
Limitations of the Rorschach as a Diagnostic Tool: A Reply to Garfield (2000), Lerner (2000), and Weiner (2000)



James M. Wood

University of Texas at El Paso



Scott O. Lilienfeld

Emory University



Howard N. Garb

*Pittsburgh V.A. Health Care System and
University of Pittsburgh*



M. Teresa Nezworski

University of Texas at Dallas

In "The Rorschach Test in Clinical Diagnosis: A Critical Review, With a Backward Look at Garfield (1947)," we have shown that the Rorschach has little validity as a diagnostic tool. In the present piece, we respond to comments by Garfield (2000), Lerner (2000), and Weiner (2000). Until very recently, Rorschach proponents have claimed that the test is useful for diagnostic purposes. It is striking, therefore, that the commentators on our article do not dispute strongly its conclusion that Rorschach scores generally are unrelated to psychiatric diagnoses. Instead, one commentator argues that the test's true usefulness consists in identifying symptoms and predicting behavioral outcomes. However, only three specific examples are given to support this assertion. Although the Rorschach may be useful for these other purposes, the burden of proof falls squarely on the test's proponents to document such claims. © 2000 John Wiley & Sons, Inc. *J Clin Psychol* 56: 441-448, 2000.

Correspondence concerning this article should be addressed to James M. Wood, Department of Psychology, University of Texas at El Paso, El Paso TX 79968; e-mail: jwood@utep.edu.

In "The Rorschach Test in Clinical Diagnosis: A Critical Review, With a Backward Look at Garfield (1947)," we surveyed the recent research literature on the Rorschach and psychiatric diagnoses (Wood, Lilienfeld, Garb, & Nezworski, 2000). At the request of the coeditor of the *Journal of Clinical Psychology*, we began our article by briefly reviewing a Rorschach study published by Sol Garfield (1947) in the same journal over fifty years ago. We were surprised and delighted when, after submitting our article, we were told that Garfield himself would write a rebuttal. It is truly impressive that, a half century after his study was published, Garfield is still making active contributions to the field of clinical psychology. We understand from his remarks that Garfield (2000) regards our treatment of his study as "overly haughty." Perhaps our tone could have been a bit gentler. We hold Sol Garfield in high esteem and intended no disrespect.

As to the substantive points, however, we still believe that our criticisms were correct. Garfield (2000) agrees that criterion contamination was a problem for his 1947 study. However, he does not find it problematic that he was the experimenter, Rorschach administrator, and scorer. In response, we can say only that the influence of administration and experimenter effects on Rorschach results has been well documented (Masling, 1960/1992). For this reason, Rorschach researchers are urged to keep test administrators and scorers blind to patient diagnoses and research hypotheses (Exner & Sendin, 1997).

Garfield (2000) contends that there were no hypotheses in his study. However, we respectfully must disagree. Whether explicitly stated or not, there was a hypothesis, specifically that the Rorschach is valid as a diagnostic tool. When he carried out his study, Garfield apparently believed in the correctness of the hypothesis: Otherwise he would not have been using the Rorschach in his clinical work or examining the agreement between his Rorschach-based diagnoses and final staff diagnoses. Even now, though disheartened by research findings, he seems to hold some respect for the test as a clinical tool. If Garfield was working from an implicit hypothesis that the Rorschach is valid, and if (as would be natural) he hoped to provide support for this hypothesis, then there was even more need for him to remain aloof from the data-gathering process. Of course, we sympathize with Garfield's point that he was administering the Rorschach as part of his work and simply tried to make use of the data for research purposes. We admire such an orientation to the scientist-practitioner model. However, the data may have been more problematic than Garfield recognized.

Garfield (2000) wonders how we came to focus on the Rorschach's relationship to 10 DSM-IV diagnoses out of a possible 350. The answer is that we focused on diagnoses that have been studied in published articles. There seems to be little replicated peer-reviewed research on the relationship of the Rorschach to such diagnoses as Encopresis, Autistic Disorder, Narcolepsy, or Dyspareunia. If there were such research, we would have included it in our article. Moreover, Garfield's point only underscores the danger of using the Rorschach to assess the other 340 DSM-IV diagnoses.

