefficients. However, after partialling first-factor variance from each scale, there was no evidence for convergent validity across methods. For the constructs of affective distress, psychosis, and wariness, the convergent correlations averaged  $-.02$  (range  $-.12$  to  $.07$ ). When a variable quantifying the interaction of Rorschach and MMPI–2 response styles was also partialled, convergent correlations remained at zero ( $M = -.02$ , range  $-.11$  to  $.07$ ).

TABLE 9  
 Different Response Styles: An Averaged Multitrait-Multimethod  
 Summary Matrix on the Relation Between Rorschach  
 and MMPI-2 Scales of Affective Distress, Psychotic Processes,  
 and Interpersonal Wariness

	<i>Rorschach Constructs</i>			<i>MMPI-2 Constructs</i>		
	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>	<i>Distress</i>	<i>Psychosis</i>	<i>Wariness</i>
R-Distress	(.67)					
R-Psychosis	.13	—				
R-Wariness	.31**	.40***				
M-Distress	<b>-.54***</b>	<b>-.25*</b>	<b>-.39***</b>	(.88)		
M-Psychosis	<b>-.44***</b>	<b>-.19</b>	<b>-.27*</b>	.71***	(.84)	
M-Wariness	<b>-.41***</b>	<b>-.20</b>	<b>-.28*</b>	.68***	.64***	(.53)

*Note.* R-Distress = Rorschach emotional distress, reflecting the average correlates of the Depression Index and the Suicide Constellation; R-Psychosis = the Rorschach Schizophrenia Index; R-Wariness = the Rorschach Hypervigilance Index; M-Distress = MMPI-2 emotional distress, reflecting the average correlates of Scale 2, Scale 7, the Depression and Anxiety Content Scales, and the Negative Emotionality/Neuroticism Scale; M-Psychosis = MMPI-2 psychosis, derived from the average correlates of Scale 8, the Bizarre Mentation Content Scale, and the Psychoticism Scale; M-Wariness = MMPI-2 interpersonal wariness, reflecting the average correlates of Scale 6, the Cynicism and Social Discomfort Content Scales, and the Inability to Disclose scale. Values within parentheses are "reliability" coefficients reflecting the average of the monotrait-monomethod correlations, coefficients in bold are convergent validity coefficients (monotrait-heteromethod), italicized coefficients are heterotrait-monomethod discriminant validity correlations, and underlined coefficients are heterotrait-heteromethod discriminant validity correlations.  $N = 78$ .

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

## DISCUSSION

This study tested five hypotheses, four of which were strongly supported and one which was partially supported. The large degree of variance in Rorschach scores that can be attributed to response styles is uncorrelated with the extensive variance in MMPI scores that can be attributed to response styles. As such, those factors that leave patients openly responsive or inhibited on the Rorschach are independent of those factors that leave patients openly responsive or defensive on the MMPI.

Independence across methods stands in contrast to response style consistency within a method family. Not only is the primary dimension from the MMPI-2 strongly related to the primary dimension from the MCMI-II ( $r = .85$ ), but the clinical validity indicators from one test are highly related to the primary dimension from the other test, having an average cross-test correlation of  $|.72|$ . The latter findings provide strong support for interpreting the very large first dimension from each test as an index of response styles.

Because response styles have a pervasive but independent effect on MMPI and Rorschach scores, they were expected to obscure empirical construct overlap. Supporting this hypothesis, when response styles were ignored in this heterogeneous sample, Rorschach and MMPI-2 scales assessing similar constructs were unrelated. Convergent validity coefficients for measures of emotional distress, psychotic propensity, and interpersonal wariness had an average magnitude of .03. These data strongly support the conclusions reached by Archer and Krishnamurthy (1993a, 1993b) and suggest each method provides unique information. Although the Rorschach and MMPI may aspire to quantify the same general constructs (e.g., depression), the constructs that are actually measured are different entities.

