

THE RORSCHACH CONTROVERSY: A COMPILATION

Cato Grønnerød

Department of Psychology, University of Oslo

First presented in Grønnerød, C. (2004). *Stability and change in the Rorschach method*. Unpublished doctoral dissertation, University of Oslo, Oslo.

Republished under [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/) on 16.05.2014.

The following outline is only intended as a brief and informal guide to the recent dispute over the Rorschach method, and does not intend to give an accurate nor a complete report of the contents of the articles.

During the past few years, several articles concerning the scientific status of the Rorschach method have been published. The debate has been long standing, but a particularly heated debate was sparked by an article in *Psychological Science* in 1996 (Wood, Nezworski, & Stejskal, 1996b) which criticized the CS for misleading estimates of interscorer reliability, lacking validity for a number of important variables, and an inappropriate reliance on unpublished studies. Replies followed in the same issue (Exner, 1996; Wood, Nezworski, & Stejskal, 1996a), and a continued exchange followed in *Psychological Assessment* (Meyer, 1997a, 1997b; Wood, Nezworski, & Stejskal, 1997) on issues of interscorer reliability raised in the original article by Wood et al. Furthermore, another article in *Psychological Science* (Garb, Florio, & Grove, 1998) criticized earlier meta-analyses of the Rorschach method for various faults and biases. Commentaries followed both directly (Garb, Florio, & Grove, 1999; Parker, Hunsley, & Hanson, 1999), and later (Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; Meyer & Archer, 2001).

As a result of these criticisms, special sections appeared in various journals. In the first part of a special section in *Psychological Assessment*, invited proponents and critics were asked to discuss the utility of the Rorschach method for clinical assessment (Dawes, 1999; Hunsley & Bailey, 1999; Meyer, 1999; Stricker & Gold, 1999; Viglione, 1999), and a meta-analysis was commissioned to address the problems pointed out in previous analyses (Hiller et al., 1999).

In a similar special section in *Assessment*, a series of papers presented at the 1999 Annual Convention of the American Psychological Association was rewritten for publication, which discussed various perspectives on the Rorschach method (Acklin, 1999; Archer, 1999; Garb, 1999; Weiner, 1999; Wood & Lilienfeld, 1999). In the same issue of *Assessment*, another critical examination of the Rorschach method was published (Wood, Nezworski, Stejskal, Garven, & West, 1999a; with erratum: Wood, Nezworski, Stejskal, Garven, & West, 1999b). This article was a response to favorable reviews of the Rorschach method published earlier (Burns & Viglione, 1996; Ganellen, 1996; Weiner, 1996), and focused on specific methodological problems in these articles. Meyer (2000) offered a rather sharp reply to Wood and colleagues in *Journal of Personality Assessment*, where he pointed out erroneous citations and faulty formulas. Later, in the same journal, Wood, Nezworski, Stejskal and Garven (2001) wrote a rejoinder to Meyer's criticism, and others were also invited to offer their comments (Bornstein, 2001; Gacono, Loving, & Bodholt, 2001; Ganellen, 2001; Wood, CATO GRØNNERØD © 2002

Lilienfeld, Nezworski, & Garb, 2001).

In parallel, a discussion in *Journal of Clinical Psychology* started with a critical review of the Rorschach method in relationship to clinical diagnoses (Wood, Lilienfeld, Garb, & Nezworski, 2000b), which was followed by comments in the same edition (Garfield, 2000; Lerner, 2000; Weiner, 2000; Wood, Lilienfeld, Garb, & Nezworski, 2000a). A discussion in *Clinical Psychology* concerned the representativeness of the CS norms (Aronow, 2001; Exner, 2001; Hunsley & Di Giulio, 2001; Meyer, 2001a; Widiger, 2001; Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b).

Finally, the second part of the *Psychological Assessment* special section came in 2001, offering comments on issues raised in the first part (Garb, Wood, Nezworski, & Grove, 2001; Hunsley & Bailey, 2001; Meyer, 2001b; Rosenthal, Hiller, Bornstein, & Berry, 2001; Viglione & Hilsenroth, 2001; Weiner, 2001). In a summary to the special section in particular, and to the whole dispute in general, Meyer and Archer (2001) offered several recommendations on future Rorschach research in order to provide better answers to the problematic issues that have been raised.

An article in *Psychological Science* (Garb et al., 1998) criticized earlier meta-analyses, and commentaries followed directly (Garb et al., 1999; Parker et al., 1999), and in later exchanges (Hiller et al., 1999; Meyer & Archer, 2001).

As a result of the initial critical article by Wood et al., several special sections appeared. The first, in *Psychological Assessment*, invited proponents and critics were asked to discuss the utility of the Rorschach method for clinical assessment (Dawes, 1999; Hunsley & Bailey, 1999; Meyer, 1999; Stricker & Gold, 1999; Viglione, 1999), and meta-analysis was commissioned (Hiller et al., 1999). The second part of the special section came in 2001 (Garb et al., 2001; Hunsley & Bailey, 2001; Meyer, 2001b; Meyer & Archer, 2001; Rosenthal et al., 2001; Viglione & Hilsenroth, 2001; Weiner, 2001).

In a similar special section in *Assessment*, a series of papers presented at the 1999 Annual Convention of the American Psychological Association were rewritten for publication, discussing various perspectives on the Rorschach method (Acklin, 1999; Archer, 1999; Garb, 1999; Weiner, 1999; Wood & Lilienfeld, 1999). In the same issue of *Assessment* another critical examination of the Rorschach method was published (Wood et al., 1999a; with erratum: Wood et al., 1999b), in response to favorable reviews published earlier (Bums & Viglione, 1996; Ganellen, 1996; Weiner, 1996). Meyer (2000) offered a rather sharp reply to this article in *Journal of Personality Assessment*. Later, in the same journal, Wood, Nezworski, Stejskal and Garven (2001) wrote a rejoinder to Meyer's criticism, and others were also invited to offer their comments (Bornstein, 2001; Gacono et al., 2001; Ganellen, 2001; Wood, Lilienfeld et al., 2001).

A discussion in *Journal of Clinical Psychology* started with a critical review of the Rorschach method in clinical diagnosis (Wood et al., 2000b), followed by comments in the same edition (Garfield, 2000; Lerner, 2000; Weiner, 2000; Wood et al., 2000a). Finally, a discussion in *Clinical Psychology* concerned the representativeness of the CS norms (Aronow, 2001; Exner, 2001; Hunsley & Di Giulio, 2001; Meyer, 2001a; Widiger, 2001; Wood, Nezworski, Garb et al., 2001a, 2001b).

The following is a summary of each of the articles mentioned, with their main points highlighted. The different exchanges are presented together, ordered by year of publication. Finally, an overview is given of available meta-analyses involving the Rorschach method.

THE 1996 AND 1997 PSYCHOLOGICAL SCIENCE AND PSYCHOLOGICAL ASSESSMENT EXCHANGE

Wood, Nezworski & Stejskal (1996a): The Comprehensive System for the Rorschach: A critical examination

Wood et al. departs from the assumption that because the CS represents a “middle ground” in the Rorschach field, it has been less scrutinized than one might expect for such a popular system. The scrutiny includes interrater reliability, validity, and the nature of the research base.