Garfield (2000) objects to our description of a .90 correlation between Rorschach diagnoses and staff diagnoses. In the original version of our article, upon which Garfield is commenting, we did report a .90 correlation. However, before submitting the final version of the article to the editor and before receiving Garfield's comments, we performed a recalculation and discovered that the correct correlation was .70. This figure coincides closely with the 71.9% agreement figure that Garfield uses. The .70 correlation still strikes us as simply too good in light of other Rorschach data. For example, in analyses of data from 263 psychiatric patients kindly provided to us by Gregory Meyer (personal communication, June 1, 1999), the Schizophrenia Index correlated .18 with schizophrenia diagnoses and .32 with psychotic diagnoses in general. We suspect that in Garfield's study (1947) the correlation was inflated substantially by some of the meth-

odological factors identified in our article, especially criterion contamination and the influence of non-Rorschach observational data. Furthermore, we disagree with Garfield that such observational data should be combined with data obtained from the Rorschach per se. Observational data indeed may be helpful in assessing some psychiatric diagnoses, but if so, the failure to eliminate such data from research designs will result in an overestimate of the Rorschach's validity.

Garfield (2000) criticizes our article for not paying attention to the training and skill of the Rorschach clinicians in the studies that we reviewed. In response, we pose the question, "If Rorschach scores are poorly related to diagnoses, can we expect that clinicians with presumed expertise in the Rorschach will be able to make more accurate diagnoses than other clinicians?" The results from empirical research indicate that the answer is no: Presumed expertise, experience, and training have not been found to be related to the validity of judgments made by clinicians using the Rorschach (Garb, 1989, 1998). The burden of proof lies with Garfield or others to demonstrate the contrary.

One of the most interesting points in Garfield's (2000) rebuttal is his suggestion that there is something a little problematic about a "psychometric approach" to the Rorschach. It is worth asking who has promoted and encouraged the idea that the Rorschach is a psychometric instrument. For the past 25 years, Rorschach proponents (particularly advocates of the Comprehensive System) have advanced the claim that the Rorschach is psychometrically sound and possesses excellent reliability and validity. Indeed, when John Exner was granted an award by the American Psychological Association a few years ago, the commendation that appeared in the *American Psychologist* claimed that he had resurrected "perhaps the single most powerful psychometric instrument ever envisioned" (Board of Professional Affairs, 1998, p. 392).

In a recent article published in the journal *Assessment*, we critically examined the evidence for the Rorschach's supposed psychometric excellence (Wood & Lilienfeld, 1999; see also Garb, 1999; Wood, Nezworski, & Stejskal, 1996). In our opinion, claims regarding the Comprehensive System's reliability and validity often have been overstated greatly. Perhaps there is common ground between Garfield and us, in that he also seems to question the prevailing view of the Rorschach as a psychometric instrument. In addition, we concur with him that the Rorschach should never be used by itself to make formal diagnostic recommendations.

Whether the Rorschach is viewed as a psychometric instrument, a clinical "tool," or a "method," the same fundamental question still remains: Which Rorschach scores and responses have shown a well-demonstrated relationship to diagnoses? As our review indicates, there seems to be little relationship between Rorschach scores and most forms of psychopathology. As either a diagnostic "test" or a diagnostic "method," the Rorschach seems to have very limited validity. We do not believe it should be used generally in the diagnostic process.

We turn next to Paul Lerner's (2000) critique. Lerner does not respond to our article by citing research or showing that our reasoning is in error. Instead, he seems to say that we are being "disrespectful" by writing about the Rorschach without first immersing ourselves in its clinical richness. In his view, we are like ill-mannered spectators at a ball game, whereas Rorschach proponents are like players full of "love and passion" for the sport. In response, we freely concede that we approach the Rorschach in a scientific spirit, without much love or passion. Indeed, the essence of the scientific method is the use of systematic procedures designed to prevent oneself from finding what one fervently wishes to find (Sagan, 1995). In the absence of supportive data, we question claims concerning the Rorschach's supposed richness. The research cited in our article indicates

that the test is generally useless for diagnosis. Lerner (2000) does not present evidence to the contrary.

Finally, we turn to Weiner's (2000) critique. We were somewhat surprised to read Weiner's contention that the Rorschach "is not a diagnostic test, if diagnosis means DSM classification." The question arises: If the Rorschach is not supposed to be related to diagnoses, then why have hundreds of researchers (many of them cited in our article) so energetically pursued this issue?