Perhaps, however, there is not evidence for convergent validity because one method yields meaningless data. Those who are in the habit of dismissing the Rorschach may have been tempted to view the preceding results in this fashion. To counter such a view, this study also proposed two patterns of convergent validity that should be evident under restricted conditions.

Even though there are good reasons to believe Rorschach-derived constructs emerge from underlying or implicit propensities while MMPI-derived constructs emerge from conscious self-schema, I anticipated there would still be sufficient overlap among constructs to observe some degree of heteromethod validity when response styles were controlled. As expected, for those patients who had parallel response styles on each method, similar constructs were strongly correlated. The average validity coefficient was .59 for emotional distress, .49 for psychotic processes, and .37 for the broadly defined construct of interpersonal wariness and mistrust. Thus, certain kinds of people have rather similar scores across methods of assessment. Consistent with expectations, the effect is most pronounced for dysphoria and least pronounced for wariness.

However, the preceding analyses did not remove the influence of method variance. In fact, my attempt to equalize response style influences by using partial correlations failed. Thus, even though response styles can be expressed as variation along a single dimension, this indicates they are not a linear function of personality characteristics. Rather, response styles probably emerge from categorically distinct "types" of patients who differ qualitatively in their characterological disturbance and/or motivated desire to bias the data in a particular fashion.

The strategy taken in the convergent validity analysis was to leave method and trait variance confounded while intentionally equalizing response styles. In essence, the question addressed by the analysis was: If we hold response styles constant across methods—as is always the case when analyses are conducted with two self-report inventories or two observer rating scales—will there be convergent validity? Clearly, the data answer this question in the affirmative.

At the same time, the resulting validity coefficients should not be considered precise quantifications of the extent to which both methods actually measure their intended constructs. The validation goal articulated by Campbell and Fiske (1959)

was to demonstrate convergent validity despite unique forms of method variance confounding the correlations. This study does not achieve that goal. Furthermore, because method variance was deliberately equalized, one could argue the observed coefficients may simply reflect manipulated method variance, rather than overlap due to genuine trait variance.

Before considering this issue in more detail, it is important to recognize the same confound always exists when convergent validity is examined between two tests drawn from within the same method family. However, in these instances, error due to method variance is more frequently overlooked. Validity coefficients between conceptually similar scales from the MMPI and another self-report inventory are often viewed as bona fide evidence of the extent to which scales are measuring similar constructs. This is not true. Presumably, some part of the obtained correlation reflects a common construct. However, one can never determine how much of the validity coefficient is due to genuine trait variance and how much is due to the artificial inflation generated by self-report method variance.

Nonetheless, it is still important to consider how matching on response styles may have affected the data in this study. Even though the first factor from the MMPI-2 and the first factor from the Rorschach were initially uncorrelated (see Table 2), selecting the upper and lower thirds on these dimensions forces the primary factors to be correlated ( $r = .70$  in the matched group,  $r = -.77$  in the discordant group). Consequently, it also forces any scales that are correlated with these dimensions to be correlated in the subsequent analysis. So the question then becomes, "To what extent are the convergent correlations in Tables 5 and 6 larger than would be predicted simply from matching subjects on the upper and lower thirds of each first factor?" To determine how much of a correlation was injected into the analysis by these selection procedures, a formula from the factor analytic literature was modified to suit the current circumstances. Briefly, the formula provides the expected correlation for two variables, each of which load separately on two correlated factors.<sup>5</sup> Once the expected degree of correlation between each MMPI and Rorschach variable is calculated, the expected value can be subtracted from the observed correlation to obtain a residual. This residual correlation reflects the extent of construct overlap remaining after first-factor variance is removed. In each instance the residual correlations were lower than those reported in Table 5, although they remained of substantial magnitude. The average values for dysphoria, psychosis, and wariness were .29, .25, and .20, respectively.