RELIABILITY

CS presents percentage of agreement, but this estimate has long been recognized as potentially inadequate and misleading measure of reliability. An example is given where a random assignment of 5% of a score yields 90% agreement. It is also pointed out that the total score level is probably more clinically relevant than individual responses. Finally, a note is made that not all variables are reported.

Data for field reliability, defined as how accurate Rorschach protocols are scores in real-life clinical situations, compared to the highly trained experts used in interrater reliability studies. Reliability of administration and recording is also unknown, even though studies show that situational factors can have a strong influence.

VALIDITY

Several variables are claimed to be related to clinically important phenomena, but empirical evidence is still often scant or negative. The most recent version of CS fail to cite negative research findings for the Egocentricity Index. Findings for D and Adj D are either unpublished or equivocal, the same for DEPI (both versions, possibly because of criterion contamination). Examples of shrinking during cross-validation are numerous, DEPI serves the same faith. S-CON showed no shrinkage to a new Exner sample, but no predictive power in an independent replication. R shows a strong correlation with numerous variables. The CS has too many reliance on single-sign interpretations. Such signs are limited by their reliability, and often lack empirical support. More research is needed to establish whether Rorschach variables yields anything in incremental validity.

RESEARCH BASE

63% percent of studies cited in CS is unpublished, only 17% in peer-reviewed journals. Three conclusions were drawn after contacts with the Rorschach Workshops: 1) most citations are

research projects, not written documents, 2) methods and results for many studies are unavailable, and 3) only raw data can be obtained, but only at a high cost.

Recommends the study of ideal and field rater reliability, clinical validation studies using well-defined diagnostic criteria, blinding of judges, and full report of all studies forming the base of the CS.

Exner (1996): A comment on "The comprehensive system for the Rorschach: A critical examination."

Originally, correlations were used, but dismissed because chance adjustments do not reflect the actual scoring rules properly. Percentage agreement is in fact percentage correct (example given, but not very clear). Field reliability is important and has been addressed in quizzes, and subsequently in refinement of score definitions. Criticism against Egocentricity index and Adj D have already been addressed. Admits problems with depression in the Rorschach. Interpretations should never be made on the basis of a single sign. Suggestions that unpublished studies are flawed or mediocre are unwarranted.

Wood, Nezworski & Stejskal (1996b): Thinking critically about the comprehensive system for the Rorschach: A reply to Exner

Interrater reliability figures are unknown, since only scoring accuracy is reported. Computation example does not clarify, and the main problem remains, which is the omission of scorer errors. Exner need to state what variables that are well validated. Meyer's article on DEPI has small good news, reported by Exner, and large bad news, which is omitted. Unpublished research continues to provide the main empirical support for the CS. Unpublished studies are not necessarily flawed or mediocre, they simply cannot be evaluated.

Meyer (1997a): Assessing reliability. Critical corrections for a critical examination of the Rorschach Comprehensive System

It is of the utmost importance to establish interrater reliability at the level of scoring definitions, not summary scores. Because random errors of measurement tend to cancel with aggregation, summary scores will be more reliable than individual responses. Field reliability is important, but actually also often incorporated into research, since many participants are clinicians working in actual clinical settings. Field reliability would be affected by many factors not directly related to scoring rules, and might therefore not have the same value as ideal reliability. The difference between percentage correct and percentage agreement is viable only at the inferential level, not psychometric.

Kappa has been criticized for its definition of chance agreement. Random assignment is defined by each scores base rate, but this cannot be held in mind by scorers over a large set of responses. An alternative way of defining chance is $1/n$, where n is the number of options for a score. Furthermore, kappa is overly sensitive to low base rates. Low base rates will restrict the maximum reliability regardless of the accuracy of the judgements. Kappa also

presupposes between-subject variance, with is in contradiction with criterion-referenced testing.

META-ANALYSIS

JPA 1992-1995, 16 studies, by entering fixed chance agreement estimates into kappa, combined with observed percentage agreement in the studies at hand. Cohen's kappa mean is .86 (.72-.96), kappa_n is .91 (.87-.96), while percentage agreement is .92 (.74-1.0). Percentage agreement does not seem problematic when agreement is substantial.

TEST-RETEST RELIABILITY

Citing Exner, Armbruster and Viglione (1978), Exner (1980; 1993), and Haller and Exner (1985), temporal stability data is reported at the level requested by Wood et al. Raters were blinded in the studies, and results show high temporal stability.

Wood, Nezworski & Stejskal (1997): The reliability of the Comprehensive System for the Rorschach: A comment on Meyer (1997)

Never stated that CS reliability is little better than random chance. Kappa example is “fishy”, since you cannot divide zero by zero. Kappa is the preferred statistic, and no other modification has received widespread acceptance. Field reliability is important to secure the quality of use in clinical settings, although no definite conclusions can be made about the scoring system from this data. Calculation methods for percentage correct in the CS is still unclear, and Meyer's attempt to clarify introduces further confusion. Meyer confused the proportion of true variance in a score with interscorer agreement. Temporal stability does not relate to interscorer reliability. Meyer proposes a segment approach which cannot indicate problematic scores, which is exactly what is needed, and is what Meyer called for when dismissing summary scores as basis for interscorer reliability. Base rate estimation was done by “best guess”, which is a highly dubious method.

Meyer (1997b): Thinking clearly about reliability. More critical corrections regarding the Rorschach Comprehensive System

Wood et al. is retreating from their initial thrust, as they clearly meant to generalize from their examples or inadequate interrater reliability estimates. Kappa = 0/0 is defined as 0 by Cohen. Percentage agreement is no necessarily a preferred estimate, but neither is kappa. Excellent temporal stability presupposes good interscorer agreement, and is therefore relevant. True score theory applies to interscorer reliability, contrary to Wood et al.'s assertions. McDowell and Acklin (1996) finds higher reliability for total scores than for single scores. Wood et al. makes a dramatic misperception when stating that “best guess” chance estimates were used in the meta-analysis. It was not, only the base rate estimates of chance were used, as is evident in the presented table.

THE 1998 AND 1999 PSYCHOLOGICAL SCIENCE EXCHANGE

Garb, Florio & Grove (1998): The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses

Previous meta-analyses cited in support of the Rorschach method are flawed. Parker (1983) did not separate reliability and validity in the analyses, Atkinson, Carrington, Alp and Cyr (1986) did not analyze magnitude of effects, and Atkinson (1986) did not aggregate estimates across studies.

Parker, Hunsley and Hanson's (Parker, Hanson, & Hunsley, 1988) meta-analysis is also problematic for several reasons. Firstly, results were not pooled according to type of statistic, but rather analyzed separately. Secondly, in some studies assessment instrument shared variance with the construct measures, thereby inflating effect sizes. Finally, small effect sizes apparently meant that a test is valid.

