

# Assessment

<http://asm.sagepub.com>

---

## **Methodological Issues in Evaluating Rorschach Validity: A Comment on Burns and Viglion (1996), Weiner (1996), and Ganellen (1996)**

James M. Wood, M. Teresa Nezworski, William J. Stejskal, Sena Garven and Stephen G. West

*Assessment* 1999; 6; 115

DOI: 10.1177/107319119900600202

The online version of this article can be found at:  
<http://asm.sagepub.com/cgi/content/abstract/6/2/115>

---

Published by:

 SAGE Publications

<http://www.sagepublications.com>

**Additional services and information for *Assessment* can be found at:**

**Email Alerts:** <http://asm.sagepub.com/cgi/alerts>

**Subscriptions:** <http://asm.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** (this article cites 38 articles hosted on the SAGE Journals Online and HighWire Press platforms):  
<http://asm.sagepub.com/cgi/content/refs/6/2/115>

# METHODOLOGICAL ISSUES IN EVALUATING RORSCHACH VALIDITY: A COMMENT ON BURNS AND VIGLIONE (1996), WEINER (1996), AND GANELLEN (1996)

James M. Wood  
University of Texas at El Paso

M. Teresa Nezworski  
University of Texas at Dallas

William J. Stejskal  
Woodbridge Psychological Associates  
Falls Church, Virginia

Sena Garven  
University of Texas at El Paso

Stephen G. West  
Arizona State University

The old controversy regarding the Rorschach Inkblot Test has recently revived. The present article suggests that the debate will be most productive if careful attention is paid to methodological issues. Three recent examples illustrate how incorrect conclusions regarding Rorschach validity may occur if methodological issues are not evaluated carefully. The present article examines (a) Burns and Viglione's (1996) conclusion that the Rorschach Human Experience Variable (*HEV*) is a predictor of interpersonal relatedness among adults; (b) Weiner's (1996) conclusion that the *D* score and Morbid Responses (*MOR*) are valid measures of "experienced distress" in Posttraumatic Stress Disorder (PTSD); and (c) Ganellen's (1996a, 1996b) conclusion that the Rorschach Depression Index (*DEPI*) and the Minnesota Multiphasic Personality Inventory (MMPI) are comparable in their power to identify diagnoses of depression.

*Keywords:* Rorschach Inkblot Test, Comprehensive System for the Rorschach, methodology, validity, posttraumatic stress disorder, depression

---

James M. Wood and Sena Garven, Department of Psychology; M. Teresa Nezworski, School of Human Development; Stephen G. West, Department of Psychology. We wish to thank Richard Revelle and Lee Sechrest for their comments regarding extreme groups studies.

Correspondence concerning this article and requests for offprints should be addressed to James M. Wood, PhD, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968. E-mail [jwood@utep.edu](mailto:jwood@utep.edu)

The validity of the Rorschach Inkblot Test (Rorschach, 1921) was a subject of heated disagreement among scholars in the 1950s and 1960s (e.g., Eysenck, 1959; Jensen, 1965; Zubin, Eron, & Schumer, 1965), but the long controversy fell dormant after the appearance of the Comprehensive System for the Rorschach (Exner, 1974; see also Exner, 1978, 1986, 1991, 1993;

Exner & Weiner, 1982, 1995). Due to the success of the Comprehensive System (CS), many psychologists came to believe that the Rorschach debate had ended forever.

The old controversy has recently revived, however (Dawes 1994; Gann, 1995; Hunsley & Bailey, in press; Sechrest, Stickle, & Stewart, 1998). For example, on the basis of meta-analyses, Garb, Florio, and Grove (1998, in press; but see Parker, Hunsley, & Hanson, in press) recently concluded that the Rorschach has less validity than the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943). The results of several empirical studies seem to support the same conclusion (Archer & Krishnamurthy, 1993a, 1993b, 1997; Krishnamurthy, Archer, & House, 1996; but see Meyer 1997b). In addition, Wood, Nezworski, and Stejskal (1996a, 1996b; see also Nezworski & Wood, 1995) have pointed out that several important CS scores have scant or even negative evidence of validity, and that much of the empirical foundation of the CS consists of unpublished studies that often cannot be obtained for scrutiny by independent scholars.

In the face of criticism, several Rorschach scholars have come forward to defend the test's validity (Exner, 1996; Ganellen, 1996a; Meyer, 1997a, 1997b; Meyer & Handler, 1997; Weiner, 1996, 1997). However, as the "fire of heated exchange" (Weiner, 1996, p. 206) again begins to burn, many proponents and critics of the test appear to agree on several important points.

1. Many critics and all proponents agree that at least a few Rorschach scores have well-established convergent validity. For example, several independent research groups have found that scores on the Schizophrenia Index (*SCZI*) are related to diagnoses of schizophrenia. Critics may question the clinical utility of the *SCZI*, or doubt whether it provides any incremental validity beyond what can be gained from a clinical interview and Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) scores (Wood et al., 1996a; see also Archer & Gordon, 1988; Archer & Krishnamurthy, 1993a, 1993b). However, such criticisms address the usefulness of the *SCZI*, not its convergent validity.
2. Many critics and proponents of the test seem to agree that there is little point in debating whether "the Rorschach," taken as a whole, is valid or invalid (Weiner, 1996, pp. 206-207; Wood et al., 1996a, p. 5). "The Rorschach" consists of several different systems and hundreds of scores. Although most scores do not seem to have well-demonstrated validity, at least a few do. Therefore, by force of logic, any global statement that "the Rorschach is valid" or "the Rorschach is invalid" must be partly wrong. However, it is still interesting that in a meta-analysis of studies that were published in the *Journal of Personality Assessment* and *Journal of Clinical Psychology* from 1970 to 1981, the MMPI explained 23% to 30% of the variance whereas the Rorschach explained only 8% to 13% of the variance (Garb, Florio, & Grove, 1998).
3. As a corollary to Point 2, critics and proponents have noted that the validity of individual Rorschach variables cannot be established by global meta-analyses that average together validity coefficients from diverse scores and systems (e.g., Atkinson, 1986; Garb, Florio, & Grove, 1998; Parker, Hanson, & Hunsley, 1988). Narrowly focused meta-analyses and literature reviews are also needed to evaluate specific Rorschach variables (e.g., Meyer & Handler, 1997; Nezworski & Wood, 1995) rather than the test as a whole.
4. As a closely related point, both critics and proponents have emphasized that the evaluation of validity is most likely to be informative if Rorschach scales are examined one-by-one, on an individual basis. Weiner (1996, p. 208) has stated:
 

...the validity of multidimensional measures resides in the applicability of their specific scores to specific purposes and does not consist of any blanket validity of the entire measure for all conceivable purposes.