There would not be a downside to this psychometric correction if one could confidently attribute all first-factor variance to response styles, rather than to actual

---

<sup>5</sup>The formula is as follows: (correlation of MMPI-2 variable  $x$  with the MMPI-2's first factor)  $\times$  (correlation of Rorschach variable  $y$  with the Rorschach's first factor)  $\times$  (correlation between Rorschach and MMPI-2 first factors within the selected subset of participants) = expected correlation for variables  $x$  and  $y$ . Thanks again to James Wood for providing this formula and for developing and running the Monte Carlo trials reported later.



trait variance. Because this is not the case, however, the residual correlations provide an underestimate of construct overlap. Nonetheless, as an estimated floor value, they provide a better coefficient for bracketing the true extent of construct overlap than the null coefficients obtained when response styles are ignored.

As a second internal check, a small Monte Carlo study was performed to estimate the impact of the selection procedures. This analysis generated two random, uncorrelated variables to represent each method's first factor. Subsequently, linear equations were used to generate simulated variables that represented each of the MMPI scales and Rorschach indices. Each simulated variable was designed to correlate properly with its "first factor" and to have the correct amount of random error variance. This simulated data was then partitioned in the same fashion as in this study (i.e., groups of simulated subjects were selected for analysis if they fell in the upper or lower third of the simulated first-factor distributions). Subsequently, each of the simulated "MMPI" and "Rorschach" variables were correlated with each other. This procedure was repeated 10 times to create 10 separate simulated data sets and 10 sets of resulting correlations. The results were then pooled to provide another estimate of the association that should be expected from the matching procedure. Although 10 runs does not provide an optimal parameter for the estimated correlations, it provided a workable approximation. As with the formula-based approach, residual correlations were created by subtracting the pooled expected values (derived from the Monte Carlo runs) from each of the observed correlations. As before, this approach treats all first-factor variance as error variance, which is not strictly warranted. Nonetheless, the results were essentially the same. Within each content domain, the residuals were substantially higher than would have been expected from mere artifact.<sup>6</sup>

Another way to address the meaningfulness of the correlations in Table 5 is to use external criteria to determine whether substantive traits are being measured by these coefficients. Optimally, one could correlate the MMPI and Rorschach scores from these patients with gold-standard criteria that quantify the constructs of affective distress, psychosis, and interpersonal wariness. Unfortunately, such criteria are not available. Nonetheless, coarse external data on patient characteristics can be considered.

*DSM* diagnoses can characterize affective distress, psychosis, and wary suspiciousness to some extent, although these are fairly crude criteria, particularly for the construct of wariness. Nonetheless, Table 10 presents data on participants who had received an externally derived diagnosis from inpatient treatment teams indicating a severe mood, psychotic, and/or personality disturbance (i.e., *DSM-IV* codes of 295.xx, 296.xx, 297.xx, 298.xx, 301.0, 301.22, or 301.83).<sup>7</sup> Data for all four MMPI-2 and Rorschach patterns are presented, although the hypothesis

---

<sup>6</sup>Because these correction procedures have other heuristic applications, more detailed findings will be presented in a forthcoming article.

TABLE 10  
 Diagnoses Indicating Severe Mood, Psychotic, or Personality  
 Disturbances for Each of the Groups Defined by the First Factors  
 From the MMPI-2 and Rorschach

MMPI-2 Group	Rorschach Group			
	Open/Engaged		Guarded/Defended	
	<i>n</i>	%	<i>n</i>	%
Open/Engaged	29/29	100.0 <sup>a</sup>	25/26	96.2
Guarded/Defended	11/13	84.6	9/21	42.9 <sup>a</sup>

Note. Diagnoses indicating a severe mood, psychotic, and/or personality disturbance included the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) codes of 295.xx, 296.xx, 297.xx, 298.xx, 301.0, 301.22, or 301.83.

<sup>a</sup>Denotes the two groups hypothesized to display a differential pattern of diagnoses;  $\chi^2(1, N = 50) = 18.78, p = .000015; r = .61$ .

focused on the contrast between the groups that displayed a similar response style on both methods. A chi-square with Yates correction revealed that those patients who scored high on the first factor from both methods were diagnosed with a severe disturbance significantly more often than those patients who scored low on the first dimension of both measures ( $p = .000015$ ; effect size expressed as  $r = .61$ ). These findings support the notion that genuine trait variance is being quantified by the convergent validity coefficients presented in Table 5.