REANALYSIS

The terms “unknown” and “convergent” validity are misleading. “Exploratory” and “confirmatory” studies are more appropriate. Effects sizes from Parker et al. were used, but only a single estimate was included from one single study, using the median value. Results indicate that MMPI is more valid than the Rorschach, contrary to Parker et al.'s conclusions.

Parker, Hunsley & Hanson (1999): Old wine from old skins sometimes tastes like vinegar: A response to Garb, Florio, and Grove

The original analysis has been cited as having settled the question of Rorschach validity, but we cannot support this conclusion. Effect sizes were not considered in the coding process. (Explanations are given to how different effects are coded as reliability or validity, unknown or convergent.) Inclusion of scales that shared method variance were done for all tests, and would therefore not affect the relationship between the tests. No clear interpretation is evident for the difference between Rorschach and MMPI found by Garb et al., and the influence of course statistics (r , F , t) may explain the differences. The detail of Garb et al.'s analyses cannot be supported by the data. Neither way does it mean that the Rorschach is invalid. Little is added to the understanding of the relationship between the Rorschach and MMPI by the critics.

Garb, Florio & Grove (1999): The Rorschach controversy: Reply to Parker, Hunsley, and Hanson

Meta-analysis text book uniformly recommends pooled effect sizes across various statistics. No significant differences were found between them. The critical elements were not

addressed, and the results are maintained. Less emphasis should be put on training with the Rorschach.

THE 1999 PSYCHOLOGICAL ASSESSMENT SPECIAL SECTION, PART 1

Meyer (1999): Introduction to the first Special Section in the Special Series on the utility of the Rorschach for clinical assessment

Meyer's point of departure for the discussion: faced with arguments both on the uselessness and the usefulness of the Rorschach method, one would probably find the truth somewhere in between: "...the Rorschach is probably not a valid test for any and every intended purpose nor an invalid test for any and every intended purpose." (Meyer, 1999, p. 236). The main question is: for what purposes is the Rorschach method useful and provides unique clinical information?

Stricker & Gold (1999): The Rorschach: Toward a nomothetically based, idiographically applicable configurational model

How to use test data? "The conversion of data from potentially useful to actively relevant depends more on the nature of the assessment task at hand, the goal of the clinician, the needs of the patient, and the clinician's theoretical perspective than on the nature of the data" (Stricker & Gold, 1999, p.240). By using depth methods and self report measures one might end up with contradictory data. This cannot however be taken as indicating poor validity of the methods used, only to conclude that the methods answer different questions.

The Rorschach method works best together with psychodynamic theories, and there is solid empirical support for the use of a problem-solving approach like the CS. It is not a method for predicting overt behavior, neither for psychiatric taxonomy. On a broad basis it is very well suited for understanding the underlying organization of the personality and for generating hypotheses about unconscious content. The Rorschach method assesses other facets of the target phenomena than MMPI/MMPI-II, Rorschach scores are minimally, if at all, correlated with MMPI/MMPI-II scores (Archer & Krishnamurthy, 1993, review 37 studies) and has a different factor structure (Meyer, 1992). Self report measures and the Rorschach method taps different aspects of phenomenon, such as dependency, depression and motivation.

Evidence exist for the identification of patients who are depressed, schizophrenic or suicidal (Exner, 1993), referring of patients to appropriate form of psychotherapy (Meyer & Handler, 1997), and evaluating intra-personal and inter-personal functioning (Stricker & Healey, 1990).

Hiller, Rosenthal, Bornstein, Berry & Brunell-Neulieb (1999): A comparative meta-analysis of Rorschach and MMPI validity

Atkinson's (1986) analysis used unsatisfactory effect sizes (r^2 , T^2 , Cramér's V). They can only assume positive values, and cannot indicate whether the direction of the result is in accordance

with the predicted association. Several effect sizes were extracted from the same studies and treated as independent, which will lead to problems in significance testing.

The data for convergent validity in Parker et al. analysis were subject to a number of restrictions. One was selection bias, since only studies from two major journals were included. Another was the exclusion of “course” statistics, which is not warranted. Garb et al.'s reanalysis suffers from the same problems regarding the inclusion of T^2 , which is altogether unsuitable for meta-analytic work. Results need not be separated into confirmatory and exploratory. Validity evidence is validity evidence, also when an author fails to provide reasonable predictions.

A meta-analysis is performed, with a random selection of studies from 1977 to 1997 from several journals, together with a few unpublished studies. Similar effect sizes are found for the Rorschach and MMPI. Somewhat higher effects are found for MMPI when self-report inventories are used as criteria, and higher effects are found for the Rorschach when objective measures are used.

Viglione (1999): A review of recent research addressing the utility of the Rorschach

TEMPORAL CONSISTENCY

The great majority of CS variables and configurations show impressive temporal consistency reliability for $R > 14$. Scores reflect development in children.

DIVERSITY

Several studies show that the Rorschach method is a valuable tool in subcultural and cross-cultural evaluations (referencing several studies in non-US cultures).

CLINICAL JUDGEMENT

Clinicians make more accurate judgements when afforded with clearly delineated criteria based on clear theoretical concepts.

TREATMENT OUTCOME

Three CS studies have successfully identified therapy progress, with more progress the longer the therapy (Abraham, Lepisto, Lewis, Schultz, & et al., 1994; Exner & Andronikof-Sanglade, 1992; Weiner & Exner, 1991). The EII has demonstrated predictive power on who would respond to medical treatment of depression (Perry & Viglione, 1991). Other studies have shown predictive power on early termination, who would benefit from therapy, psychotic symptoms, and treatment outcome.

SCHIZOPHRENIA AND PSYCHOSIS

The TDI, EII and SCZI have all proven incremental and discriminative validity for schizophrenia, psychosis and affective disorders over other test and clinical interviews. The SCZI still has a problem of false positives.

DEPRESSION

DEPI is not recommended for routine application for diagnostic purposes, as it is not sensitive enough to clinical depression.

RORSCHACH RESPONSE CONTENT

A large group of studies indicate that human representation content, dependency content, and aggressive content have significant links to pathology measures on behavioral criteria, longitudinal designs, and self report.

CONCLUSION

The Rorschach method can and should be used to predict outcome, to monitor change, to plan treatment, to survey for unanticipated problems, and to set ultimate, intermedia, and instrumental intervention goals. In contrast to Garb's (1984) conclusion, Rorschach variables have shown incremental validity with many utility-relevant criteria, beyond information derived from interviews, diagnoses, self-report tests, and ability tests.

The lack of relationship between the Rorschach method and self-report tests supports the utility of the Rorschach method, in making it more likely to yield incremental validity, and by avoiding problems with false self-report presentations.

Hunsley & Bailey (1999): The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future

The Rorschach method is one of the most expensive test used, as it has a modal time usage of three hours. The extent and type of use of the CS in clinical settings is unclear. Many clinicians choose to rely more on their personal experience than the scientific base and psychometric properties when choosing tests. Given the accumulating evidence of the negative effects of biases and heuristics on clinical judgement, and overwhelming evidence demonstrating the superiority of empirically derived decision-making rules over clinical judgement, assessment practices must to a larger extent be grounded in empirical evidence.

