Call for a Moratorium on the Use of the Rorschach Inkblot Test in Clinical and Forensic Settings

Howard N. Garb

Assessment 1999; 6; 313
DOI: 10.1177/107319119900600402

ASSESSMENT
1999, Volume 6, Number 4, 313-315

CALL FOR A MORATORIUM ON THE USE OF THE RORSCHACH INKBLOT TEST IN CLINICAL AND FORENSIC SETTINGS

Howard N. Garb
Pittsburgh V.A. Health Care System
and
University of Pittsburgh

A call is issued for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic (but not research) settings. The moratorium should last until we have determined which Rorschach scores are valid and which ones are invalid. Unfortunately, for most Rorschach scores, results from meta-analyses have been uninformative. Also, incremental validity has not been studied for most Rorschach scores. Furthermore, positive findings for Rorschach scores have rarely been independently replicated. Finally, selective reporting of results has been a problem: Some investigators report significant results but not nonsignificant results. The magnitude of this problem has not been determined. Unless a moratorium is adopted, clinicians will continue to interpret invalid scores along with valid scores.

Keywords: Test, validity, Rorschach, projective test, psychological test

Trying to decide whether the Rorschach is valid is like looking at a Rorschach Inkblot. The results from research are ambiguous just as Rorschach Inkblots are ambiguous. Different people look at the research and see different things, just as clients look at inkblots and see different things. My own point of view is that after 70 years, we still know little about the validity of the Rorschach.

This article is based on a talk that was presented as part of a symposium on the Rorschach at the 107th Convention of the American Psychological Association in Boston on August 21, 1999. The symposium was sponsored by the Society for a Science of Clinical Psychology (Section III of the Society of Clinical Psychology of the American Psychological Association). The author appreciates the critical comments he received from Teresa Nezworski and James Wood on an earlier draft of this article.

Correspondence concerning this article and requests for offprints should be addressed to Howard N. Garb, Behavioral Health (116A-H), V.A. Medical Center, 7180 Highland Drive, Pittsburgh, PA 15206-1297. E-mail: garb.howard@pittsburgh.va.gov

During the first part of my speech, I will talk about meta-analyses. Later, I will address three other topics: incremental validity, the need to replicate findings, and the selective reporting of results.

Meta-Analyses

At least 10 meta-analyses have been conducted on the validity of the Rorschach. To a surprising degree, the results are ambiguous.

The most widely known meta-analysis was conducted by Parker, Hanson, and Hunsley (1988). Their results were published in *Psychological Bulletin*. They reported that the Rorschach is as valid as the MMPI. Their article is frequently cited to defend the Rorschach. For example, just last year, the Parker et al. meta-analysis was cited in a report sponsored by the Board of Professional Affairs of the American Psychological Association

(Meyer et al., 1998). The work group concluded that the Rorschach is as valid as the MMPI. This conclusion was based largely on the results of the Parker et al. article.

Unfortunately, the meta-analysis conducted by Parker et al. (1988) is flawed. One problem is that results that were obtained by calculating correlations were not pooled with results that were obtained by calculating *t* tests, *F* ratios, or other statistics. Textbooks on meta-analysis (e.g., Cooper & Hedges, 1994) uniformly recommend that these types of results be pooled.

My colleagues and I reanalyzed the data from the Parker et al. (1988) meta-analysis (Garb, Florio, & Grove, 1998; also see Garb, Florio, & Grove, 1999; Parker, Hunsley, & Hanson, 1999). Our results were published in *Psychological Science*. We used the effect size estimates that Parker et al. (1988) had calculated, but we pooled the *t*, *F*, and chi-square results with the correlation results. When we reanalyzed the Parker et al. (1988) data, the MMPI was more valid than the Rorschach. We obtained correlation coefficient equivalents of .29 for the Rorschach and .48 for the MMPI. The difference is statistically significant.

Though we found that the MMPI is more valid than the Rorschach, limitations of our meta-analysis have to be acknowledged. The studies used in our reanalysis are not current: They were published between 1970 and 1981. Also, questions can be raised about the effect size estimates that were calculated by Parker et al. (1988) and used in our reanalysis. Parker et al. used ω^2 as an effect size measure. Problems can occur when ω^2 is used for this purpose. Thus, the results from our meta-analysis are more suggestive than definitive.