Although hypotheses were not developed for the other two groups in Table 10, these groups do not differ in their diagnostic frequency,  $\chi^2(1, N = 39) = 0.41, p = .52$ . Importantly, however, this argues against the skeptical position that it is the MMPI in isolation that is the workhorse pulling the preceding analysis. If the Rorschach was immaterial, the guarded MMPI-engaged Rorschach group should have had a much lower frequency of these diagnostic codes.

The final hypothesis for this study anticipated negative correlations between Rorschach and MMPI constructs when patients displayed opposing response styles on each method. This hypothesis was also supported, although the effect was clearly stronger when measures were targeting a broad construct of negative affectivity ( $M r = -.54$ ), rather than psychosis ( $M r = -.19$ ) or wariness ( $M r = -.27$ ). In fact, when residual correlations were calculated by subtracting the extent of correlation expected from the matching procedure, the average lower boundary coefficient for estimating construct correspondence remained salient for dysphoria ( $-.24$ ), while it was small for psychosis ( $-.11$ ) and trivial for wariness ( $-.04$ ).

<sup>7</sup>Outpatients were excluded from this analysis because they are typically diagnosed by a single clinician rather than a treatment team. Had I included their data in the analyses, the results would have actually been more robust.

Negative cross-method correlations were expected for two clinical reasons. First, some patients were expected to have discrepant scores if they intentionally manipulated their data. Second, and more important, patients were expected to have disparities because personality characteristics are not always organized in a uniform fashion across levels of awareness. Some patients were expected to have idealized conscious beliefs about their personal attributes that would be very different from their underlying dynamics. Others were expected to genuinely report their psychic distress on the MMPI but to have characterological limitations that precluded behavioral engagement with the Rorschach. Given this is a psychiatric sample, I expected that which was denied or defended against on one method of assessment would be evident on the alternative method, as long as the patient remained openly responsive or engaged on the alternative method.

The results indicate my hypothesis was not sufficiently differentiated. The pattern does not hold across all constructs. In retrospect, this makes considerable sense. Character defenses are most often in place to ward off troubling and disruptive affective states, not to keep psychotic operations or wary hypervigilance out of awareness (e.g., Gabbard, 1990; Shedler, Mayman, & Manis, 1993; Watson & Clark, 1984). Thus, the differential support for dysphoria is most likely because affective discomfort is more heavily defended against than other constructs.

When clinicians are confronted with conflicting data on how extensively depressed a patient is, at least two possibilities must be considered. First, the communication factors that may have lead to disparities should be evaluated (Response Styles 1 and 4). This would be a primary consideration when there is some external incentive to appear healthier or more pathological than is actually true (e.g., custody evaluations, insanity defenses, etc.), as these motivations should have a particular influence on MMPI data (cf. Bornstein et al., 1994). One should also consider these factors when an unusual interpersonal dynamic may have produced a situationally guarded Rorschach protocol (e.g., an examiner who generates unusual defensiveness in a patient, etc.).

The second reason some patients present with opposing response styles on the MMPI and Rorschach is because of characterological defensive structures (Styles 2 and 5). Under these circumstances, method variance becomes synonymous with trait variance because intrinsic characteristics related to the management of negative affect actually produce the response styles. Because these patients are more important to consider for construct validity, each pattern of scores will be discussed in turn.

The first pattern consists of patients who are openly responsive on the MMPI (Styles 3-M or 5-M) but guarded or constricted on the Rorschach (Style 2-R). Many patients like this will accurately convey their emotional distress on the MMPI but not engage with the Rorschach for one of several reasons. Some may have genuinely impoverished coping resources, whereas others may be chronically guarded in an interpersonal context. Other patients, particularly those with anxiety driven problems, may defensively inhibit spontaneity and produce constricted

Rorschachs because they are defended against the in vivo experience of the distress they carry with them, unwilling or unable to engage with the troubling percepts that jump out at them on the inkblots, even though they are quite able to talk about their affective distress in a more detached or abstract fashion. Finally, some patients may be so depressed they are simply unable to muster the energy and effort that is required to engage with the Rorschach and articulate many responses and determinants, although they remain capable of describing themselves on the MMPI.