Satisfactory interrater agreement seems to be established, although the calculation should be standardized to a greater degree.

META-ANALYSES

Parker's (1983) and Atkinson's (1986) meta-analyses are not valid, as their statistical technique has shown to be flawed. They therefore only have historical merits. Parker et al. (Parker et al., 1988) used appropriate statistics, but focused only on nine scales from the CS, and subsequent analyses of their data (Garb et al., 1998) showed that the typical validity of the Rorschach scales were significantly lower than for the MMPI. Atkinson et al. (1986) found that about half of the variables in 120 studies proved significant for conceptually based studies, only one fifth for those lacking. Ornberg & Zalewski (1994) presented limited evidence to the clinical utility of the Rorschach method. The mentioned studies have all been cited frequently by others as evidence to the utility of the Rorschach method, but the evidence is invalid or extremely limited.

CS

Response frequency is a major problem, and might account for half of the explained variance in Rorschach variables. Research using the CS is however more powerful than non-CS

research (Acklin, McDowell, & Orndoff, 1992). As to the lack of correspondence with MMPI, claims that the Rorschach method's low convergent validity reflects its virtues rather than its limitations must be supported by empirical evidence. Lacking evidence for the consistent and stringent integration of the multitude of data given by the Rorschach method, there is more reason to lend weight to studies showing that two main underlying factors in the Rorschach method are general psychological health or distress, and affective control. The DEPI does not show diagnostic sensitivity. The SCZI shows similar lack of support, although it might be usefully in distinguishing psychotic and nonpsychotic disorders. Clinical validity is low. No evidence is shown for incremental validity.

Research on the Rorschach Oral Dependency Scale may serve as a useful guide for future CS research, as it includes varied research samples and varied measures for relevant validity criteria. The reliability and validity of the CS have been greatly overstated.

THE RORSCHACH IN PRACTICE

Combining information from different scales based on intuition is inferior to combining information using statistical formulas. Using cutoff scores on populations with base rates departing from around 50% is uncertain, in cases of 10-20% inadvisable. Clinical validity will decrease if findings from carefully selected, homogenous samples are applied to diverse clinical populations.

The main conclusion is that neither the Rorschach method nor the CS have a firm enough basis for widespread clinical use.

Dawes (1999): Two methods for studying the incremental validity of a Rorschach variable
Demonstrates two designs for investigating incremental validity with a reanalysis of previous EII data.

THE 1999 ASSESSMENT EXCHANGE

Archer (1999): Introduction to a special section: Perspectives on the Rorschach

Weiner (1999): What the Rorschach can do for you: Incremental validity in clinical applications

There is evidence that a complementarity exists between structured and unstructured tests like the Rorschach method and MMPI. In this complementarity we search for divergence between test results to provide different kinds of information about some characteristic. The Rorschach method is not a diagnostic test, and is not expected to do well in identifying diagnoses.

Acklin (1999): Behavioral science foundations of the Rorschach Test: Research and clinical applications

Rorschach critics persist in making global assessments of the Rorschach method without taking into account the complexity and multidimensional nature of the test's yield. Scale focused research and reviews are more illuminating than considerations for the global validity of the Rorschach.

STATISTICAL POWER, RELIABILITY, AND VALIDITY

The Rorschach method in general compared favorably with behavior science research overall, and the CS yielded research that was more powerful than non-CS research.

Applying Applied behavior analysis to the Rorschach method shows that a great majority of CS codes and variables yield acceptable levels of interobserver agreement and interrater reliability, using kappa and ICC.

Hiller et al. (1999) concluded that the Rorschach method yielded highly favorable findings compared to the MMPI.

THOUGHT DISORDER INDEX

The TDI is not influenced by age or race, showing discriminative power in psychotic and non-psychotic blacks and whites, and children and adolescents. The TDI distinguishes between borderline and schizophrenia, showing a continuum of severity for thought disorders, and marking quality differences in schizophrenic and manic conditions beyond the accuracy of clinical interviews. There are significant neuropsychological and biological markers and correlates for the TDI, and it is sensitive to the effects of medication. The TDI reflect a continuum of thought disorders in non-patients, siblings of schizophrenic patients, and schizophrenics. The studies presented adequate to excellent interrater reliability. As a conclusion the TDI has demonstrated a robust and clinically meaningful factor based on a priori and empirical factors, using objective validity criteria.

Garb (1999): Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings

The Parker et al. (Parker et al., 1988) meta-analysis is flawed. Garb's (1998) reanalysis shows that the MMPI is more valid than the Rorschach method. The results are however only suggestive. Hiller et al.'s (Hiller et al., 1999) meta-analysis is also flawed, as the criteria for inclusion affected the results. Meyer & Hadler (Meyer & Handler, 1997) failed to report that the Health-Sickness rating scale did better than the Rorschach method. The clinical relevance of Bornstein's (1999) findings are unclear, although he did find significant results. The West meta-analysis (1998) contains a terrible flaw by only reporting significant findings.

Positive results have never been obtained for the Rorschach in studies on clinical judgement and incremental validity. Some evidence exists for the EII, ROD, and RPRS. It is poor for DEPI, V, Col-Shd Bld, Ego/R, Afr, MOR, Sum-Sh, AG, COP, S, P, T, and D. incremental validity has not been studied for other variables.

Wood & Lilienfeld (1999): The Rorschach Inkblot Test: A case of overstatement?

NORMS

The normative sample for the CS is smaller than for other tests. The children sample is problematic because protocols with fewer than 14 responses were excluded, thereby reducing the number of protocols for certain age groups considerably. There is virtually no data for cultural minorities and non-US samples.

INTERRATER RELIABILITY

Claims are made that the CS has interrater reliability comparable to that of the WAIS. However, percentage agreement, as reported by the CS might be seriously inflated. Recalculations using kappa shows considerably lower levels, ranging from 1.0 to below 0.20. Several clinically important variables have unacceptably low levels, e.g. SCZI and Adjusted D.

TEST-RETEST RELIABILITY

Exner seems to be reporting the same data for the same set of 40 variables over and over again. This leaves 85 variables unreported. Test-retest is also an unsuitable method for testing temporal stability.

FACTOR STRUCTURE

The biggest factor in the Rorschach method is R, accounting for approximately 50% of the variance. Several important variables show correlations with R ($r=.25$ to $.60$).

UNPUBLISHED STUDIES

63% of the studies cited in Exner (1993) are unpublished works from the Rorschach Workshops. The studies are not available upon request. This seriously compromises the empirical foundation of the system.

THE 2000 JOURNAL OF CLINICAL PSYCHOLOGY EXCHANGE

Wood, Lilienfeld, Garb & Nezworski (2000b): The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947)

Two major sources of methodological flaws, namely failure to blind test administrators and scorers, and criterion contamination of diagnoses, are still common in Rorschach studies, despite the repeated warnings by Rorschach scholars.

DIAGNOSES

Schizophrenics often shows signs of deviant verbalization and bad form. This is only to be expected, as they are striking and serious symptoms of schizophrenia. No incremental validity is to gained from the Rorschach in this area.