Another meta-analysis can also be described. Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib conducted a meta-analysis that is in press at *Psychological Assessment*. They reported that unweighted mean validity coefficients were .29 for the Rorschach and .30 for the MMPI, but mean validity coefficients weighted by degrees of freedom were .37 for the MMPI and .26 for the Rorschach. Unfortunately, their meta-analysis is also flawed. For example, interrater reliability for the coding judges was poor. Coding judges decided which results would be included in the meta-analysis. They were given the Method sections for each study, and they were to indicate whether one could reasonably expect the predictors to be related to a criterion. Importantly, coding books were not constructed to help the coding judges make their ratings. Even more important, interrater reliability was poor, as reflected by a phi coefficient of only .35. Not only is this a low value, but it is probably an overestimate of reliability. This is because phi will overstate reliability unless the two coders have equal acceptance rates. Instead of (or in addition to) calculating phi, a kappa coefficient should have been calculated.[1]

In addition to problems with interrater reliability, coding judges were not blind to the results of all of the studies. According to Hiller et al. (in press):

> ...it is likely that judges were nevertheless familiar with at least some of the studies included. In fact, several studies included in the meta-analysis were authored by [the coding] judges. (p. 30)

Since coding judges were not blind, it is possible that the selection of results for inclusion in the meta-analysis was biased.

Additional meta-analyses have also been done. The Rorschach Prognostic Rating Scale was supported in a meta-analysis conducted by Meyer and Handler (1997). However, Meyer and Handler did not acknowledge that when the Rorschach Prognostic Rating Scale was compared to the Health-Sickness Rating Scale in the same studies using the same samples, the Health-Sickness Rating Scale did better (Luborsky, Mintz, & Christoph, 1979; also see Luborsky, Crits-Christoph, Mintz, & Auerbach, 1988; Luborsky et al., 1993). The Health-Sickness Rating Scale is not a projective measure. The results of a different meta-analysis were recently published by Bornstein (1999). He looked at the assessment of dependency. Though he found

[1]Following my speech, Jordan Hiller calculated a kappa coefficient for his data. He obtained a value of .34. Though this value is only slightly lower than the value obtained for phi, I still recommend that kappa or an intraclass correlation coefficient be calculated when measuring interrater reliability.

positive results for the Rorschach, the clinical relevance of his findings is unclear. He found that, "Projective test scores were positively correlated with behavioral ratings obtained in laboratory, field, and classroom settings but not in *clinical settings*" [italics added] (p. 52).

One last meta-analysis will be described. Last year in an article published in *Child Abuse & Neglect*, West (1998) described the results of a meta-analysis and concluded that projective techniques can be used to detect child sexual abuse. Studies on the TAT and projective drawings were included along with studies on the Rorschach. West located 12 studies. She obtained a large overall effect size (d = .81).

Unfortunately, the meta-analysis conducted by West (1998) contains a *terrible* flaw (Garb, Wood, & Nezworski, in press). When conducting a meta-analysis, one should not arbitrarily exclude data. West routinely included results that were statistically significant, but excluded results from the same studies that were nonsignificant. She excluded results from 9 of the 12 studies. In 8 of the studies, West excluded half or more than half of the data. Furthermore, in her article, she never stated that she excluded nonsignificant results. We discovered this only when we examined the original studies.

Results from research on incremental validity also offer little support for the Rorschach (Hunsley & Bailey, in press). In fact, some of the results suggest that we should not be using the Rorschach. In an incremental validity study, we investigate whether the addition of the Rorschach to other types of information leads to an increase in accuracy. Given that it takes 2 to 3 hours to administer, score, and interpret a Rorschach, one would hope that the addition of the Rorschach would lead to increased accuracy. However, with the exception of using the Rorschach to predict IQ (Potkay & Ward, 1972), positive results have *never* been obtained for the Rorschach in studies on clinical judgment and incremental validity (Garb, 1984, 1998). In these studies, judgments were made by clinicians. The validity of judgments made by clinicians did not increase when the Rorschach was added to demographic data, other test results, or biographical information (Barendregt, 1961; Bilett, Jones, & Whitaker, 1982; Cochrane, 1972; Gadol, 1969; Golden, 1964; Perez, 1976; Potkay & Ward, 1972). In many instances, the addition of the Rorschach led to a decrease in validity, though the decrements in accuracy were not always tested for statistical significance (e.g., Barendregt, 1961; Gadol, 1969; Golden, 1964; Potkay & Ward, 1972; Sines, 1959).