Although these suppositions refer to dynamic operations and focused characteristics more than *DSM* diagnoses, patients with this MMPI-Rorschach pattern should generally receive diagnoses indicating chronic interpersonal guardedness (i.e., *DSM-IV* codes 295.3, 297.1, or 301.00), anxiety driven problems (i.e., 300.01, 300.21, 300.22, 300.3, or 300.02), or severe and immobilizing depression (i.e., 296.23, 296.24, 296.33, 296.34, 296.53, 296.54, 296.63, or 296.64). Consistent with this expectation, inpatients who were open and forthcoming on the MMPI-2 but constricted on the Rorschach had a greater frequency of these diagnoses (18 of 26, or 69.2%) compared to patients with the opposing pattern of scores (4 of 13, or 30.8%), although this result was not quite statistically significant,  $\chi^2(1, N = 39) = 3.77, p = .052$ ; effect expressed as  $r = .31$ .

The other pattern consists of patients who are engaged with the Rorschach (Styles 3-R and 5-R) but defended on the MMPI (Style 2-M). Most of these patients genuinely need to see themselves as being as virtuous, ideal, and emotionally healthy as they report on their MMPI, despite the fact they are often beset by considerable underlying psychological distress. As before, affective distress can no longer be considered a single global construct that encompasses conscious as well as underlying experiences. Rather, affective distress is bifurcated. It is disavowed and defended against at the level of conscious awareness but quite evident and pressing in less consciously mediated behaviors.

Some of the psychological conditions that should give rise to this pattern of scores include conversion or somatoform disorders, masked depressions, or one of the many disorders that are characterized by defensive self-idealization and/or grandiosity, such as bipolar disorder (in a hypomanic to manic state); delusional disorder; or hysterical, narcissistic, antisocial, obsessive, or paranoid personality disorder. Considering only inpatients with an externally derived diagnosis, Table 11 presents data on the proportion who received one of these diagnoses for each of the four cross-method patterns (*DSM-IV* codes of 300.81, 300.11, 316.0, 307.8, 307.89, 300.7, 311, 301.0, 301.7, 301.4, 301.5, 301.81, 296.4x, and 297.1). Although data for all four groups are presented, the hypothesis focused on the contrast between the groups that displayed discordant response styles across methods. A chi-square with Yates correction revealed that those patients with defended MMPI-2s and engaged Rorschachs were significantly more likely than patients with the opposing pattern to be diagnosed with a condition where conscious awareness is protected from underlying affective distress ( $p = .00126$ ; effect size

TABLE 11  
 Diagnoses Indicating the Defensive Protection of Conscious Awareness From Affective Distress for Each Group Defined by First Factors From the MMPI-2 and Rorschach

MMPI-2 Group	Rorschach Group			
	Open/Engaged		Guarded/Defended	
	n	%	n	%
Open/Engaged	2/29	6.9	1/26	3.8 <sup>a</sup>
Guarded/Defended	7/13	53.8 <sup>a</sup>	11/21	52.4

Note. Diagnoses indicating a defensive protection against affective distress included the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) codes of 300.81, 300.11, 316.0, 307.8, 307.89, 300.7, 311, 301.0, 301.7, 301.4, 301.5, 301.81, 296.4x, and 297.1.

<sup>a</sup>Denotes the two groups hypothesized to display a differential pattern of diagnoses;  $\chi^2(1, N = 39) = 10.40, p = .00126; r = .52$ .

expressed as  $r = .52$ ). Although not an a priori hypothesis, those who were defended on both methods also received one of these diagnoses significantly more often than those who were open on both methods,  $\chi^2(1, N = 50) = 10.84, p = .00099$ ; effect size expressed as  $r = .47$ .