No conclusive evidence can be found regarding PTSD, based on several studies. Findings have not been replicated. Patients with panic disorders might have a lower SumC than other patients, but no conclusive evidence is given. There is therefore no established relationship between Rorschach scores and anxiety. A probable relationship might exist between the total number of movement responses and Dissociative Identity Disorder.

Only one study reports relationships between Rorschach Oral Dependency Scale and Dependent Personality Disorder. Virtually all other studies on ROD has been conducted by one researcher (Masling) and his former student (Bornstein). Replication is therefore needed. In setting diagnosis there are problems of contaminations or chart based evaluations in all studies regarding Narcissistic Personality Disorder.

Only ten studies on Borderline Personality Disorder is discussed, the rest eliminated for reasons of methodology. Also in the ten remaining, methodological elements are questionable: Administrators have scored the Rorschachs, t-test or ANOVA is used for skewed data, alpha inflation has not been warded against, negative findings have not been reported, unclear selection procedure have been applied, and "fishing expedition" have been organized in the pool of data. There is virtually no overlap or replication of findings.

No replicated findings have confirmed relationships between Rorschach scores and Anti-Social Personality Disorder, Conduct Disorder, or psychopathy. In positive findings, the probability of Type I error is substantial, as a large number of variables were investigated and only a few significant results were found.

CONCLUSIONS

Contrary to common claims, only a few Rorschach scores have a well-demonstrated relation to psychiatric disorders, except deviant verbalization and bad form. This is supported by meta-analyses, showing that the Rorschach method does not contribute to incremental validity in making diagnostic decisions, compared to MMPI. Rorschach scales lack replication, and in many instances methodological soundness. Published claims of the validity of the Rorschach method has by unjustifiably optimistic and even overstated.

RECOMMENDATIONS

The Rorschach method should not be used when formulating diagnoses. It should not be used in forensic contexts. Training programs in clinical, counseling, and school psychology may want to consider eliminating the test from their assessment curricula. Rorschach researcher should pay more attention to methodological issues, as exemplified by the following six recurring problems: a) comparing diagnostic groups to normative data, b) basing criterion diagnoses on procedures other than clinical or structured interviews, c) failing to blind diagnosticians to both direct and indirect influence from Rorschach scores, d) failing to blind administrators and scorers to research hypotheses and diagnostic groups, e) performing large numbers of statistical tests without adequate adjustments for alpha, and f) using parametric rather than non-parametric statistics for skewed data and small samples.

PREDICTIONS

Negative consequences are likely to arise if Rorschach proponents do not begin to pay closer attention to methodological issues, and future research will probably not uncover important new relationships between the Rorschach method and psychiatric disorders.

Garfield (2000): "The Rorschach test in clinical diagnosis": A brief commentary

Garfield refutes the statement that 71.9% is "near-perfect", and cites newer research on the diagnostic hit rate with the MMPI-2 of 70%. Further inaccuracies are identified regarding the statement that Rorschach results contaminated the diagnostic decisions. Diagnostic criteria are not well suited for clinical research, also because they are often inaccurate. It is important to consider the totality of the protocol, and not focus too intensely on scores.

Lerner (2000): A nonreviewer's comment: On the Rorschach and baseball

A nonreviewer's comment: "If someone chooses not to play, if his or her interests lies elsewhere that is fair enough. However, to not have played, to not have immersed oneself in a protocol, carefully sifted through the rich imagery, and then struggled to translate the data into psychological meaning so as to understand and help a troubled human being; to not have done this, but instead, to sit outside the ballpark in front of a television set and cavalierly pronounce the process an invalid waste of time, unfit for patients and student consumption, that is not ok. It is disrespectful and unacceptable."

Weiner (2000): Using the Rorschach properly in practice and research

"Associations between Rorschach indices and DSM categories have little bearing on the utility of the RIM for achieving its intended purposes." Many DSM diagnoses are unreliable, and the definitions change in each revision. Correlations between the Rorschach and DSM are not meaningful and of little interest. Rorschach data is however meaningfully related to clinically important behavior, such as dependency (ROD) and thought disorders (TDI), regardless of diagnostic status. Wood et al. present only parts of the Hiller et al. meta-analysis data, which suits their argument, overlooking the basic conclusion that the Rorschach method and the MMPI are equally well suited in appropriate contexts.

Wood, Lilienfeld, Garb & Nezworski (2000a): Limitations of the Rorschach as a diagnostic tool: A reply to Garfield (2000), Lerner (2000), and Weiner (2000)

The influence of administrator and experimenter effects on the Rorschach has been well documented, and it is important to keep administrators and scorers blind to diagnoses and hypotheses. The 10 diagnoses were mentioned because there was data available. 71.9% is very high, also considering other more recent correlations. Clinical training and experience has not been related to the validity of Rorschach findings. Many Rorschach proponents have endorsed the method for diagnostic purposes. Inanimate movement, pure shading and the D

score are not as clearly related to stress as previously indicated. The TDI does probably not add any significant information over an interview and an MMPI.

THE 2001 JOURNAL OF PERSONALITY ASSESSMENT EXCHANGE FOLLOWING WOOD ET AL. (1999B) AND MEYER (2000)

Wood, Nezworski, Stejskal, Garven & West (1999b) with erratum (1999a):
Methodological issues in evaluating Rorschach validity: A comment on Burns and
Viglione (1996), Weiner (1996), and Ganellen (1996)

Many critics and all proponents agree that some Rorschach scores have well established convergent validity. Many critics and proponents agree that the reliability and validity of "The Rorschach" is not an issue, rather its specific scores. More narrow and focus meta-analyses are needed. The debate is likely to achieve greater clarity and focus when more specific evidence is offered regarding the validity of specific scores.

HEV VALIDITY

In Burns & Viglione (1996), the EMRF is inadequately used as a measure of interpersonal relatedness. Further more, EMRF, BORI and SORI scores are not presented separately, making interpretations difficult. Two incompatible methods were used to calculate HEV in the same study, one Z score and one weighted method. Extreme groups of the highest and lowest values were defined and entered into a logistic regression analysis, rather than using the continuous data in a linear regression analysis. This design offers greater statistical power to detect a true effect, but is only recommended in a few, very specific instances. In the present design, the consequences for clinical decision making are limited by the exclusion of the middle group. Dichotomized continuous data often yield inflated estimates. Hierarchical and backward stepwise regression procedures are mixed in manner not suited for the data at hand. The inclusion of HEV into the equation did not significantly improve predictive power, contrary to the conclusions in the text.

PTSD

Two independent studies confirmed the relationship between PTSD, and D score and MOR, but both compare their data to reference material, not control groups. The combat veterans were different from the reference group in many demographic variables, and differences in interscorer reliability could explain the relationship.

DEPRESSION

Ganellen (1996) argued that the MMPI, MCMI-II and the Rorschach have comparable ability to identify depression. However, all independent replication studies of the DEPI have failed to find any such relationship. Ganellen relied on a "between-group" comparison, which is inconclusive because of the inability to control for extraneous variables.