In fact, Rorschach proponents have encouraged strongly the view that Rorschach scores are related to diagnoses. For example, in a recent article, Weiner himself (1997, pp. 10–11) included a section entitled "Differential Diagnosis." He stated:

At present the Rorschach Comprehensive System provides indices for schizophrenia (SCZI) and depression (DEPI) that can prove helpful in identifying these two conditions . . . Recent work by Gacono and Meloy (1994) suggested that a similarly sound and useful index of psychopathic personality can now be constructed . . . In addition, although further documentation is needed, accumulating data indicate that there are on the horizon adequately conceptualized and empirically valid Rorschach indices for bipolar disorder, borderline and schizotypal personality disorder, and acute and chronic stress disorder . . .

Other Rorschach proponents have made similar claims. For example, Exner (1991, p. 146) has reported that a high score on the Depression Index "correlates very highly with a diagnosis that emphasizes serious affective problems." Levin (1993, pp. 189–190) says, "the Rorschach is ideally suited for assessment of PTSD." Meloy and Gacono (1995, p. 414) claim that the Rorschach is "a sensitive instrument to discriminate between psychopathic and nonpsychopathic subjects."

As may be seen, until very recently Rorschach advocates enthusiastically have promoted the Rorschach as a diagnostic tool. Why does Weiner (2000) now seem to retreat from such claims? Perhaps because the evidence cited in our article is so bleak. The Depression Index has shown little or no relationship to diagnoses of depression. Controlled studies have failed to identify any Rorschach variables with a consistent relationship to PTSD. The supposed markers of psychopathy identified by Meloy and Gacono (1995) generally have failed to demonstrate validity in replication studies.

Negative findings not only undermine the Rorschach as a diagnostic tool, but also cast doubt on the construct validity of several important scores. For example, inanimate movement responses (*m*), diffuse shading responses (*Y*), and the *D* score are supposedly related to anxiety and stress (McCown, Fink, Galina, & Johnson, 1992; Perry et al., 1995; but see Frank, 1978, 1993a, 1993b). Yet research does not indicate that these Rorschach variables are related significantly to PTSD or other anxiety disorders. To our thinking, these null findings strongly suggest that *m*, *Y*, and *D* really are not related to anxiety and stress at all. Similarly, Space responses (*S*) are supposedly related to negativism, oppositionality, and an angry attitude toward the environment (Exner, 1991, p. 199), whereas an absence of Texture responses (*T*) is thought to indicate reluctance "about creating or maintaining close emotional ties with others" (Exner, 1991, p. 184). One would therefore expect Space responses and Textureless protocols to be more common among psychopaths or individuals with antisocial personality disorder or conduct disorder than among other individuals (Gacono & Meloy, 1991). However, research has not corroborated these hypotheses. Again, the null findings seem to indicate that *S* and *T* do not measure what they are purported to.

Weiner (2000) suggests that null findings regarding the Rorschach and psychiatric diagnoses may reflect problems with the diagnoses themselves, rather than with the Rorschach. However, it is important to note that MMPI and MMPI-2 scales have shown a

consistent and replicable correlation of over .45 with psychiatric diagnoses (Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999). Such findings indicate that the diagnoses possess sufficient reliability and validity to allow at least moderate correlations with test scores. The question arises, "Why have substantial positive correlations been obtained for the MMPI and MMPI-2, but not for the Rorschach?" The reason, it would seem, is not the invalidity of the diagnoses, but rather the invalidity of the Rorschach for most diagnostic purposes.

In our article, we discuss a meta-analysis by Hiller et al. (1999), which found that the correlation with diagnoses was .18 for the Rorschach versus .47 for the MMPI. Weiner (2000) accuses us of "reporting the results of the Hiller et al. meta-analysis selectively." In our article, we cited the Hiller et al. meta-analysis on the single point that was relevant to our topic, the poor relationship of the Rorschach to psychiatric diagnoses. There was no selective reporting: The other meta-analytic findings were peripheral to our subject. However, now that Weiner has opened the door and reported additional findings from the Hiller meta-analysis, we will report several more. First, as Weiner correctly states, the unweighted mean-effect sizes in the Hiller meta-analysis were .29 for the Rorschach versus .30 for the MMPI. However, Weiner fails to mention that the *weighted* mean-effect sizes were .26 for the Rorschach versus .37 for the MMPI. Major textbooks on meta-analysis (e.g., Hedges & Olkin, 1985, chapter 6; Hunter & Schmidt, 1990, Chapter 11; Shadish & Haddock, 1994) do not recommend or even mention calculating unweighted estimates of effect sizes. Weighted mean-effect sizes generally are recommended because more weight is given to studies with large sample sizes (which typically yield more accurate parameter estimates) than to studies with small samples.