A Rorschach skeptic may take the full array of findings in Table 11 as an indication the MMPI tells the full story because the diagnostic pattern is similar whenever there is a guarded or defended MMPI profile. However, this is not the case. Because these patients are expected to defend against underlying depression, depression should not be evident on their MMPI although it should be present on Rorschach depressive indices. Thus, the MMPI should not quantify the full scope of their condition. As anticipated, the seven patients carrying such diagnoses who were defended on the MMPI-2 yet still engaged with the Rorschach had an average *T* score of 44.1 across the five MMPI-2 scales of affective distress. No patient had a *T* score greater than 59 on any scale. However, these patients had an average *DEPI* value of 5.14, and 57% were above the critical *DEPI* value of 5. Thus, the Rorschach can be sensitive to the underlying distress of these patients in a manner that the MMPI cannot.

Similar clinical observations can be made when considering the 11 patients who were guarded or constricted on both methods and also received a diagnosis indicating defenses against underlying affective distress. They had an average *T* score of 46.7 on the five MMPI-2 scales of dysphoria. Two patients had a *T* score greater than 64 on Scale 2 because they endorsed many somatic items, but no patient had a *T* score exceeding 58 on any of the four remaining scales. It may be recalled that a constricted Rorschach can be the result of intrinsic deficiencies that preclude engagement with the task. Exner (1993) conceptualized these deficiencies as leading to a form of helplessness-based depression, which is measured by the

Coping Deficit Index. This ineffectual form of depression differs from the implicit, cognitive-affective depression measured by the *DEPI* and *S-CON*. When both forms of depression are considered, 64% of the patients receiving these diagnoses exceeded the critical cutoff score for at least one of these Rorschach indices. Again, the Rorschach captured some of the underlying distress that theoretically should be present in these patients, even though this was not quantified by the MMPI.

Overall, the diagnostic data indicate genuine clinical phenomena are being captured by the negative convergent validity coefficients given in Table 6. The findings are particularly impressive because the *DSM* diagnoses are coarse and global indicators, rather than focused criteria evaluating dynamics and defenses.

## CONCLUSIONS

Several general conclusions emerge from this research. First, researchers and clinicians should adjust their conceptualization of the constructs being measured by each method so it remains clear that they are distinct latent entities. One way to facilitate this process is to develop more differentiated scale terminology. Although somewhat cumbersome, even a procedure as simple as attaching the name of the method to the name of the scale would have beneficial effects, as it is much less likely that the construct measured by the Performance-Derived Rorschach *DEPI* would become confused with the construct measured by the Self-Report-Derived MMPI-2 *DEP* Scale. Second, psychologists should recognize the limitations of each assessment method and appreciate each method's distinct range of effectiveness. Both methods can provide useful information about clinical constructs, although the data indicate neither method can consistently illuminate the full scope of any construct (see Meyer, 1996). Under optimal circumstances, the MMPI reveals what a patient understands about him or herself and is willing to convey, whereas the Rorschach reveals underlying characteristics and propensities as they are manifest through the articulation of subtle perceptual qualities. Third, because MMPI and Rorschach data are globally independent, a personality assessment would be incomplete if it relied on only one method of assessment. Thus, clinicians should competently use at least both of these measures when conducting a broad personality assessment.

Clinicians must obviously consider a range of extra-test information when sorting through Rorschach and MMPI data, perhaps particularly when confronted with cross-method discrepancies. Even in this study, which grossly simplified the issues by selecting patients with well-defined response styles, many different reasons were postulated for observing the four basic patterns of test scores. In actual practice, the complexity of the issues increases significantly. Previously, Archer and Krishnamurthy (1993a) encouraged clinicians to suppress the least reliable or valid piece of data when confronted with cross-method discrepancies. This advice is similar to the instruction many clinicians receive in training to place the greatest

interpretive confidence in those findings that emerge on several different measures. Although both guidelines can certainly be appropriate at times, they offer an incomplete solution to the problem of cross-method disagreement. In fact, these suggestions can be problematic even when there is cross-method agreement, as the agreement may be due to the confluence of communication-driven response styles that inaccurately drive observed scores on both methods higher (or lower) than is true. A better solution then may be to reconceptualize the role of the clinician in personality assessment.