RECOMMENDATIONS

Future studies should avoid using CS normative data for comparison. Backward stepwise regression should not be used to test incremental validity, rather hierarchical regression. Extreme group designs are not recommended for studies with relevance for clinical decision making.

Meyer (2000): On the science of Rorschach research

even though Wood et al. have many relevant points, they probably also knew some of them were inaccurate. Regarding the faulting of Weiner's logic, a similar type of argument is frequent in other tests, and studies without control groups provide more evidence than no studies at all. Ganellen devoted several pages to a discussion of the limitations of his material, also the Rorschach data.

ERRONEOUS CITATIONS

Wood et al. erroneously cited Meyer et al. (1998), but printed an erratum for this. Also McCann (1998) was cited as supporting the use of DEPI in forensic settings, despite his clear warnings to do so. Meyer (1993) was erroneously cited in support of the incremental validity of the SCZI, despite the fact that no such topics were raised. Archer & Gordon (1998) was cited for similar support, even though they only briefly reported an additional analysis in their Discussion section. On the other hand, other features indicated that the SCZI is actually a better univariate predictor than MMPI Scale 8.

Additional analyses of the Meyer (1993) data reveals that the SCZI adds incremental validity to the MMPI's Scale 8 and Bizarre Mentation. The same results were obtained from a similar analysis of Scale 2 and DEPI, although with more modest gains for the DEPI. The procedure to test incremental validity with a regression analysis would in fact impose demands for statistical validity for the SCZI, but not the MMPI. Cohen and Cohen (1983) do not see hierarchical and stepwise procedures as incompatible. Forced entry of variables is highly problematic because of suppressor effects and the inclusion of sample variance in the models. Archer and Krishnamurty (1997) found that vista and affective ratio did add incremental validity over the MMPI, though not the whole DEPI.

BURNS AND VIGLIONE

The alpha levels chosen in the equation building were in line with recommendations ($p = .15-.20$). The HEV did show predictive power related to interpersonal competence.

A regular experimental and control group design often effectively creates an extreme group design, because criteria selections will disregard those with "some" of the criteria. Disregard for base rates can also produce similar effects. When a number with a rare condition is compared to the same number without, the variance in the rare group will be inflated. It seems that Wood et al. put up different standards for the Rorschach than for the MMPI. Burns and Viglione made specific limitations clear regarding the generalizability of their results.

Composite measures are most valuable when they give unique contributions, that is they are not highly correlated. A pilot study showed that the composite measure was in fact relevant and useful. Two studies overlooked by Wood et al. support the validity of EMRF.

“Incompatible HEV formulas did not exist until Wood et al. created a faulty formula for their article.” They entered faulty values for means and standard deviations, even though they were told in a review process in a different journal that the computations were erroneous. Meyer himself also supplied a simulation study for the faulty formula to Wood resulting in $+0.99$ correlations, but he still disregarded the errors. The same data was discussed by Meyer and Wood on the Rorschach Discussion List. In the face of this data, Wood still claimed there was “considerable doubt”.

Wood, Nezworski, Stejskal & Garven (2001): Advancing scientific discourse in the controversy surrounding the Comprehensive System for the Rorschach: A rejoinder to Meyer (2000).

“It is untrue that we knowingly published a faulty formula for the HEV.” Burns and Viglione stated explicitly that the “erroneous” data was used, and they mixed different versions of the HEV developed for experimental purposes. Upon request for reanalysis, the Burns and Viglione data had been lost. The simulations were unconvincing, because they were not based on real life data with similar skewness, and not on the actual data itself. Calculations show that the two versions can produce clinically important differences, and in some cases give different directions for the results. Still, the significance of the difference was overstated.

Ganellen (2001): Weighting the evidence the Rorschach's validity: A response to Wood et al. (1999)

Meyer (1997a) have refuted Wood et al.'s claims of lacking interscorer reliability for the CS. Claims that the data is not relevant for clinical practice has never been made for other tests. Archer and colleagues have mainly focused on the relationship between the Rorschach and the MMPI, not incremental validity.

Several of the studies with negative findings for the DEPI was excluded because they used the first version. The two versions must be regarded as different indexes, and not interchangeable. Archer and Krishnamurty (1997) contained an error which led Wood et al. to a faulty conclusion. The correct data indicates that DEPI and MMPI Scale 2 are equally good at predicting depression. The sample characteristics are also somewhat deviant, leaving the conclusions uncertain. All shortcomings pointed out by Wood et al. in Ganellen were addressed in the article.

Gacono, Loving & Bodholdt (2001): The Rorschach and psychopathy: Toward a more accurate understanding of the research findings.

It is important to differentiate between psychopathy, defined in a psychoanalytical setting, and Antisocial Personality Disorder (ASPD) in the DSM. (Goes on to explain differences between

the two.) Diagnoses are poor constructs, not intended to be linked with Rorschach results(!). Many of the studies cited by Wood et al. (2000b) are in fact inconclusive, because of poor methodology and lacking regard for the ASPD/psychopathy distinction.

Bornstein (2001): Clinical utility of the Rorschach Inkblot method: Reframing the debate

In the backlash period of the Rorschach (after Wood et al's 1999 article), proponents and critics have often reached diametrically opposite conclusions based on the same available data. The debate cannot be resolved because it is not framed properly. Five principles are offered to increase the quality of Rorschach research.

Recognize what the test can and cannot do. Current approaches focus on implicit motives, cognitive/perceptual style, and coping style. These are more or less inaccessible to verbal report, and the Rorschach should therefore be better suited than questionnaires.

Choose appropriate outcome criteria. The Rorschach and self-report measures cannot be expected to give the same correlations, as different things are being measured. Motivational and emotional experience that are filtered through analytic thought and various concepts of self and others are at the heart of the Rorschach method.

Use multi method, multi criterion matrices to place results in an appropriate context. Both the Rorschach and a self-report test should be compared to at least two other outcome measures.

Consider limitations in the outcome criteria themselves. Diagnoses are poor criteria.

Use experimental – not just correlational – data to contrast the results obtained with projective and self-report tests. Manipulations are better than correlations.

In some cases, the absence of a relationship actually supports the validity of a test. This is a strength for many of the lacking relationships with self-report measures, not a weakness.

Wood, Lilienfeld, Nezworski & Garb (2001): Coming to grips with negative evidence for the Comprehensive System for the Rorschach: A comment on Gacono, Loving and Bodholt; Ganellen; and Bornstein.

The scoring reliability of the CS is at “a level which is questionable for clinical or forensic work”. Rorschach norms are problematic, and recent evidence has shown deviance from the Exner (1993) norms. Not so long ago, it was predicted that Rorschach variables should correlate with self-report measures. This has been retracted. Similarly, a relationship with diagnoses was expected, which has also been retracted. Again, the CS was thought to be superior to other approaches, which has not proven right. There is still ample evidence to refute the relationship between the Rorschach and psychopathy. The revised DEPI has not fared better than the previous (citing articles), also when Archer and Krishnamurthy's revised data is considered. Individual scores can have low reliability even when the category is high, despite Meyer's contentions. Despite Bornstein's acknowledgment that the Rorschach is not

related to diagnoses, many proponents have argued to the contrary. Some confusion is evident in the arguments on convergent and discriminative validity, e.g., Bornstein.