Several other findings also can be reported. The Hiller et al. meta-analysis (1999) found that the Rorschach did not correlate highly with self-report instruments. The mean validity coefficient was .28 for the Rorschach. By comparison, the corresponding validity coefficient for the MMPI was .48. Another finding was more surprising: The Rorschach correlated only .03 with other projective tests. The corresponding figure for the MMPI was .20. Thus, the findings indicated that the Rorschach had poor concurrent validity with both self-report and other projective tests. Lastly, Hiller et al. did not find significantly greater effect sizes for the Comprehensive System (CS) than for other Rorschach methods. This finding, and a similar finding by Garb, Florio, and Grove (1998), cast doubt on the common assumption that the CS is more valid than other Rorschach approaches.

As we note in our article, the Hiller et al. meta-analysis (1999) contains serious methodological flaws, so its findings cannot be considered definitive (see also Garb, 1999). However, if Rorschach proponents wish to cite this meta-analysis, they need to discuss all its findings, not just those that portray the Rorschach in a positive light.

If the Rorschach is unrelated to most psychiatric diagnoses, then what is it useful for? Weiner (2000) argues that the Rorschach is related to psychiatric symptoms and is useful for predicting behavioral outcomes. Of course this assertion invites the question, "Which symptoms and behavioral outcomes specifically?" Weiner names only three: He says that the Rorschach is useful for measuring thought disorder, predicting treatment outcomes, and measuring dependence. All three of these examples involve Rorschach scales or scores that are *not* included in the Comprehensive System as it has been described in recent books (Exner, 1991, 1993; Exner & Weiner, 1995). We will discuss each of Weiner's three examples in turn.

First, Weiner (2000) asserts that the Thought Disorder Index (TDI; Johnston & Holzman, 1979) is related to thought disorder. We agree. In our article, we discuss how deviant verbalizations on the Rorschach (which are measured by the TDI) are related to

schizophrenia, and probably bipolar disorder, as well as schizotypal and borderline personality disorders. However, it is not clear that the TDI can improve the accuracy of psychiatric diagnoses in clinical work beyond what can be learned from a standard interview and a self-report instrument such as the MMPI. The TDI may be useful mainly as a research tool. Moreover, Weiner's invocation of the TDI findings is puzzling. Earlier in his comment, Weiner criticizes findings based on psychiatric diagnoses, largely on the ground that such diagnoses are heterogeneous. Yet he acknowledges that thought disorder is highly nonspecific and is observed in a variety of different psychiatric conditions. Thus Weiner dismisses negative findings because the criteria (e.g., diagnoses) are heterogeneous, but embraces positive findings when the criteria (e.g., thought disorder) are probably even more heterogeneous.

Second, Weiner (2000) asserts that Rorschach scores are related to treatment outcome. Though he provides no citation to support this claim, Weiner is apparently referring to the Rorschach Prognostic Rating Scale (RPRS). We agree that the research literature generally supports the validity of the RPRS (Goldfried, Stricker, & Weiner, 1971; Meyer & Handler, 1997). However, most studies on the topic are rather old and have serious methodological limitations (Hunsley & Bailey, 1999). Furthermore, the RPRS lacks current norms and depends on complicated scoring rules from the Klopfer system. Rigorous, replicated research is needed to determine whether the RPRS is valid when used as part of the CS. If it is, then normative data will need to be collected, cut points established, and sensitivity and specificity determined.

Third, Weiner (2000) asserts that the Rorschach Oral Dependency scale (ROD) is a valid measure of dependency. As we discuss in our article, the research completed thus far on the ROD is encouraging (Bornstein, 1996). However, there are two problems. First, only one published study has examined the relationship of the ROD to pathological dependency as described in the DSM criteria for Dependent Personality Disorder. Diagnoses were established by questionnaires, a problematic procedure as we discuss in our article. Second, a single researcher and his former graduate student have conducted virtually all research on the ROD during the past two decades. As our article argues, independent replications are extremely important in evaluating the Rorschach's validity, because such replications render less likely the possibility that positive findings are attributable to systematic methodological errors on the part of one research team. We urge researchers in all parts of the country to undertake replications using the ROD.