Limitations of these studies must be acknowledged. Most of the clinical judgment studies are old, and it is unclear if the Comprehensive System (Exner, 1993) was used by any of the clinicians in the studies. However, one can still conclude that the results offer little support for the Rorschach.

Incremental validity has also been studied by using statistical prediction rules. When judgments have been made by using statistical prediction rules, there is some evidence supporting the incremental validity of the Ego Impairment Index (Perry, Moore, & Braff, 1995, reanalyzed by Dawes, in press; Perry & Viglione, 1991), the Rorschach Oral Dependency Scale (Bornstein, Bowers, & Robinson, 1995), and the Rorschach Prognostic Rating Scale (Meyer & Handler, 1997). In other studies, incremental validity was poor when judgments were made by using statistical prediction rules (Archer & Krishnamurthy, 1997; also see Archer & Gordon, 1988). Incremental validity was poor for the following scores: DEPI, Vista, Col-Shd Bld, Egocentricity Index, Afr, MOR, Sum Shading, AG, COP, S, Populars, Texture responses, and D. Overall, for the vast majority of Rorschach scores, incremental validity has not been studied. Thus, it seems fair to conclude that few results on incremental validity support the Rorschach.

The Rorschach controversy will not be resolved until positive findings have been replicated by independent investigators. Three years ago, Wood, Nezworski, and Stejskal (1996b, p. 15) asked the Rorschach community to describe scores that have shown "a consistent relationship to a particular psychological symptom or disorder...in several methodologically adequate validation studies that were...conducted by unrelated researchers or groups." Unfortunately, the Rorschach community has not responded to this invitation. My own opinion is that positive findings have been replicated

for a small number of Rorschach predictors. However, this does not seem to be true for the majority of Rorschach scores that are used by clinicians (Wood, Nezworski, & Stejskal, 1996a).

One reason why positive findings need to be replicated by independent investigators is because some research investigators engage in selective reporting of results. Research investigators who have obtained results for a large number of test scores may report only the statistically significant findings. Selective reporting can be very difficult to detect: If investigators do not report that they collected data on particular test scores, then it will usually be impossible to demonstrate that they did so. However, in a few cases it has been possible to determine that investigators selectively reported statistically significant results. For example, Wixom, Ludolph, and Westen (1993) reported that girls with borderline diagnoses have higher scores than depressed girls on two Rorschach measures. Their article was based on a dissertation by Wixom (1988/1989), which reported results for additional Rorschach variables. Significant results were *not* found for measures of egocentricity or narcissistic injury. For reasons that are unclear, the article by Wixom et al. reported studying only the two Rorschach variables that could be used to differentiate the borderline and depressed girls. The authors omitted mentioning the Rorschach variables with negative findings. For additional examples of selective reporting, see Garb, Wood, & Nezworski (1999), Garb, Wood, Nezworski, Grove, & Stejskal (1999), and Wood, Lilienfeld, Garb, & Nezworski, (1999).

In summary, meta-analyses have provided little support for the Rorschach, results on incremental validity have been disappointing, and positive findings have rarely been replicated by independent investigators. Also, selective reporting of results has been a problem. Similar conclusions were reached by Hunsley and Bailey (in press):

> The reliability and validity of the Comprehensive System has been greatly overstated. The over-reliance on unpublished research, the limited nature of current meta-analytic results, the paucity of replicated evidence, the questionable standards used for evaluating reliability, the problem of variation in response frequency, and the nature of the factor structure of Comprehensive System scales weakens the claims of many that the Comprehensive System has finally legitimized the Rorschach.... At present, the Comprehensive System, as a whole, does not meet the requirements set out in professional standards of practice such as the *Standards for Psychological and Educational Testing* (APA, 1985). (pp. 21-22 of manuscript accepted for publication)

Given these findings, one has to wonder if psychologists should be using the Rorschach. While positive findings have been independently replicated for a small number of Rorschach scores, it is likely that most clinicians also interpret Rorschach scores that are invalid. Thus, I am calling for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic (but not research) settings. This moratorium should last until we have determined which Rorschach scores are valid and which ones are invalid.

References

American Psychological Association. (1985). *Standards for psychological and educational testing*. Washington, DC: Author.

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. (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.

Barendregt, J. T. (1961). *Research in psychodiagnostics*. The Hague, Netherlands: Mouton.