THE 2001 CLINICAL PSYCHOLOGY EXCHANGE

Wood, Nezowski, Garb & Lilienfeld (2001a): The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach

Shaffer, Erdberg and Haroian (1999) presented normative data that deviated from the CS normative samples in many important variables: High SCZI (>4) in one of six, 29% reflection responses, X+% was .51 (3 SD's lower than CS .79), X-% was .21 (two SD's higher than CS .07). Other deviating variables were ambivalent style, Afr, FC:CF+C, P, Y, T=0, SumC, MOR, Wsum6, Lambda>.99, and PureH<2. Nearly all were in the direction of pathology.

In a search for other normative samples, 32 studies were identified which contained nonpatient adult samples. Data was compared on variables showing discrepancies in the Shaffer et al. data. All variable means and standard deviations differed significantly from the CS data, while only two variable means and three standard deviations differed from the Shaffer et al. data. In effect size terms, the differences were substantial.

POSSIBLE EXPLANATIONS

Administration and scoring was poor in the 32 studies. No specific evidence support this. On the contrary, many prominent researchers were involved, an interscorer reliability data did not give such indications.

The samples are not representative for normal American adults. All samples were distinctive, and did not pretend to be representative. On the other hand, they all showed a uniform deviance, despite their diversity.

The dissertations included were a biased sample. They were not randomly selected for the review. Statistical tests showed that 10 of the 14 variables were not different in the dissertations and the published studies, and the remaining four were still very different from the CS.

The normative data was accurate in 1970, but the population has changed. Exner stated that the data has accumulated over a period of more than 20 years. Data presented in different version of the CS shows little change. On the contrary, some shift can be seen leading away from the discrepancies. Data from the 1970's does not indicate any changes through the years.

The CS norms are not representative. This seems like the most plausible explanation. Poor administration and scoring could be a cause, but no data is given to make such an evaluation possible. Another explanation could be that the sample was super-normal, and excluded all subjects with any type of psychopathology. Indications are given to the contrary, however, and this should be ruled out.

IMPLICATIONS

The validity of the CS is put into question. Gacono and Meloy's (1994; 1992) findings may not be valid, the same for Viglione's (1999) claims regarding Y as an indicator of stress. Hiller et al.'s (1999) meta-analysis included four studies which compared their data to the CS norms, which may have lead to inflated estimates.

When used in clinical settings, people might look more pathological than they are. Discrepancies between normal MMPI's and pathological Rorschach's may be caused by this. Other tests with more well established norms should be used.

Widiger (2001): The best and the worst of us?

The Wood et al. data can actually be used as norms. Garb and other critics may be as excessive in their dismissal fo the Rorschach method as Weiner and other proponents are in their embracement. Wood et al. are not making suggestions for how Exner can further advance the Rorschach; Wood et al are indicating that the interests, perspectives and efforts of Exner and his colleagues are no longer wanted. Perhaps these attacks will only result in the removal of it weaker components. The method may be more suited as a therapeutic technique. Perhaps the Rorschach is now in the best of times rather than the worst of times, given the assumption that scientific progress proceeds with the survival of severe criticism.

Hunsley & Di-Giulio (2001): Norms, norming, and clinical assessment

Wood et al. only examined a limited set of variables. Still, Shafer et al. also found deviations in other variables, indicating that they are not limited to the examined variables. The CS sample was not gathered as a probability sample designed to be representative, but rather a selection was made afterwards to match the demographic makeup of the 1980 U.S. Census. Weiner is retracting from previous claims on behalf of the CS that the sample is representative. Compared to the normative samples for MMPI-II and WAIS-III, the CS normative sample is substandard. Before better norms are established, the existing ones should be used with caution in clinical practice. Rorschach users tend to overestimate the value of Rorschach data when it is in conflict with other tests. Given the deficiencies, a discontinuation of the use of the Rorschach in clinical and forensic settings must be considered.

Aronow's (2001): CS norms, psychometrics, and possibilities for the Rorschach technique

Wood et al.'s general point seems justified. The effort to transform the clinically rich Rorschach into a super-psychometric tool was not an appropriate goal from the start. The Rorschach is basically a fine tool for subjects to reveal aspects of self-concept, significant emotional conflicts, and view of significant others. Scoring must be secondary to projective analysis. Hermann Rorschach's own conceptions about the method has been deified.

Exner (2001): A comment on "The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach"

Wood et al. did not report on the data gathering procedures of his studies. The studies also show considerable variation among themselves, which may not warrant the calculation of a single mean. The deviations for Afr and T is linked to the high Lambda, which again is connected to the highly skewed R in the direction of few responses. The deviations are not necessarily connected to pathology, but are rather elements in a total pattern of variables. Normative data is not used directly to determine pathology.

Meyer (2001a): Evidence to correct misperceptions about Rorschach norms

Both the WAIS and the MMPI have changed their base norms in the direction of higher IQ and more symptoms compared to previous versions. These effects are of the same magnitude as the differences reported by Wood et al. Research shows that pathology has increased over the years, and that the life time incidence of pathology is 48% for limited set of disorders. The recruitment procedures for the CS sample indicated that a super-normal sample was found. Many of Wood et al.'s studies did not exclude pathology, and were often recruited from groups where pathology is more prevalent. College students are more disturbed than other samples, even psychiatric populations. Clinical samples are often only minimally different from census-matched samples. Changes in the form level tables have lead to a decrease in good form level responses, and lower X+% must be expected. Many determinants were scored more liberally in previous versions of the CS. Significant deviancies in R were apparent in many of the studies, which also had an effect on many of the other variables. Several errors, omissions and inaccuracies are evident in the Wood et al. data. Analysis of international CS data (Erdberg, 1999) indicates that, on average, the deviancies are about 0.1 of an SD for 69 variables.

Wood, Nezworski, Garb & Lilienfeld (2001b): Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations

Rorschach administration and scoring needs to be secured, as many others also have pointed out. This, however, casts doubt on the current use of the Rorschach, and the CS. No recommendations are made to abandon the Rorschach all together, only temporarily halt the use in clinical and forensic settings until further clarification is achieved on the validity of Rorschach variables. The method may still be quite useful as a therapeutic tool. Exner and his efforts are welcome. Aronow has a realistic view on the Rorschach.

Exner has recently reported that 221 of the 700 protocols were duplicates. The consequences are unknown. Clinicians must rely on normative data to identify deviant scores. Overpathologation will be the result, even when a global approach is used. Mittman showed a high degree of false-positives.

The data on the 14 variables in the international data is as deviant as the initial study, and actually slightly more pathological. The data therefore confirms, not refutes, the main

argument. 15-20% of the CS sample had been in contact with a mental health professional at one time or another.