 Similarly, we have posed a "central question" that emphasizes the need to evaluate Rorschach scores individually according to three criteria:

Which Comprehensive System scores have shown (a) a consistent relationship to a particular psychological symptom or disorder, (b) in several methodologically adequate validation studies that were (c) conducted by unrelated researchers or research groups? (Wood et al., 1996b, p. 15)

We posed this question 3 years ago and offered our own appraisal of the research literature: Some individual CS scores (e.g., the *SCZI*) have a well-established relationship to schizophrenia, and several others (e.g., *R*) appear related to intellectual disabilities. Beyond that, however, we cannot identify any CS score with a well-demonstrated relationship to a psychological symptom or diagnosis, according to the three criteria in our “central question” (consistent validity, methodological quality, independent replications). So far, Rorschach proponents have not challenged this negative appraisal by publishing a list of individual CS scores that meet these criteria with citations to the relevant scientific literature. However, if such a list is ever published, the Rorschach debate is likely to achieve greater clarity and focus.

In the present article we focus on methodological quality, the second of our proposed criteria. Our argument is straightforward: Disagreements regarding Rorschach validity will be easier to resolve if participants base their conclusions on methodologically sound research. We discuss three recent articles that have relied on problematic research or methodologically unsound reasoning to bolster claims of Rorschach validity. We offer constructive suggestions that may help avoid such problems in the future.

### **The Validity of the Human Experience Variable as a Measure of Interpersonal Relatedness**

In the article “The Rorschach Human Experience Variable, Interpersonal Relatedness, and Object Representation in Nonpatients,” Burns and Viglione (1996) reported that the Human Experience Variable (*HEV*) is a useful predictor of interpersonal relatedness as assessed by non-Rorschach measures. Asserting that their study

provides “strong quantitative support for both object relations theory and the Rorschach” (p. 97), Burns and Viglione propose that the *HEV* can be used to “enhance” the CS for the Rorschach.

Despite the positive conclusions of the article by Burns and Viglione (1996), there are several important methodological problems with the study it describes. After summarizing the aims and methods of the study, we will discuss problems with its Rorschach and non-Rorschach measures and its “extreme groups design.” Finally, we will critically examine its central statistical analyses.

### **The Aims and Methods of Burns and Viglione’s Study**

“Form quality” in the Rorschach may be roughly defined as “fit”: Do the images reported by a subject actually “fit” the appearance of the inkblots? Form quality has long been recognized as an important aspect of the Rorschach (Exner, 1993, pp. 150, 186-187) and is measured in the CS by scores such as Conventional Form ( $X+$ %) and Distorted Form ( $X-$ %).

However, Perry and Viglione (1991, p. 491) have proposed a new Rorschach variable, the *HEV*, which focuses specifically on the quality of *human* imagery. *HEV* scores reflect the extent to which an individual reports “poor” rather than “good” human images when viewing the Rorschach cards. Citing object relations theory, Burns and Viglione (1996, pp. 92-93, 97) have hypothesized that the quality of *human* imagery, as measured by the *HEV*, has a special relationship to interpersonal relatedness. They hypothesize that *HEV* scores can provide information about interpersonal relatedness, above and beyond what can be learned from traditional form quality scores (e.g.,  $X-$ %), or from other Rorschach indicators of psychopathology (e.g., the Weighted Sum of the First Six Special Scores, *WSUM6*, a measure of deviant speech patterns).

To test their hypothesis, Burns and Viglione (1996) recruited 105 married women, who were administered the Rorschach using the Comprehensive System, and the Bell Object Relations Inventory (BORI; Bell, Billington, & Becker, 1986), a self-report instrument intended to measure

object relationships. The women's husbands were given the spouse-rated BORI (SBORI), a version of the BORI modified for the present study by changing items from first- to third-person, and the Emotional Maturity Rating Form (EMRF; Bessell, 1984). The BORI, SBORI, and EMRF scores for each woman were converted to  $z$  scores, then summed to yield a composite variable that Burns and Viglione labeled "Interpersonal Relatedness." In addition, the women and their husbands were administered measures of social desirability, intelligence, and demographic characteristics. Before discussing the statistical analyses of the study, we will comment on the choice of measures.

### Validity of the "Interpersonal Relatedness" Variable

Of the three interpersonal relatedness measures used in the study by Burns and Viglione (1996), the most problematic is the Emotional Maturity Rating Form (EMRF). The EMRF is taken from a popular press book called *The Love Test* (Bessell, 1984) and its psychometric properties are largely unknown. Bessell explicitly warns that the EMRF "is not a standardized test" (p. 16). Apparently there is only one validity study on the EMRF: In a dissertation, Tilden (1989, cited in Burns & Viglione, 1996, p. 95) asked expert judges to rate the content (or face) validity of EMRF items as indicators of "emotional maturity."

According to *The Love Test*, the EMRF can be used to assess one's romantic or marital partner. The 63 items ask about a variety of topics (Bessell, 1984, pp. 70-75). For example, items ask if the partner is "inquisitive and investigative" or has a "well-developed imagination." Eighteen of the test items inquire about work competence.

Is the EMRF helpful for evaluating the maturity of prospective marriage partners? In the absence of scientific evidence and considering only the information in *The Love Test*, we think that the EMRF might be useful for such a purpose. But as a measure of "interpersonal relatedness" it appears to be inadequate. Besides lacking demonstrated validity, many of its items appear to assess qualities that bear little relation to interpersonal relatedness. For example, over 25% of the items focus explicitly on competence in work tasks.

In the study by Burns and Viglione (1996), problems of interpretation are further complicated because separate results are not reported for EMRF, BORI, and SBORI scores. Instead, all three variables are lumped together into one composite variable labeled "Interpersonal Relatedness." The use of composite variables in research is not uncommon. For example, if two self-report measures of anxiety are used in a study and correlate highly, then the researcher may reasonably combine them into one composite measure of anxiety. The composite measure is likely to be more reliable than either of the two separate anxiety measures (Cohen & Cohen, 1983, pp. 169-171). Furthermore, by combining two variables into one, the researcher can reduce the number of variables and statistical tests, and so avoid the inference problems that arise when multiple tests are performed on the same data set.