In closing, we wish to express our gratitude to Sol Garfield, Paul Lerner, and Irving Weiner for participating in the present debate. We hope that dialogues between Rorschach proponents and Rorschach critics prove to be fruitful. We also wish to thank Larry Beutler, coeditor of this journal, who has fostered the present exchange of ideas.

References

- Board of Professional Affairs. (1998). Awards for distinguished professional contributions: John Exner. *American Psychologist*, 53, 391-392.
- Bornstein, R.F. (1996). Construct validity of the Rorschach Oral Dependency Scale: 1967-1995. *Psychological Assessment*, 8, 200-205.
- 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., & Sendin, C. (1997). Some issues in Rorschach research. *European Journal of Psychological Assessment*, 13, 155-163.

- Exner, J.E., & Weiner, I.B. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.
- Frank, G. (1978). On the validity of hypotheses derived from the Rorschach: III. The relationship between shading and anxiety. *Perceptual and Motor Skills*, 46, 531–538.
- Frank, G. (1993a). On the meaning of movement responses on the Rorschach. *Psychological Reports*, 73, 1219–1225.
- Frank, G. (1993b). On the validity of hypotheses derived from the Rorschach: The relationship between shading and anxiety, update 1992. *Psychological Reports*, 72, 519–522.
- Gacono, C.B., & Meloy, J.R. (1991). A Rorschach investigation of attachment and anxiety in anti-social personality disorder. *Journal of Nervous and Mental Disease*, 179, 546–552.
- Garb, H.N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, 105, 387–396.
- Garb, H.N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H.N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test. *Assessment*.
- 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.
- Garfield, S.L. (1947). The Rorschach test in clinical diagnosis. *Journal of Clinical Psychology*, 3, 375–381.
- Garfield, S.L. (2000). The Rorschach test in clinical diagnosis—A brief commentary. *Journal of Clinical Psychology*, 56, 431–434.
- Goldfried, M.R., Stricker, G., & Weiner, I.B. (1971). *Rorschach handbook of clinical and research applications*. Englewood Cliffs, N.J.: Prentice-Hall.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
- Hiller, J.B., Rosenthal, R., Bornstein, R.F., Berry, D.T.R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11, 278–296.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11, 266–277.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Johnston, M., & Holzman, P.S. (1979). *Assessing schizophrenic thinking*. San Francisco: Jossey-Bass.
- Lerner, P.M. (2000). A nonreviewer's comment: On the Rorschach and baseball. *Journal of Clinical Psychology*, 56(3), 439.
- Levin, P. (1993). Assessing posttraumatic stress disorder with the Rorschach projective technique. In J.P. Wilson and B. Raphael (Eds.), *International handbook of traumatic stress syndromes* (pp. 189–200). New York: Plenum Press.
- Masling, J. (1992). The influence of situational and interpersonal variables in projective testing. *Journal of Personality Assessment*, 59, 616–640 (Original work published in *Psychological Bulletin*, 1960, 57, 65–85).
- McCown, W., Fink, A.D., Galina, H., & Johnson, J. (1992). Effects of laboratory-induced controllable and uncontrollable stress on Rorschach variables *m* and *Y*. *Journal of Personality Assessment*, 59, 564–573.
- Meloy, J.R., & Gacono, C.B. (1995). Assessing the psychopathic personality. In J.N. Butcher (Ed.), *Clinical personality assessment* (pp. 410–422). New York: Oxford University Press.
- 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.

- Perry, W., Sprock, J., Schaible, D., McDougall, A., Minassian, A., Jenkins, M., & Braff, D. (1995). Amphetamine on Rorschach measures in normal subjects. *Journal of Personality Assessment*, 64, 456–465.
- Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. New York: Random House.
- Shadish, W.R., & Haddock, C.K. (1994). Combining estimates of effect size. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.
- Weiner, I.B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment*, 68, 5–19.
- Weiner, I.B. (2000). Using the Rorschach properly in practice and research. *Journal of Clinical Psychology*, 56, 435–438.
- Wood, J.M. and Lilienfeld, S.O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment*, 6, 341–349.
- Wood, J.M., Lilienfeld, S.O., Garb, H.N., & Nezworski, M.T. (2000). The Rorschach Test in clinical diagnosis: A critical review with a backward look at Garfield (1947). *Journal of Clinical Psychology*, 56, 395–430.
- Wood, J.M., Nezworski, M.T., & Stejskal, W.J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science*, 7, 3–10.