One vision of the assessor's role is to tally scores, ratios, and indices from one or more tests. These scores can then be taken to a cookbook for amplification or entered into a computer for an expert's descriptive report. The distillation from this process are statements about personality along with some diagnostic impressions and treatment implications. Contradictory pieces of information can be handled by omitting less frequently mentioned descriptions, or by selecting information that fits with the assessor's impression of the patient from observation or interview. Within this model, the assessor can begin to think he or she has sampled a relatively large domain of personality because most cookbooks and computer summaries are not shy about providing comments about many facets of functioning. However, the assessor in this scenario is essentially a technician who happens to work with testing information. He or she needs very little clinical acumen, minimal understanding of the strengths and limitations of each testing measure, minimal diagnostic skills, and essentially no advanced training.

The data presented here argue against this vision and instead suggest the assessor needs to be highly skilled and capable of bringing clinical acumen; integrative conceptual abilities; and extensive knowledge of tests, assessment methodology, and psychopathology to bear on each patient being evaluated. He or she must recognize what type of information each method can reveal under optimal circumstances, while also having a realistic appreciation of what information cannot be quantified. When using different methods, the assessment clinician must work to create a sophisticated portrait of the patient that makes use of all data—even when the methods are superficially discrepant, as it is just such disagreements that often provide a richer understanding of defensive structures and struggles to adapt.

Because clinical validity does not reside solely within the scales of an instrument, and because all measures are inherently limited, the clinician should not try to decide which scale or method is the most accurate in some ultimate sense. Rather, the goal should be to figure out what conditions in nature are most likely to give rise to the observed pattern of scores across all methods. This means assessment clinicians must develop a range of inferences that could explain the pattern of scores on all tests. Once this is done, these inferences need to be compared to the clinician's knowledge of psychopathology and understanding of the many complex ways people are actually put together. Through a process of successive approximation, and incorporating as much additional information as possible while still staying

true to all of the test data, the clinician must then prune and shape test-derived inferences, terminating the process when an understanding of the data comes into sharper focus but before inferences extend beyond what the data will support. At times, the reasoning associated with this process should even generate impressions of the patient that are at odds with those drawn from an interview or history. If tests can truly provide nontrivial information then, at times, they should provide knowledge that is not obvious from other sources. If the alternative is to only report findings that are consistent with impressions derived from an interview or behavioral observations, then there is really no point to embarking on testing in the first place.

To achieve the skills discussed earlier, additional training beyond that offered in most graduate schools and internships will probably be necessary. However, relevant data indicate clinicians have trouble making the kinds of complex judgments that would be required by an assessment clinician even after advanced training (e.g., Dawes, 1994; Meehl, 1954). At least in part, Dawes attributed this difficulty to the fact that clinicians generally do not receive immediate feedback about their judgments and decisions. Consequently, it may be most important for assessment clinicians to consistently discuss test-based impressions with patients and openly solicit feedback about what seems accurate and inaccurate in order to critically evaluate the reasoning that led to particular conclusions (see Finn, 1996).

Finally, these data speak to the need for nomothetic research that more accurately incorporates the type of complex idiographic reasoning that would accompany skilled use of the MMPI and Rorschach in clinical practice (Meyer, 1996). In particular, the pervasive impact of response styles must be built into the planning and implementation of research. Although a fair amount of attention has been given to malingering on the MMPI, more MMPI research is needed on the intrinsic characteristics that lead to response styles and more research is needed to explore how response styles moderate external validity coefficients. Parallel research designed to achieve a better understanding of these factors on the Rorschach would also be quite beneficial.