If a researcher combines two or more measures of the same construct (e.g., one anxiety measure with another), then the interpretation of the composite variable is usually straightforward. However, if a measure of one construct is mixed with a measure of another construct (e.g., work competence with object relations), the composite variable may be nearly impossible to interpret. This seems to be the problem with the composite Interpersonal Relatedness variable in the article by Burns and Viglione (1996, p. 94). For example, they report that the correlation between BORI and EMRF scores was small ( $r = -.28$ ), although the absolute correlation between the SBORI and EMRF was larger ( $r = -.64$ ). Given this pattern of findings, one cannot assume that the BORI and EMRF are measuring the same thing, or that the BORI, SBORI, and EMRF should have all been combined into one composite measure.

### The Two Versions of the HEV

We turn next to the *HEV*, the central Rorschach variable in Burns and Viglione's (1996) study. Here an important problem reveals itself: Two different and incompatible methods were used to compute the *HEV* variable, although this problem was not noted in the original article.

The first method for calculating *HEV* scores (Burns & Viglione, 1996, pp. 92, 99) might be

called the “z score method”: (a) The “good” human responses (*Good H*) in the Rorschach protocol are counted, including human imagery that is “accurate, popular, whole, benevolent, cooperative, realistic and logical” (p. 92); (b) The “poor” human responses (*Poor H*) are counted, including human imagery that is “distorted, partial, damaged, aggressive, imaginary and confused” (p. 92); (c) *Good H* and *Poor H* scores are converted to standardized z scores using the means and standard deviations of these variables as reported by Perry and Viglione (1991); and (d) The standardized score for *Good H* is subtracted from the standardized score for *Poor H*, to yield the *HEV* score.

The second method for calculating the *HEV* (Burns and Viglione, 1996, pp. 92, 99) may be called the “weighting method”: The *Good H* and *Poor H* responses are counted and inserted into a weighting formula (pp. 92, 99). A few simple multiplications and additions are then performed, to yield the *HEV* score.

The “z score method” and “weighting method” are intended to be different versions of the same formula, and are supposed to yield identical results (Burns & Viglione, 1996, p. 92). However, the two methods do *not* yield identical results, as can be found by anyone who performs the calculations. For instance, the Appendix of Burns and Viglione’s article (p. 99) gives a hypothetical example of a protocol with 3 *Poor H* and 5 *Good H* responses. Using the weighting method, an *HEV* score of 2.18 is derived for the example (though -2.18 seems to have been intended). By contrast, using the z score method, with the means and standard deviations taken from Perry and Viglione (1991, p. 495, Table 2), we obtained an *HEV* score of -1.59 for the same example. Our calculations are shown in the Appendix of this article. The z score and weighting methods do not yield *HEV* scores that are identical or even very close. Most importantly, the two methods can change the order of *HEV* scores. For example, when the “weighting method” is used, Person A may have a higher *HEV* score than Person B. But when the z score method is used, Person B may have the higher score. Thus, the statistical results of the study by Burns and Viglione (1996) could

change depending on which scoring method was actually used.

Contacting the authors, we asked for clarification regarding the apparently discrepant methods for calculating the *HEV* (James Wood, personal communication, January 6, 1997). They published a Correction in *Psychological Assessment* (Burns and Viglione, 1997), which explained that two discrepant formulas had indeed been used to calculate the *HEV* in different tables in the original article (Burns & Viglione, 1996). The Correction (Burns and Viglione, 1997) minimized this problem and argued that these two discrepant versions of the *HEV* were “highly correlated” and “should produce the same results.”

The Correction (Burns and Viglione, 1997) asserted that the two versions of the *HEV* “should” produce the same results, not that they actually “did” produce the same results when the original data were reanalyzed. This assertion appears problematic for two reasons. First, if the two versions of the *HEV* “should produce the same results,” why were analyses performed separately for both versions of the *HEV* in the first author’s dissertation (Burns, 1993/1994, p. 84)? Second, and more importantly, when questions of this type arise regarding research findings, they should be addressed directly by an actual reanalysis of the data, not by assertions about what a reanalysis “should” show.

### The Extreme Groups Design

Burns and Viglione (1996) were primarily interested in the relationship of *HEV* and interpersonal relatedness scores. Because both variables were continuously distributed, multiple regression might have seemed a logical way to analyze the data. However, another approach was used instead: (a) The 105 participants were divided into three groups (high, medium, low) according to interpersonal relatedness scores; (b) The medium group of 35 participants was dropped from the analyses; and (c) logistic regressions were performed on the remaining participants, with interpersonal relatedness group (high or low) as the dichotomous criterion variable, and *HEV* as the main predictor. Thus, although the Abstract of the

article mentions 105 participants (p. 92), the number of participants included in the analyses was actually 70 (p. 95).

Below we discuss the logistic regressions in detail. First, however, it may be helpful to comment on the research strategy of dropping the middle third of participants in the study, and changing interpersonal relatedness from a continuous variable into a dichotomous one. The "extreme groups design," as this strategy is called, has a long history in psychology (Feldt, 1961; McNemar, 1960, p. 298) but entails certain problems that are often unrecognized.

In an extreme groups study, all participants are measured on a continuous variable (in this case, interpersonal relatedness). Participants are excluded from the analyses if their scores fall into a middle range, so that the remaining participants have scores that are either very low or very high. These "extreme groups" are compared on a second variable (in this case, *HEV* scores).

Methodologists have recommended the extreme groups design under two very specific conditions (Abrahams & Alf, 1978; Alf & Abrahams, 1975; Feldt, 1961; Garg, 1983; McNemar, 1960; Pitts & West, 1998). First, the researcher should *only* be interested in testing whether X and Y are related, and not be interested in the shape or size of their relationship. Second, the measure of X is already available, or is easily and cheaply administered to large numbers of participants, whereas the measure of Y is very time consuming or costly to administer to the full group (e.g., a CAT scan, a neuropsychological test). If these two conditions are true, then researchers can select the highest and lowest scoring individuals on the measure of X, and administer the costly measure Y only to these individuals.

