June 2001

# Advancing Scientific Discourse in the Controversy Surrounding the Comprehensive System for the Rorschach: A Rejoinder to Meyer (2000)

James M. Wood
*University of Texas at El Paso*

M. Teresa Nezworski
*University of Texas at Dallas*

William J. Stejskal
*University of Virginia*

Sena Garven
*University of Nebraska–Lincoln*

JOURNAL OF PERSONALITY ASSESSMENT, *76*(3), 369–378


ARTICLES

# Advancing Scientific Discourse in the Controversy Surrounding the Comprehensive System for the Rorschach: A Rejoinder to Meyer (2000)

James M. Wood
*Department of Psychology*
*University of Texas at El Paso*

M. Teresa Nezworski
*School of Human Development*
*University of Texas at Dallas*

William J. Stejskal
*Institute of Law, Psychiatry, and Public Policy*
*University of Virginia*

Sena Garven
*Department of Psychology*
*University of Nebraska–Lincoln*

A recent commentary by Meyer (2000) in the *Journal of Personality Assessment* alleged that Rorschach critic Wood and his colleagues had intentionally published information that they knew to be in error. To substantiate this contention, Meyer's commentary published information that was part of the peer review process at another journal. In this rejoinder, we present factual information that shows we have consistently acted in good faith. This rejoinder suggests that the scientific debate regarding the Comprehensive System for the Rorschach is unlikely to be advanced by speculat-

ing about the intentions of Rorschach critics, or by publishing information from the peer review process that is usually kept confidential.

Recently, Wood, Nezworski, Stejskal, Garven, and West (1999) published a 15-page article in *Assessment* that discussed certain methodological points relevant to research on the Comprehensive System for the Rorschach (CS) and critically examined claims made by Burns and Viglione (1996), Weiner (1996), and Ganellen (1996a, 1996b).

Recently, Meyer (2000) published a 36-page commentary in the *Journal of Personality Assessment* on the article by Wood et al. (1999). The commentary disagreed with many of the methodological points made by Wood et al. and presented new data regarding the incremental validity of some Rorschach scores.

In this rejoinder, we wish to respond briefly to Meyer's (2000) commentary. Meyer's commentary presented some new material that is informative and highly interesting (e.g., the data on incremental validity of the Schizophrenia Index and the Depression Index), but other discussions strike us as questionable (e.g., regarding stepwise procedures for testing incremental validity or the construction of composite measures). If Meyer's commentary had been limited to such methodological points and the presentation of data, however, we probably would not have felt that a rejoinder was necessary. Interested readers could have read our article, Meyer's commentary, and the relevant statistical and methodological references, and then come to their own conclusions.

However, Meyer's (2000) commentary went beyond methodology and data. It alleged that in our article we had published information that we knew to be in error at the time we submitted it for publication. To substantiate this allegation, Meyer's commentary cited reviews that had been written when our article had been earlier submitted to another journal.

In light of Meyer's allegations, which suggest that we acted dishonestly or in bad faith, we believe that we have no choice but to write this rejoinder. In the following pages, we present factual information that shows we have consistently acted in good faith. We then discuss whether the current scientific debate regarding the Rorschach is likely to be advanced by questioning the intentions of critics of the CS for the Rorschach or by publishing materials from the peer review process that are usually kept confidential.

## THE "FAULTY" FORMULA FOR THE *HEV*

### Meyer's Commentary

In our *Assessment* article (Wood et al., 1999), we discussed a study of the Rorschach Human Experience Variable (*HEV*) by Burns and Viglione (1996). Among other things, we argued (pp. 118–119, 129) that a formula for the *HEV* published by Burns and Viglione failed to yield the results that it was supposed to.

Meyer's (2000) commentary contended that the formula for the *HEV* in our article was "faulty" (p. 72). Calculation of the *HEV* score requires that means and standard deviations from a distribution of scores be inserted into the *HEV* formula. Meyer's commentary argued that we had erroneously inserted the means and standard deviations from a study by Perry and Viglione (1991), whereas we should have inserted the means and standard deviations from a study by Haller (1982).