Bilett, J. L., Jones, N. F., & Whitaker, L. C. (1982). Exploring schizophrenic thinking in older adolescents with the WAIS, Rorschach, and WIST. *Journal of Clinical Psychology, 38*, 232-243.

Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment, 11*, 48-57.

Bornstein, R. F., Bowers, K. S., & Robinson, K. J. (1995). Differential relationships of objective and projective dependency scores to self-reports of interpersonal life events in college student subjects. *Journal of Personality Assessment, 65*, 255-269.

Cochrane, C. T. (1972). Effects of diagnostic information on empathic understanding by the therapist in a psychotherapy analogue. *Journal of Consulting and Clinical Psychology, 38*, 359-365.

Cooper, H., & Hedges, L. (Eds.). (1994). *The handbook of research synthesis*. New York: Sage.

Dawes, R. M. (in press). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment*.

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

Gadol, I. (1969). The incremental and predictive validity of the Rorschach test in personality assessments of normal, neurotic, and psychotic subjects. *Dissertation Abstracts, 29*, 3482-B. (University Microfilms No. 69-4469)

Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4*, 641-655.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

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. (1999). The Rorschach controversy: Reply to Parker, Hunsley, and Hanson. *Psychological Science, 10*, 293-294.

Garb, H. N., Wood, J. M., & Nezworski, M. T. (1999). *Projective techniques and the detection of child sexual abuse*. Manuscript submitted for publication.

Garb, H. N., Wood, J. M., & Nezworski, M. T. (in press). Projective techniques and the detection of child sexual abuse. [Letter to the editor]. *Child Abuse & Neglect*.

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (1999). *Towards a resolution of the Rorschach controversy*. Manuscript submitted for publication.

Golden, M. (1964). Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology, 28*, 440-446.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (in press). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*.

Hunsley, J., & Bailey, J. M. (in press). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*.

Luborsky, L., Crits-Christoph, P., Mintz, J., & Auerbach, A. (1988). *Who will benefit from psychotherapy*? New York: Basic Books.

Luborsky, L., Diguer, L., Luborsky, E., McLellan, A. T., Woody, G., & Alexander, L. (1993). Psychological health-sickness (PHS) as a predictor of outcomes in dynamic and other psychotherapies. *Journal of Consulting and Clinical Psychology, 61*, 542-548.

Luborsky, L., Mintz, J., & Christoph, P. (1979). Are psychotherapeutic changes predictable? Comparison of a Chicago Counseling Center project with a Penn Psychotherapy Project. *Journal of Consulting and Clinical Psychology, 47*, 469-473.

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 health-care delivery: Report of the Board of Professional Affairs, Psychological Assessment Work Group, Part I*. Washington, DC: American Psychological Association.

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.

Parker, K. C. H., Hanson, R., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103*, 367-373.

Parker, K. C. H., Hunsley, J., & Hanson, R. K. (1999). Old wine from old skins sometimes tastes like vinegar: A response to Garb, Florio, and Grove. *Psychological Science, 10*, 291-292.

Perez, F. I. (1976). Behavioral analysis of clinical judgment. *Perceptual and Motor Skills, 43*, 711-718.

Perry, W., Moore, D., & Braff, D. (1995). Gender differences on thought disturbance: Measures among schizophrenic patients. *American Journal of Psychiatry, 152*, 1298-1301.

Perry, W., & Viglione, D. J., Jr. (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.

Potkay, C. R., & Ward, E. F. (1972). Clinical judgment under varied informational conditions: Rorschach, personal data, and best guess. *Journal of Consulting and Clinical Psychology, 39*, 513.

Sines, L. K. (1959). The relative contribution of four kinds of data to accuracy in personality assessment. *Journal of Consulting Psychology, 23*, 483-492.

West, M. M. (1998). Meta-analysis of studies assessing the efficacy of projective techniques in discriminating child sexual abuse. *Child Abuse & Neglect*, 22, 1151-1166.

Wixom, J. (1989). The depressive experiences of adolescents with borderline personality disorder. (Doctoral dissertation, University of Michigan, 1988). *Dissertation Abstracts International, 49*, 5038B.

Wixom, J., Ludolph, P., & Westen, D. (1993). The quality of depression in adolescents with borderline personality disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 1172-1177.

Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (1999). *The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947)*. Manuscript submitted for publication.

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.