The chief advantage of the extreme groups approach over other sampling designs is that it often has greater statistical power to detect a true effect (but see McClelland & Judd, 1993). To illustrate, imagine that the correlation between children's height and IQ were .10. If the height of 1,000 school children were known, and the 50 tallest and 50 shortest children were selected and measured on intelligence (IQ), a *t* test based on

the comparison of these 100 children would have very high power to detect the effect. In contrast, the test of the correlation based on a random sample of 100 children from the sample would have very low power to detect the effect. Particularly in the early stages of research on a particular question, the use of such designs can be very valuable.

Although the extreme groups design can have benefits, it can also present significant disadvantages, particularly for research that bears on clinical decision making. We discuss two methodological problems that the extreme groups design created for the study of Burns and Viglione (1996).

First, if the relationship between the *HEV* and interpersonal relatedness was nonlinear, then any ability to estimate the form of the relationship was lost when the middle participants were removed. Without such information, clinical inferences about *HEV* scores may be difficult or impossible to make. For example, the study by Burns and Viglione (1996, p. 94) found that participants with low interpersonal relatedness scores had higher *HEV* scores than participants with high interpersonal relatedness scores. However, it is entirely conceivable that the highest *HEV* scores would be associated with moderate levels of interpersonal relatedness, although this possibility cannot be explored because the data were eliminated from the analysis. Without the missing information, clinical statements about interpersonal relatedness based on *HEV* scores are risky and potentially misleading.

Second, any ability to estimate the size of the relationship between the *HEV* and interpersonal relatedness without making additional stringent assumptions was lost when data from participants were eliminated from analysis. Effect sizes can theoretically be estimated if the distribution of X and Y and their relationship in the population is known. Formulas have been developed that provide appropriate corrections if X and Y are measured continuously, are normally distributed in the population, and have a linear relationship (Alf & Abrahams, 1975; Garg, 1983). If one of the variables has been dichotomized, as in the Burns and Viglione study (1996), Feldt (1961, p. 315) provides a correction formula. However, Garg has found that this formula generally yields estimated correlations

that are more than twice as high as the true correlations. Thus, attempts to estimate the population correlation from dichotomized data are likely to be unsuccessful and potentially misleading (Garg, 1983, p. 370; McNemar, 1960, p. 298).

A further complication arises in the study by Burns and Viglione (1996) because the central analyses involve logistic regression equations with multiple predictors. No correction formula has been developed for this complex multivariate case.

The drawbacks of the extreme groups design are particularly relevant to clinical practitioners. Important clinical questions cannot be addressed because in the Burns and Viglione (1996) study data from the middle one-third of the participants were not analyzed. If one assumes a linear or monotonic relationship between the *HEV* and interpersonal relatedness, what is the size of that relationship? Could a practicing psychologist conclude that the relationship is strong enough to be clinically useful? The extreme groups design cannot provide answers to these critical clinical questions.

#### **“Hierarchical Logistic Regression With Backward Elimination”**

As already described, logistic regression was used to test the central hypotheses in Burns and Viglione’s (1996) article. The statistical approach was rather unusual: Although medical and epidemiological studies often use logistic regression when the criterion variable is truly dichotomous (e.g., Did the patient develop lung cancer, yes or no?), few use logistic regression to analyze continuous data that has been artificially dichotomized.

The results of the logistic regression analyses are reported in Tables 4, 6, and 8 (Burns & Viglione, 1996). However, these tables are ambiguous on a critical point: Although “backward elimination procedures” are mentioned in the text of the article (p. 96), the numbers in the tables appear inconsistent with either a backward stepwise or hierarchical selection procedure.

The ambiguity of these tables is important because backward stepwise and hierarchical selection procedures have entirely different purposes and interpretations. The central hypothesis in the Burns and Viglione (1996) study was that *HEV* scores would

provide information about interpersonal relatedness, above and beyond what could be learned from older Rorschach scores such as *X-%* and *WSUM6*. In short, it was hypothesized that *HEV* scores would demonstrate incremental predictive power, after controlling for *X-%* and *WSUM6*.

As Cohen and Cohen (1983, pp. 120-125) explain, a hierarchical selection procedure is appropriate to test a hypothesis of this type regarding incremental validity. In a hierarchical procedure, the researcher (a) specifies beforehand that variables will be forced into the regression equation in discrete stages in a particular order, and (b) predicts that when a particular variable (or set of variables) is forced into the equation, the predictive power of the equation will incrementally and significantly improve over the previous stage. For example, the researcher might specify that *X-%* and *WSUM6* would be entered into the regression equation during the first stage of a hierarchical procedure, that the *HEV* would be entered during the second stage, and that the predictive power of the equation would incrementally and significantly improve when the *HEV* was entered.

By contrast, a backward stepwise selection procedure would not be appropriate for such hypothesis testing. Backward stepwise selection is used primarily for data reduction, when a researcher wants to screen a large number of predictors to identify the best predictors (Cohen & Cohen, 1983; Hosmer & Lemeshow, 1989, p. 106). The researcher specifies a certain group of variables to be used as potential predictors in a regression equation. A computer algorithm then eliminates these variables one by one (i.e., step by step) from the regression equation, if their removal does not seem to reduce the equation’s predictive power. The researcher hopes that the computer will thus “weed out” nonpredictive variables and eventually leave only a subset of the best predictors. However, the procedure does not always work, is not designed for hypothesis testing, and cannot address issues of incremental validity.

We were able to clarify the nature of the logistic regressions in Burns and Viglione’s (1996) article by consulting the first author’s dissertation, which indicates that the tables in the article actually represent



both types of selection procedures simultaneously, in what she terms “hierarchical logistic regression with backward elimination” (Burns, 1993/1994, pp. 83, 87, 138; Donald J. Viglione, personal communication, March 10, 1997). In most well-known books on regression (Aiken & West, 1991; Cohen & Cohen, 1983; Darlington, 1990; Hosmer & Lemeshow, 1989; Pedhazur, 1982), hierarchical and backward stepwise procedures are described as two separate approaches with completely different uses. “Hierarchical logistic regression with backward elimination” appears to be an unusual combination of two different procedures.