However, Meyer's (2000) commentary went well beyond the simple assertion that we were in error: It alleged that we had known our *HEV* formula was wrong at the time we submitted our article to *Assessment.* To support this allegation, which implied intellectual dishonesty on our part, Meyer's commentary revealed that before being sent to *Assessment,* our article had been submitted to another journal, where it was reviewed by Meyer and four other reviewers. Meyer (2000) published the following quote from a review written by Reviewer 4:

> The authors refer to Perry and Viglione (1991). These authors should know, per Perry and Viglione (p. 491) that the EII and the *HEV* weights are derived from data from Haller (1982). One of the strengths of Perry and Viglione (1991) was that the *HEV* and EII weights were derived from one sample (Haller, 1982) and cross-validated with another. Thus without recognizing this clearly stated fact (in Perry & Viglione, 1991), no wonder the authors are confused and mistaken in their calculations. (p. 71)

Meyer (2000) then commented:

> For some reason, Wood et al. did not attend to this corrective input from the peer review process and continued to promulgate a *z*-score formula that they had been told was derived from improper means and standard deviations. … Wood et al. … championed a formula they knew was incorrect. (pp. 71–72)

## Our Response

It is untrue that we knowingly published a faulty formula for the *HEV.* In fact, the formula we published was correct and not faulty at all. As we stated in our article (Wood et al., 1999) and continue to insist, the formula for the *HEV* in the article by Burns and Viglione (1996) should be calculated by inserting the means and standard deviations reported by Perry and Viglione (1991), and not the means and standard deviations of the scores reported by Haller (1982). We are quite confident in our assertion that the *HEV* formula used by Burns and Viglione should be based on the means and standard deviations of Perry and Viglione because Burns and Viglione explicitly said so:

> These *z*-score transformations [used to calculate the *HEV*] are derived from means and standard deviations from the original Perry and Viglione (1991) distributions. (p. 92)

Burns and Viglione unambiguously stated that the means and standard deviations of the Perry and Viglione distributions were used. The study by Haller was not even cited by Burns and Viglione.

It is true, as Meyer (2000) asserted, that Reviewer 4 claimed that Burns and Viglione (1996) used the means and standard deviations of Haller (1982). However, this "corrective input from the peer review process" (Meyer, 2000, p. 71) flatly contradicted what Burns and Viglione had said in their article. Faced with this contradiction, we had to choose whether to believe the clear statement by Burns and Viglione on this issue, or instead to accept the undocumented assertion of an anonymous reviewer. We chose to believe the statement in Burns and Viglione's article.

Much confusion regarding the *HEV* formula has arisen because there are actually several different ways of computing the *HEV* that yield different results. Specifically, in the dissertation that formed the basis for the article by Burns and Viglione (1996), Burns (1993/1994) reported results for three different versions of the *HEV*, each of which was calculated using a different formula based on a different set of means and standard deviations. Apparently she was experimenting with these different formulas to determine which would yield the best results. When Burns's dissertation was transformed into the article by Burns and Viglione, the results for the different versions of the *HEV* were mixed together, apparently by accident. As a result, one table in the article reported results for one version of the *HEV*, and another table reported results for a different version, but without stating that the results were for different versions of the variable. When we compared the dissertation with the article, we noted the discrepancy and contacted Viglione (J. M. Wood, personal communication, January 6, 1997). Burns and Viglione (1997) subsequently published an erratum in *Psychological Assessment* and acknowledged that they had mistakenly used two versions of the *HEV* formula in their article.

## THE INCOMPATIBILITY OF THE DIFFERENT VERSIONS OF THE *HEV*

### Meyer's Commentary

In our *Assessment* article (Wood et al., 1999), we pointed out that Burns and Viglione's (1996) original article had used different versions of the *HEV* in different tables (see also the erratum published by Burns & Viglione, 1997). We argued that this mix-up had undesirable consequences because the different formulas used to calculate the *HEV* are "incompatible [and] do not yield *HEV* scores that are identical or even very close" (p. 119). We illustrated this point (pp. 119, 129) by showing how one formula yielded an *HEV* score of –2.18, whereas another formula yielded an *HEV* score of –1.59 (a difference of approximately .3 *SD*). We further asserted that the different formulas “can change the order of the *HEV* scores” among different subjects, and thus “the statistical results of the study by Burns and Viglione (1996) could change depending on which scoring method was actually used” (p. 119).