#### ACKNOWLEDGMENTS

I am grateful to David S. Nichols, Radhika Krishnamurthy, Rhonda R. Dallas, and two anonymous reviewers for their constructive comments on an earlier draft of this article. I am also indebted to James M. Wood for his helpful comments and for his consistent challenges to the data presented here. I am also thankful to him for suggesting a formula to correct potential method artifacts and for designing and implementing the Monte Carlo trials reported in this article.

A version of this article using smaller samples was presented at the Annual Meeting of the Society for Personality Assessment, Chicago, April, 1994; and at the 29th Annual Symposium on Recent Developments in the Use of the MMPI-2, Minneapolis, MN, May 1994.



## REFERENCES

- Akhtar, S. (1987). Schizoid personality disorder: A synthesis of developmental, dynamic, and descriptive features. *American Journal of Psychotherapy, 41*, 499–518.
- Akhtar, S. (1990). Paranoid personality disorder: A synthesis of developmental, dynamic, and descriptive features. *American Journal of Psychotherapy, 44*, 5–25.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Archer, R. P., & Krishnamurthy, R. (1993a). Combining the Rorschach and the MMPI in the assessment of adolescents. *Journal of Personality Assessment, 60*, 132–140.
- Archer, R. P., & Krishnamurthy, R. (1993b). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment, 61*, 277–293.
- 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.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117*, 187–215.
- Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63*, 363–386.
- 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.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
- Cronbach, L. J. (1995). Giving method variance its due. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 145–157). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Edwards, L. K., & Edwards, A. L. (1991). A principal-components analysis of the Minnesota Multiphasic Personality Inventory factor scales. *Journal of Personality and Social Psychology, 60*, 766–772.
- Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.
- Finn, S. E. (1996). Assessment feedback integrating MMPI-2 and Rorschach findings. *Journal of Personality Assessment, 67*, 543–557.
- Gabbard, G. O. (1990). *Psychodynamic psychiatry in clinical practice*. Washington, DC: American Psychiatric Press.
- Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual*. Boston: Allyn & Bacon.
- Guntrip, H. (1958). *Schizoid phenomena, object-relations, and the self*. New York: International University Press.
- Harkness, A., McNulty, J., & Ben-Porath, Y. (1995). The Personality Psychopathology Five (PSY-5): Constructs and MMPI-2 scales. *Psychological Assessment, 7*, 101–114.
- Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology, 47*, 105–114.
- Kohut, H. (1971). *The analysis of the self: A systematic approach to the psychoanalytic treatment of narcissistic personality disorders*. New York: International Universities Press.

- Kohut, H. (1984). *How does analysis cure?* Chicago: University of Chicago Press.
- Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist*, 39, 451-454.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meyer, G. J. (1992a). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment*, 58, 231-244.
- Meyer, G. J. (1992b). 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.
- Miller, S. B. (1987). A comparison of methods of inquiry: Testing and interview contributions to the diagnostic process. *Bulletin of the Menninger Clinic*, 51, 505-518.
- Millon, T. (1987). *Manual for the Millon Clinical Multiaxial Inventory-II (MCMI-II; 2nd ed.)*. Minneapolis, MN: National Computer Systems.
- Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367-373.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 143-165). Berlin: Springer-Verlag.
- Piotrowski, C., & Keller, J. W. (1992). Psychological testing in applied settings: A literature review from 1982-1992. *The Journal of Training and Practice in Professional Psychology*, 6, 74-82.
- Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. *Journal of Personality Assessment*, 61, 394-405.
- Rappaport, D., Gill, M. M., & Schafer, R. (1968). *Diagnostic psychological testing* (Rev. ed.). New York: International Universities Press.
- Schafer, R. (1954). *Psychoanalytic interpretation in Rorschach testing: Theory and application*. New York: Grune & Stratton.
- Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist*, 48, 1117-1131.
- Sternberg, R. J. (1992). *Psychological Bulletin's* top 10 "hit parade." *Psychological Bulletin*, 112, 387-388.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465-490.

Gregory J. Meyer  
 Department of Psychology  
 University of Alaska Anchorage  
 3211 Providence Drive  
 Anchorage, AK 99508-8224

Received May 20, 1996