As an illustration of this approach, consider Table 6 in the article by Burns and Viglione (1996, p. 95). This table portrays a logistic regression with steps labeled “1,” “2,” “3,” and “Final.” As the second author clarified (Donald J. Viglione, personal communication, March 10, 1997), Steps 1, 2, and 3 in Table 6 represent a hierarchical logistic regression, with new variables forced into the equation at each step. However, the step labeled “Final” is not the final step of the hierarchical procedure in Steps 1 to 3, as many readers might assume. Rather, the “Final” step is the last step of a *separate* backward stepwise elimination procedure for all variables in Steps 1 to 3. The same approach to data presentation was used in Tables 4 and 8 of the article. Both these tables represent a similar blending of hierarchical and backward stepwise procedures, although the article does not clearly inform readers that the numbers come from this hybrid approach.

Once the format of Tables 4, 6, and 8 is understood, interpretation becomes easier. The numbered steps in the tables represent hierarchical analyses, which are appropriate for testing questions regarding incremental validity. The steps labeled “Final” represents the endpoint of separate stepwise analyses, which are not appropriate for addressing such questions, as we have explained above.

When the tables are reexamined in this new light, an important insight emerges: The numbers in the tables do not support the conclusions that the article (Burns and Viglione, 1996) draws from them. First, consider the logistic regression in Table 6 (p. 95), which tests the hypothesis that the *HEV*

will have incremental validity, after controlling for *X-%*, and *WSUM6*. Table 6 shows that when the *HEV* was forced into the logistic regression equation in Step 3 (after entering demographic variables, *X-%*, and *WSUM6*), the improvement in prediction was not significant ( $p < .19$ ). Thus the *HEV* did not add significantly to predictive power; The research hypothesis was not confirmed. Despite the clearly negative results, however, the article concludes that the *HEV* “is making a unique contribution” (p. 96) and “predicts the quality of interpersonal relationships beyond that which is signified by other Rorschach measures of psychopathology, such as *WSUM6* and *X-%....*” (p. 97).

A second, similar discrepancy between findings and conclusions appears somewhat later in the article. According to Burns and Viglione (1996), object relations theory predicts that the quality of specifically *human* imagery should have a special relationship to interpersonal relatedness. To test this idea the authors created a special variable, the Nonhuman Experience Variable, to measure non-human (e.g., animal) Rorschach imagery. It was hypothesized that the *HEV* would predict interpersonal relatedness, above and beyond the predictive power of this Nonhuman Experience Variable. The first three steps of Table 8 (p. 96) present a hierarchical logistic regression to test this hypothesis. Table 8 clearly shows that when the *HEV* was forced into the equation (after entering demographic variables and the Nonhuman Experience Variable) the improvement in prediction of interpersonal relatedness was not significant ( $p < .11$ ). Nevertheless, Burns and Viglione (p. 97) interpret the findings as confirming the initial hypothesis: “...human Rorschach representations specifically are salient in understanding the nature and quality of interpersonal relatedness.”

As may be seen from these two examples, the conclusions drawn by Burns and Viglione (1996) are inconsistent with their findings: The analyses actually failed to support their main hypotheses.

#### **Summary Regarding the Study by Burns and Viglione (1996)**

Two methodological problems have been identified in the study by Burns and Viglione (1996): (a) The central measures in the study are problematic;

(b) When the statistical analyses are properly interpreted, they do not support the article's conclusions. Although the article argues that the results provide "strong quantitative support for both object relations theory and the Rorschach" (p. 97), in fact the study is inadequate to support such a conclusion.

### **The Validity of the Rorschach as a Measure of "Experienced Distress" in Posttraumatic Stress Disorder**

In the article "Some Observations on the Validity of the Rorschach Inkblot Method," Weiner (1996) argues that "experienced distress" (p. 212) in Posttraumatic Stress Disorder (PTSD) is related to two Rorschach variables, morbid content (*MOR*) and the *D* score (a variable derived from Rorschach determinants and believed to measure current tolerance for stress; see Exner, 1993, p. 388). Weiner (p. 112) presents data from three studies of veterans (Hartman et al., 1990; Sloan, Arsenault, Hilsenroth, Harvill, & Handler, 1995; Swanson, Blount, & Bruno, 1990), compares them with normative data from Exner (1993), and concludes

...data collected from Vietnam and Persian Gulf war veterans with PTSD have yielded (a) average *D* scores substantially lower than normative expectation and well into the minus range and (b) average *MOR* scores substantially higher than normative expectation.

The three PTSD studies cited by Weiner (1996) come from independent researchers and are consistent in their findings regarding *D* and *MOR*. However, all three studies share the same methodological flaw: None had a control group. Instead, they all compared the Rorschach scores of veterans to general population norms for the CS published by Exner (1986, 1993).

There are two main problems with using the CS normative data, rather than true control groups, in the PTSD studies. First, the combat veterans in the studies apparently differed from the CS normative group in a variety of ways (age, occupation, socioeconomic status, educational level, and so on). These differences, or other uncontrolled factors, could account for the between-group differences in

Rorschach scores. The critical scientific and clinical question is whether *D* and *MOR* can discriminate combat veterans with PTSD distress from combat veterans without the disorder. However, this question cannot be addressed by studies that use normative data, rather than veterans, as a control group.

Second, the interrater reliability of most Comprehensive System scores, including *D* and *MOR*, has never been adequately demonstrated (Wood et al., 1996a, 1996b, 1997; see also McDowell & Acklin, 1996; but see Exner, 1996, and Meyer, 1997a, 1997c). Even if *D* and *MOR* can be coded reliably (e.g., with an intraclass correlation coefficient of at least .80), there would still be room for substantial and systematic scoring differences between the Rorschach raters for the PTSD studies and the raters for the CS normative group. Unless the same set of blinded raters were used for both groups, the between-groups differences in scores might simply reflect the different scoring habits or biases of different groups of raters, or be due to "observer drift" (Haynes, 1978, pp. 152-157).

In a *Psychological Assessment* article published about a year before his discussion of the PTSD studies, Weiner (1995, p. 336) himself voiced similar concerns about the use of normative data in Rorschach research:

...formal statistical comparisons of data from a delimited sample with the normative data should be avoided. Comparison groups should be similar in size and composition, and should be examined in similar ways at similar places and points in time as the target or experiment groups in a study. The Comprehensive System nonpatient reference group is a larger and more diverse group of people than the usual samples of individual investigators, and their protocols were collected in many different places at different times using many different examiners. These differences in the nature of the samples makes statistical comparisons between them inappropriate.