In response, Meyer’s (2000) commentary presented tables (pp. 73–74) showing that three different formulas for the *HEV* are highly correlated ($r > .99$). The commentary asked rhetorically:

> Are these claims [of Wood et al., 1999] true? Does the faulty Wood et al. *z*-score formula produce results that are so dramatically at odds with the correct formula? The answer to both questions is *no*. (p. 72)

However, Meyer’s (2000) commentary again went beyond the simple assertion that we were mistaken. It asserted that we had knowingly published false statements:

> Wood et al. knew their statements were not true before they submitted their final article for publication …. Given the remarkable association between these formulas, it is troubling to consider that Wood et al. … were aware of these findings before they submitted their article for final publication. That is, before going to press … the authors had been told that, at worst, they were describing correlations greater than .9985. (pp. 72–73)

To support the allegation that we intentionally published statements that we knew to be untrue, Meyer’s (2000) commentary presented two pieces of evidence. First, Meyer (p. 74) revealed that when he had reviewed an earlier version of our article in Spring 1998, for a journal other than *Assessment,* he included in his anonymous review the results of seven simulation studies (i.e., Monte Carlo studies) that showed that different formulas for the *HEV* yielded very highly correlated scores in artificially generated data. Second, Meyer reported (p. 74) that in a debate on the Internet in October 1998, he supplied Wood with analyses based on real data from 232 patients. According to these analyses, the different formulas for the *HEV* yielded very highly correlated scores.

## Our Response

It is simply untrue that we intentionally included anything in our *Assessment* article that we knew to be false. When the article was published, we accurately and forthrightly stated the views that we held at that time. Although we had read Meyer’s arguments and examined his analyses, we did not find them compelling.

As background, readers need to understand that in 1996, within a few months after Burns and Viglione (1996) had published their article regarding the *HEV* in *Psychological Assessment,* we contacted the corresponding author, Donald Viglione, and requested their data for reanalysis (J. M. Wood, personal communication, April 5, 1996). In response, Viglione informed us that the data had been lost and were therefore unavailable for reanalysis (D. J. Viglione, personal communications, October 18, 1996, March 10, 1997).

Because Burns and Viglione's (1996) data were lost, any claim (e.g., Burns & Viglione, 1997) that the various *HEV* formulas should yield "the same" or "highly correlated" (p. 82) results in Burns and Viglione's data set struck us as highly problematic when we wrote our article. Neither Burns's (1993/1994) dissertation nor the article by Burns and Viglione reported the correlations among the various *HEV* formulas: Apparently these correlations were never calculated while the data set was still in existence. Now that the data set is lost, the correlations will probably never be known. Thus, instead of real correlation coefficients calculated with Burns and Viglione's real data, there are only conjectures and speculations about what the data might have shown.

In early 1998, we submitted an early version of our article to a journal where it was ultimately rejected. In this version, we reported that Burns and Viglione's (1996) data were lost, and we raised concerns about their use of different formulas for the *HEV*. In response, an anonymous reviewer (whom Meyer has now identified as himself) stated that he had carried out several Monte Carlo simulations using artificial data. These simulations indicated that the various versions of the *HEV* formula yielded different *HEV* scores (as we had said) but these scores were highly correlated with each other, $r > .99$. These analyses based on artificial data were reproduced in the recent commentary by Meyer (2000).

We found these analyses to be unconvincing for two reasons. First, although the artificial variables in the simulations were apparently uncorrelated and normally distributed, in reality the Rorschach variables used to calculate the *HEV* (*Good H* and *Poor H*) are apparently intercorrelated, with moderate to high skewness (Burns, 1993/1994). Thus, the Monte Carlo artificial variables did not seem to model the real-world variables very well. Second, the Monte Carlo simulations struck us as a very poor substitute for the original data from Burns and Viglione (1996). In our view, any claim that the different *HEV* formulas would yield the same results in the Burns and Viglione data set should be substantiated by an actual reanalysis of that data set, not by Monte Carlo studies of artificial data. We expressed our views on these issues in a letter to the action editor at the journal:

> Reviewer 5 [Meyer] seems to think that fictional data, generated with erroneous assumptions, are an acceptable substitute for the real data. We disagree. (J. M. Wood, personal communication, May 22, 1998)