In this earlier article, Weiner (1995) articulates the reasons that comparisons to normative data may

yield inappropriate inferences (see also Exner, Kinder, & Curtiss, 1995; Ritzler & Exner, 1995; Viglione, 1997; Viglione & Exner, 1995). If similar principles are applied to the PTSD studies in the later article (Weiner), the methodology of those studies can be seen as inadequate: The three PTSD studies cited by Weiner are flawed, and are inadequate for establishing the validity of *D* and *MOR* as measures of “experienced distress” in PTSD.

### **The Validity of the Rorschach and MMPI as Predictors of Depression Diagnoses**

In “Comparing the Diagnostic Efficiency of the MMPI, MCMI-II, and Rorschach: A Review,” Ganellen (1996a; see also Ganellen, 1996b) argues that the MMPI, the MCMI-II, and the Rorschach “have comparable ability to correctly identify patients diagnosed with depression as being depressed” (p. 235). The present discussion focuses on the assertion that the MMPI and the Depression Index (*DEPI*) of the CS are comparable for identifying diagnoses of depression.

As Ganellen’s article (1996a) discusses, many peer-reviewed studies have documented that MMPI scales are related to diagnoses of depression (see also Archer & Krishnamurthy, 1997). However, the same is not true for the *DEPI*. The most recent editions of *The Rorschach: A Comprehensive System (TRACS)* report that the *DEPI* is both sensitive and specific as an indicator of depression diagnoses (Exner, 1993, pp. 260-264, 309-311), and that elevated scores on the *DEPI* “correlate very highly with a diagnosis that emphasizes serious affective problems” (Exner, 1991, p. 146). However, independent peer-reviewed studies of the *DEPI*, published both before and after Ganellen’s (1996a) article, have nearly all found that *DEPI* scores are unrelated to diagnoses of depression in either adolescents or adults (Archer & Gordon, 1988; Archer & Krishnamurthy, 1997; Ball, Archer, Gordon, & French, 1991; Carlson, Kula, & St. Laurent, 1997; Carter & Dacey, 1996; Meyer, 1993; Viglione, Brager, & Haller, 1988; see also Sells, 1990/1991; but see Jansak, 1996/1997; Singer & Brabender, 1993).

If peer-reviewed studies have consistently found that MMPI scores are validly related to diagnoses of depression, and that *DEPI* scores are not, how did Ganellen (1996a) arrive at the apparently paradoxical conclusion that the MMPI and Rorschach are comparable in this respect? His methodology appears to have been weak in at least three respects.

First, in estimating sensitivity and specificity figures for the *DEPI*, Ganellen (1996a) relied exclusively on tables published in *TRACS* (Exner, 1991). As we have discussed, however, the findings in *TRACS* appear to be unusual, even anomalous, compared to the findings regarding the *DEPI* in peer reviewed articles. Thus, Ganellen may have based his conclusions on an atypical data set.

Second, if a scientist wishes to compare the relationship of Tests A and B to a particular diagnosis, the best strategy is to examine studies that have administered both tests to the same group of participants. This may be called the “within-groups” approach. A more problematic strategy, which may be called the “between-groups” approach, is to examine studies that have administered Test A to some groups of participants, and studies that have administered Test B to other groups of participants, and compare the results. This “between-groups” approach is apt to be unreliable because the various groups of participants may differ in important ways (e.g., validity of diagnoses, severity of symptoms). Any differences between Tests A and B will be entangled with differences between the groups of participants. Ganellen (1996a) relied on the problematic between-groups approach, rather than the within-groups approaches, when he concluded that the MMPI and Rorschach are comparable for evaluating diagnoses of depression.

Third, in reaching his conclusions Ganellen (1996a) minimized the importance of existing negative research findings regarding the *DEPI*. Specifically, Archer and Gordon (1988) had found that the *DEPI* and the Schizophrenia Index (*SCZI*) do not provide incremental validity beyond MMPI scores in prediction of diagnoses. However, Ganellen (p. 240) discounted the study by Archer and Gordon because it had used the original MMPI and earlier versions of the *DEPI* and *SCZI*:

It is likely that results based on the original MMPI, original *DEPI*, and original *SCZI* will differ from results using the MMPI-A and updated *SCZI* and updated *DEPI*. Thus, Archer and Gordon's results have little bearing on the diagnostic efficiency of the MMPI-A, revised *DEPI*, and revised *SCZI*, the versions of the MMPI and Rorschach currently in use and appropriate for an adolescent population.

Ganellen (1996a) argued that Archer and Gordon's (1988) study had "little bearing," and speculated that the results of future research were "likely" to be different. However, untested speculations tend to be risky, especially when they conflict with prior empirical findings. In fact, Ganellen's speculations were soon disconfirmed by research. Using the revised *DEPI* and the MMPI-A, Archer and Krishnamurthy (1997) reported results similar to those of Archer and Gordon: The new *DEPI* did not significantly predict diagnoses of depression or increase incremental validity beyond the predictive power of MMPI-A scores.

In all fairness, it should be noted that Ganellen (1996a) himself acknowledged some of the same methodological limitations identified here, although he apparently did not consider them as weighty as we do. Despite the potential methodological pitfalls, his problematic conclusions about the *DEPI* and the MMPI were published in both an article and book (Ganellen 1996a, 1996b), and have been cited as justification for using the Rorschach in forensic and other contexts (McCann, 1998; Meyer et al., 1998).

Although we have not discussed Ganellen's (1996a; 1996b) conclusions regarding the *SCZI* and the MMPI, many of the same methodological considerations apply. The sensitivity and specificity figures for the *SCZI* reported by Ganellen (1996a) must be viewed cautiously. In fact, studies that have compared the first and second versions of the *SCZI* with the MMPI (Archer & Gordon, 1988; Meyer, 1993) have found that the *SCZI* does not add incremental validity to the prediction of

schizophrenia diagnoses, beyond what can be obtained using the MMPI.

## Recommendations

We have discussed several methodological issues in the articles by Burns and Viglione (1996), Weiner (1996), and Ganellen (1996a). Most of these problems have an obvious remedy and need not be discussed further. However, a few are more complex and their solution may not always be obvious. Therefore, in closing we would like to offer three specific recommendations for future Rorschach studies and literature reviews.

First, empirical studies and literature reviews of Rorschach variables should avoid using CS normative data for comparisons. Weiner's (1995) recommendations regarding appropriate control groups should be followed in empirical studies. Administration and scoring of Rorschach tests should be the same for both experimental and control groups and blind to group membership.

Second, when using multiple regression and logistic regression analyses, backward stepwise selection procedures should not be used to test hypotheses about incremental validity. Hierarchical selection procedures should be used instead.

Finally, Extreme groups designs are not generally recommended for Rorschach studies that might be used as a basis for clinical decision making. Positive findings based on this design should be interpreted very conservatively, and should not be used as a basis for clinical decision making. Conversely, because the design typically has high statistical power, failure to detect a relationship between two variables in a well-done extreme groups study usually indicates that the relationship is small and therefore unlikely to be of clinical significance. It should be noted that these comments regarding extreme groups designs do not apply to studies in which group membership is based on diagnostic categories (e.g., schizophrenics vs. non-schizophrenics, Alzheimer's patients vs. normal elderly).

## References

- Abrahams, N. M., & Alf, E. F. (1978). Relative costs and statistical power in the extreme groups approach. *Psychometrika*, *43*, 11-17.
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Alf, E. F., & Abrahams, N. M. (1975). The use of extreme groups in assessing relationships. *Psychometrika*, *40*, 563-572.
- Archer, R. P., & Gordon, R. A. (1988). MMPI and Rorschach indices of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment*, *52*, 276-287.
- 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.
- Archer, R. P., & Krishnamurthy, R. (1997). MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment*, *69*, 517-533.
- Atkinson, L. (1986). The comparative validities of the Rorschach and the MMPI: A meta-analysis. *Canadian Psychology*, *27*, 238-247.
- Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach depression indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment*, *57*, 465-476.
- Bell, M. D., Billington, R., & Becker, B. (1986). A scale for the assessment of object relationships: Reliability, validity, and factorial invariance. *Journal of Clinical Psychology*, *42*, 733-741.
- Bessell, H. (1984). *The love test*. New York: William Morrow.
- Burns, B. (1994). An object relations study: Relationship between the Rorschach Human Experience Variable and interpersonal relatedness among nonpatients (Doctoral dissertation, California School of Professional Psychology, San Diego, 1993). *Dissertation Abstracts*, *54*, 3847B.
- Burns, B., & Viglione, D. J. (1996). The Rorschach Human Experience variable, interpersonal relatedness, and object representation in nonpatients. *Psychological Assessment*, *8*, 92-99.
- Burns, B., & Viglione, D. J. (1997). Correction to Burns and Viglione (1996). *Psychological Assessment*, *9*, 82.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Carlson, C. F., Kula, M. L., & St. Laurent, C. M. (1997). Rorschach revised DEPI and CDI with inpatient major depressives and borderline personality disorder with major depression: Validity issues. *Journal of Clinical Psychology*, *53*, 51-58.
- Carter, C. L., & Dacey, C. M. (1996). Validity of the Beck Depression Inventory, MMPI, and Rorschach in assessing adolescent depression. *Journal of Adolescence*, *19*, 223-231.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.
- Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Exner, J. E. (1974). *The Rorschach: A Comprehensive System. Volume 1*. New York: Wiley.
- Exner, J. E. (1978). *The Rorschach: A Comprehensive System. Volume 2. Current research and advanced interpretation*. New York: Wiley.
- Exner, J. E. (1986). *The Rorschach: A Comprehensive System. Volume 1: Basic foundations* (2nd ed.). New York: Wiley.
- Exner, J. E. (1991). *The Rorschach: A Comprehensive System. Volume 2: Interpretation* (2nd ed.). New York: Wiley.
- Exner, J. E. (1993). *The Rorschach: A Comprehensive System. Volume 1: Basic foundations* (3rd ed.). New York: Wiley.
- Exner, J. E. (1996). A comment on "The Comprehensive System for the Rorschach: A critical examination." *Psychological Science*, *7*, 11-13.
- Exner, J. E., Kinder, B. N., & Curtiss, G. (1995). Reviewing basic design features. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 145-158). Mahwah, NJ: Erlbaum.
- Exner, J. E., & Weiner, I. B. (1982). *The Rorschach: A Comprehensive System. Volume 3: Assessment of Children and Adolescents*. New York: Wiley.
- Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A Comprehensive System. Volume 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.
- Eysenck, H. J. (1959). The Rorschach Inkblot Test. In O. K. Buros (Ed.), *The fifth mental measurements yearbook* (pp. 276-278). Highland Park, NJ: Gryphon Press.
- Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika*, *26*, 307-316.
- Ganellen, R. J. (1996a). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: A review. *Journal of Personality Assessment*, *67*, 219-243.
- Ganellen, R. J. (1996b). *Integrating the Rorschach and the MMPI-2 in personality assessment*. Mahwah, NJ: Erlbaum.
- Gann, M. K. (1995). The Rorschach and other projective methods. In J. Ziskin (Ed.), *Coping with psychiatric and psychological testimony. Vol. II: Special topics* (5th ed.; pp. 823-884). Los Angeles: Law and Psychology Press.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses. *Psychological Science*, *9*, 402-404.