In October 1998, in an exchange on the Internet, Meyer presented analyses of his own Rorschach data from 232 patients. These analyses (which were also reproduced in the commentary by Meyer, 2000) suggested that although the different *HEV* formulas might yield different scores, nevertheless these scores were highly correlated. These analyses of real data seemed more relevant to Burns and Viglione's (1996) study than the earlier simulated analyses of artificial data. However, it still seemed to us that the analysis of Meyer's 232 patients was no substitute for a reanalysis of Burns and Viglione's data. For example, Meyer's sample consisted of patients, whereas Burns and Viglione's sample consisted of nonpatients in the community. Furthermore, before analyzing their data, Burns and Viglione had dropped 33% of their participants who had scored in the middle range on a measure of interpersonal relatedness. By contrast, Meyer's sample was not administered this measure, and he did not drop the middle 33% of his participants before he carried out his analyses. Meyer contacted us by e-mail within a few days of the date that our article was published and mailed out. In response, we explained our viewpoint to him:

> We requested Burns and Viglione's original data, but were informed that the data were lost. Later you posted analyses to a Rorschach list regarding a *separate* [emphasis in original] data set (not the data set of Burns and Viglione). In my opinion, your posting was an insufficient and unconvincing substitute for the real thing i.e. a chance to examine Burns and Viglione's data set for ourselves.
>
> In your message, you ask how "in good conscience" we could criticize Burns and Viglione on this point, in light of your analyses. Although you seem to see it as an ethical or moral issue, we see it as an intellectual issue: In our view, we are acting reasonably even if we fail to find your analyses as compelling as you do. There is no issue of "conscience" here: You find your numbers highly convincing, but we are still in considerable doubt. (J. M. Wood, personal communication, June 1, 1999)

As may be seen, at the time that our article was published, we firmly believed that what we had published was correct. However, in preparing this response to Meyer's (2000) critique, we decided that it was important to conduct a closer examination of the mathematical issues that he had raised regarding the *HEV* formulas. Accordingly, in September 2000, approximately a month after the publication of Meyer's critique, we conducted several Monte Carlo studies and other analyses similar to those he had reported. We compared the *HEV* formula based on Haller's (1982) means and standard deviations (Haller *HEV*) with the *HEV* formula based on the means of Perry and Viglione (1991; Perry *HEV*). In light of these new analyses, we find that we agree with Meyer on some issues but continue to disagree on others. The following five points seem most important.

First, it is definitely true, as we said in our article (Wood et al., 1999), that the Haller *HEV* formula and the Perry *HEV* formula can yield scores that are not "identical or even very close" (p. 119). Interested readers are invited to examine this issue for themselves. In Meyer's (2000, p. 72) critique, the relevant formulas are given for the Haller *HEV* (which is designated in the critique as Correct *HEV* traditional *z* score) and for the Perry *HEV* (which is designated as Faulty Wood et al. *HEV z* score). If a *Poor H* score of 4 and a *Good H* score of 6 are inserted into the formulas, the Haller formula yields an *HEV* score of –2.44, whereas the Perry formula yields a score of –1.73. The difference between the two scores is .71 (i.e., approximately .4 *SD*). Because the two *HEV* formulas do not yield equivalent results, they should not be substituted for each other in clinical work or published research (e.g., the article by Burns & Viglione, 1996).

Second, it is also true, as we said in our article (Wood et al, 1999), that the two formulas "can change the order of the *HEV* scores" (p. 119). For example, consider two patients, A and B. Patient A gives no *Poor H* responses and three *Good H* responses, whereas Patient B gives seven *Poor H* responses and eight *Good H* responses. When these numbers are inserted into the Haller formula, Patient A has an *HEV* score of –2.21, which is higher than Patient B's score of –2.43. However, when the same numbers are inserted into the Perry formula, Patient A has an *HEV* score of –1.73, which is lower than Patient B's score of –1.60. As may be seen, the order of scores can change depending on which *HEV* formula is used. However, our recent analyses indicate that order is likely to change only for scores that are already very close to each other. For instance, the *HEV* scores of Patients A and B are less than .1 *SD* apart. If the scores were much further apart, they probably would not change order from one *HEV* formula to the other.