- Garb, H. N., Florio, C. M., & Grove, W. M. (in press). The Rorschach controversy: Reply to Parker, Hunsley, and Hanson. *Psychological Science*.
- Garg, R. (1983). An empirical comparison of three strategies used in extreme group designs. *Educational and Psychological Measurement*, 43, 359-371.
- Hartman, W. L., Clark, M. E., Morgan, M. K., Dunn, V. K., Fine, A. D., Perry, G. G., & Winsch, D. L. (1990). Rorschach structure of a hospitalized sample of Vietnam veterans with PTSD. *Journal of Personality Assessment*, 54, 149-159.
- Hathaway, S. R., & McKinley, J. C. (1943). *Manual for the Minnesota Multiphasic Personality Inventory*. New York: The Psychological Corporation.
- Haynes, S. N. (1978). *Principles of behavioral assessment*. New York: Gardner Press.
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley & Sons.
- Hunsley, J., & Bailey, J. M. (in press). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*.
- Jansak, D. M. (1997). The Rorschach Comprehensive System Depression Index, depression heterogeneity, and the role of self-schema (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). *Dissertation Abstracts International*, 57, 6576B.
- Jensen, A. R. (1965). The Rorschach Inkblot Test. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (pp. 501-509). Highland Park, NJ: Gryphon Press.
- Krishnamurthy, R., Archer, R. P., & House, J. J. (1996). The MMPI-A and Rorschach: A failure to establish convergent validity. *Assessment*, 3, 179-191.
- McCann, J. T. (1998). Defending the Rorschach in court. An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment*, 70, 125-144.
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376-390.
- McDowell, C. J., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interobserver reliability: Conceptual and empirical foundations. *Journal of Personality Assessment*, 66, 308-320.
- McNemar, Q. (1960). At random: Sense and nonsense. *American Psychologist*, 15, 295-300.
- Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment*, 60, 153-180.
- Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment*, 9, 480-489.
- Meyer, G. J. (1997b). On the integration of personality assessment methods: The Rorschach and MMPI. *Journal of Personality Assessment*, 68, 297-330.
- Meyer, G. J. (1997c). Thinking clearly about reliability: More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment*, 9, 495-498.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69, 1-38.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W., Moreland, K. L., Eisman, E. J., & Dies, R. R. (1998). *Benefits and costs of psychological assessment in healthcare delivery: Report of the board of professional affairs psychological assessment work group, Part 1*. Washington, DC: American Psychological Association.
- Nezworski, M. T., & Wood, J. M. (1995). Narcissism in the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice*, 2, 179-199.
- 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.
- Parker, K. H. C., Hunsley, J., & Hanson, R. K. (in press). Old wine from old skins sometimes tastes like vinegar: A response to Garb, Florio, and Grove. *Psychological Science*.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- Perry, W., & Viglione, D. J. (1991). The Ego Impairment Index as a predictor of outcome in melancholic depressed patients treated with tricyclic antidepressants. *Journal of Personality Assessment*, 56, 487-501.
- Pitts, S. C., & West, S. G. (1998). *Alternative sampling designs to detect interactions in multiple regression*. Unpublished manuscript, Arizona State University.
- Ritzler, B. A., & Exner, J. E. (1995). Special issues in subject selection and design. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 123-143). Mahwah, NJ: Erlbaum.
- Rorschach, H. (1921). *Psychodiagnostic*. Bern, Switzerland: Bircher (Transl. Han Huber Verlag, 1942).
- Sechrest, L., Stickle, T. R., & Stewart, M. (1998). The role of assessment in clinical psychology. In A. Bellack, M. Hersen, (Series Eds.) & C. R. Reynolds, (Vol. Ed.), *Comprehensive clinical psychology. Vol. 4: Assessment*, (pp. 1-32).
- Sells, J. E. (1991). A validity study of the DEPI index: The Rorschach Comprehensive System (Doctoral dissertation, University of Utah, 1990). *Dissertation Abstracts International*, 51, 5590B.
- Singer, H. K., & Brabender, V. (1993). The use of the Rorschach to differentiate unipolar and bipolar disorders. *Journal of Personality Assessment*, 60, 333-345.
- Sloan, P., Arsenault, L., Hilsenroth, M., Harvill, L., & Handler, L. (1995). Rorschach measures of posttraumatic stress in Persian Gulf War veterans. *Journal of Personality Assessment*, 64, 397-414.
- Swanson, G. S., Blount, J., & Bruno, R. (1990). Comprehensive System Rorschach data on Vietnam combat veterans. *Journal of Personality Assessment*, 54, 160-169.
- Tilden, R. S. (1989). *Predicting marital adjustment with level of object relations, romantic love, and emotional maturity*. Unpublished doctoral dissertation, United States International University, San Diego, CA.

Viglione, D. J. (1997). Problems in Rorschach research and what to do about them. *Journal of Personality Assessment*, 68, 590-599.

Viglione, D. J., Brager, R. C., & Haller, N. (1988). Usefulness of structural Rorschach data in identifying inpatients with depressive symptoms: A preliminary study. *Journal of Personality Assessment*, 52, 524-529.

Viglione, D. J., & Exner, J. E. (1995). Formulating issues in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 53-71). Mahwah, NJ: Erlbaum.

Weiner, I. B. (1995). Methodological considerations in Rorschach research. *Psychological Assessment*, 7, 330-337.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment*, 8, 206-213.

Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment*, 68, 5-19.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996a). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, 7, 3-10.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996b). Thinking critically about the Comprehensive System for the Rorschach. A reply to Exner. *Psychological Science*, 7, 14-17.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System for the Rorschach: A Comment on Meyer (1997). *Psychological Assessment*, 9, 490-494.

Zubin, J., Eron, L. D., & Schumer, F. (1965). *An experimental approach to projective techniques*. New York: Wiley.

## Appendix

### Comparison of the “z Score” and “Weighting Formula” Methods of Calculating *HEV* Scores

(1) In a hypothetical example, Burns and Viglione (1996, p. 99) state that a Rorschach protocol with 3 *Poor H* and 5 *Good H* responses has an *HEV* score of 2.18.

(2) Perry and Viglione (1991, p. 495, Table 2) give the mean of *Poor H* as 3.80 and the standard deviation as 2.48. Therefore if there are 3 *Poor H* responses in a protocol, then

$$z \text{ score of } \textit{Poor H} = \frac{3 - 3.80}{2.48} = -0.32$$

(3) Perry and Viglione give the mean of *Good H* as 2.63 and the standard deviation as 1.86. Therefore if there are 5 *Good H* responses, then

$$z \text{ score of } \textit{Good H} = \frac{5 - 2.63}{1.86} = 1.27$$

(4) Using the “z score method,” then

$$\begin{aligned} \textit{HEV score} &= z \text{ score of } \textit{Poor H} - z \text{ score of } \textit{Good H} \\ &= -0.32 - 1.27 = -1.59 \end{aligned}$$

(5) Thus the *HEV* score of -1.59 calculated using the “z score method” is different than the *HEV* score of 2.18 that Burns and Viglione (1996, p. 99) arrived at for the same hypothetical protocol using the “weighting formula method.”