Third, although the Perry and Haller *HEV* formulas do not yield identical results, our analyses nevertheless indicate that the scores calculated with the Perry formula are an approximate (but not exact) linear transformation of scores calculated with the Haller formula. Specifically, the Perry *HEV* score of a particular patient can be estimated with considerable accuracy from that patient's Haller *HEV* score, using the following linear formula:

$$\text{Perry } HEV \text{ score} = (.74 \times \text{Haller } HEV \text{ score}) - .15$$

Fourth, because the Perry *HEV* score is an approximate linear transformation of the Haller *HEV* score, it is likely that any statistical test will yield very similar (but not identical) results, whichever of the two scores is used. In our article (Wood et al., 1999) we asserted that "the statistical results of the study by Burns and Viglione (1996) could change depending on which scoring method was actually used" (p. 119). Although this assertion was technically correct, we now believe that it overstated the case. Specifically, in light of our recent analyses, we believe that if Burns and Viglione's data were available, and if the relevant logistic regressions were performed, then the statistical results would be very similar, whether the Perry formula or the Haller formula were used.

Fifth and finally, it is worth reemphasizing the conclusion of our original article (Wood et al., 1999). The logistic regressions in the study by Burns and Viglione (1996) yielded negative results: The *HEV* did not significantly increase predictive power, after controlling for other Rorschach variables. In light of our recent analyses, it appears that the same negative results would have been obtained no matter which version of the *HEV* had been used. Thus, even though our views have changed regarding the relation of the various *HEV* formulas, our article's substantive conclusions regarding the incremental predictive power of the *HEV* remain the same.

## WHAT WILL ADVANCE THE SCIENTIFIC DEBATE REGARDING THE CS FOR THE RORSCHACH?

The CS for the Rorschach (Exner, 1991, 1993) is currently the subject of intense scientific controversy. For example, in the past 2 years, exchanges between Rorschach proponents and Rorschach critics have appeared in three major journals (*Assessment*, *Journal of Clinical Psychology*, *Psychological Assessment*), and further exchanges are forthcoming in three other journals (*Clinical Psychology: Science and Practice, Journal of Personality Assessment, Journal of Forensic Psychology Practice*).

In concluding this rejoinder, we wish to suggest that certain tactics of argumentation are unlikely to advance the scientific debate regarding the CS. First, it seems unlikely that the debate will be advanced by questioning the intentions or intellectual honesty of Rorschach critics, as in Meyer's (2000) commentary. Ad hominem arguments that focus on personalities rather than substantive issues contribute little to the scientific evaluation of the CS and may even act as a distraction. Moreover, it is generally unproductive to attribute scientific disagreements to intentional acts of dishonesty: Such attributions are often mistaken (as in this case) and can introduce unnecessary emotion into scientific discourse.

Second, we question whether it is wise to publish materials from the peer review process that are usually kept confidential, as in Meyer's (2000) commentary. When preparing this rejoinder, we contacted several colleagues in the field with considerable editorial experience, including the editor and the action editor of the journal to which we originally submitted our article (Wood et al., 1999). They affirmed the view that the materials connected with the peer review process are provided to reviewers only for purposes of review, and should otherwise be kept confidential. It would seem that such confidentiality probably should not be breached without the permission of all participants.

In our opinion, the debate regarding the CS is best carried out according to the traditional procedures of scientific discourse, which include the presentation of data and rational arguments. The scientific method advances by vigorous debate. It is best if that debate remains civil and restrained, with a focus on substantive issues rather than individuals.

## ACKNOWLEDGMENT

We wish to thank Debra Corey, Howard Garb, John Hunsley, Scott Lilienfeld, and Stephen West for their helpful comments on an earlier version of this article.

## REFERENCES

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 International, 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.

Exner, J. E. (1991). *The Rorschach: A comprehensive system: Vol. 2. Interpretation* (2nd ed.). New York: Wiley.

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

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: Lawrence Erlbaum Associates, Inc.

Haller, N. (1982). *The reliability of Rorschach depressive indices in major depressive disorder.* Unpublished doctoral dissertation, United States International University, San Diego, CA.

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

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.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach inkblot method. *Psychological Assessment, 8,* 206–213.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). *Assessment, 6,* 115–120.

James M. Wood
Department of Psychology
University of Texas at El Paso
El Paso, TX 79968
E-mail: jawood@utep.edu

Received October 21, 2000