THE 2001 PSYCHOLOGICAL ASSESSMENT SPECIAL SECTION, PART 2

Meyer (2001b): Introduction to the final Special Section in the Special Series on the utility of the Rorschach for clinical assessment

Weiner (2001): Advancing the science of psychological assessment: The Rorschach Inkblot Method as an exemplar

The standardization provided by the CS ensures comparable data from any circumstance. Ill-informed interpretations reflect the clinicians incompetence, and say nothing about the soundness and value of the Rorschach method. The data is not bound by theory, and can be interpreted in different contexts. The stimulus material is cultural free, but interpretations must rely on ethnic differences. The Rorschach lends itself both to nomothetic and ideographic interpretations.

Hunsley and Bailey's (1999) review consisted of 128, of which only 28 appear to be original research reports. Viglione's (1999) review consisted of 195 original research reports. He demonstrated that interscorer agreement was adequate in about every report. Hunsley and Bailey did not review specific evidence that it is poor. Further that, they question filed reliability as if it is a property of the method, when it is in fact reflecting the clinicians' competence. Temporal stability data also indicate that interscorer reliability is good.

The normative sample may represent a rather well-adjusted segment of the U.S. population. Erdberg and Shaffer's (1999) data show remarkably similar ranges for many variables. The data also provide evidence of construct validity by their representativeness and their changes with age groups.

Each score has its own validity coefficient. Rorschach data must be related to personality process data, and preferably to observed rather than inferred data. Self-report instruments are inadequate as target criteria. The choice of target variables must be informed, focusing on relevant criteria, not any criteria.

Considering Stricker and Gold's (1999), Viglione's (1999), and Hiller et al.'s (1999) clear evidence to support the validity of the Rorschach, Hunsley and Bailey's assertions are without substance. Their arguments of faulty meta-analyses are now rendered moot. The strong evidence is reported despite the possible confounding problems regarding the number of responses. The relationship between the Rorschach and the MMPI cannot be used to infer the validity of the Rorschach.

Garb (1984) concluded that the Rorschach does not yield incremental validity, but was not considering the CS, and only data from 1954 to 1982. Moreover, faulty conclusions about the

use of the Rorschach and MMPI were presented by Hunsley and Bailey, which in fact showed that the Rorschach method outperformed the MMPI. Dawes (1999) showed, contrary to his previous dismissal of the Rorschach method as “shoddy”, that the EII was better than the MMPI at distinguishing psychosis from neurosis. Incremental validity studies too often base their comparison on a MMPI-Rorschach combination.

Valid concerns include the need to promote and document the coding reliability of clinicians and researchers, resolving the impact of response frequency, and more well-designed studies for many variables that have not so far been adequately examined.

Garb, Wood, Nezworski & Grove (2001): Toward a resolution of the Rorschach controversy

Stricker and Gold (1999) fail to report studies that does not support the Rorschach. Other proponents have identified variables that are not well founded (DEPI, SCZI). Others have failed in independent replications (S-Con). RPRS results, although favorable, need to be replicated. Lacking performance on part of the Rorschach is often rationalized in post-hoc arguments. Negative evidence cannot be given as support, such as lack of correlations.

Viglione's (1999) claim for impressive temporal consistency cannot be supported by data (exemplified by Adair & Wagner, 1992; Perry, McDougall, & Viglione, 1995; Schwartz, Mebane, & Malony, 1990). Only 40% of the CS variables are reported. Acklin et al. (2000) showed that half fo the CS variables fell short of 85% interscorer agreement. Cross-cultural studies have identified several shortcomings of the Rorschach, often with deviancies in the direction of pathology for ethnic groups. Viglione consistently ignored negative evidence on clinical judgement. The Rorschach might be of value when identifying psychosis, but other instruments are equally suited and less time-consuming. Other evidence reported by Viglione overstated the original authors' conclusions.

Some evidence is available for the incremental validity of Rorschach variables, but only for a limited number of them.

The Hiller et al. (1999) meta-analysis had several flaws. Interrater reliability was poor (.35), probably because of lacking coding guidelines. Judges were not blind to the results when selecting studies. Effect size aggregations relied on subjective estimations of intervariable correlations.

RECOMMENDATIONS

Four problems are frequent, and must be avoided in future studies. Control groups are needed for each study, and the normative data should not be used as a reference. Rorschach results must not be a part of the criterion ratings, to avoid criterion contamination. Nonsignificant results need to be reported, and alpha-levels must be adjusted when conducting large numbers of significance tests. More replication studies are needed. Incremental validity must be more thoroughly examined. Meta-analyses should include unpublished studies, and include all predictors in the study. The data from the Rorschach Workshops must be available.

Rosenthal, Hiller, Bornstein & Berry (2001): Meta-analytic methods, the Rorschach, and the MMPI

Several indices of central tendency is available, and should be reported. The MMPI showed variability in different indices, while the Rorschach showed consistency. Corrections for the problems identified by Garb et al. did not lead to any changes, trimming lead to small, but similar decreases in effect sizes. MMPI studies were skewed in their distribution of N.

Viglione & Hilsenroth (2001): The Rorschach: Facts, fictions, and the future

Interscorer reliability is good. The available data for temporal consistency is good, the studies mentioned are methodologically problematic. Compared to data from other tests, the Rorschach shows impressive temporal stability. The normative sample is rather young, 73% is under 36, and they are relatively well-educated. Normative data on R is variable, Lambda is "quite close" to the CS data. Form quality interpretations need to be modified, and definitions should be more restricted. Differences found in various ethnic samples may in large be explained by demographic differences. A new normative sample should be collected.

Several studies demonstrate incremental validity of the Rorschach, often outperforming MMPI. Dawes (1999) support the same conclusion, even with a biased reanalysis of the EII. The number of responses and EQUAL was entered into the regression analysis before the EII, even though R is not a valid predictor, and FQ- and R is part of the EII. The result is reduced predictive power. The DxSum developed for the reanalysis is not a valid criterion.

The TDI has proven valid in the assessment of schizophrenia. S-Con should still be considered as potentially positive.

The CS, with its combination of actuarial and clinical methods, is consistent with thinking in the field of judgement and decision making. Hunsley and Bailey is oversimplifying the routines of clinical decision making by ignoring the interaction with other sources of data. More ecologically valid studies are needed, where the actuarial-clinical approach is tested in realistic circumstances. Students, internship directors and practitioners all regard the Rorschach as an essential method, together with the MMPI and WAIS. The Rorschach is widely accepted in forensic use, and the scientific status is questioned in a minority of cases. Clinicians continue to use the method, despite time constraints and limited third party payments, because of its clinical utility. The Rorschach method is especially useful in contexts where the subjects are unwilling or unable to self-report problems, and where the clinicians are trying to predict real life behavior.

The critics are not applying the same weight of evidence on other tests. Viglione (1999) cited published research in response to the criticism of the reliance on unpublished reports. The critics did not mention positive findings and only based their critic on a small selection of studies. Viglione's review contained a wide selection, and also included negative findings. Double standards are used when the critics argue that clinicians used as criterion raters is a fallible approach, while at the same time use studies with negative findings which use such an approach to argue against the Rorschach. "These recent criticisms are undermined by methodological double standards, confirmatory bias, incomplete coverage of the literature,

failure to integrate positive contributions over the last five years that have specifically addressed earlier criticism from these authors, and a lack of any original data to support their positions.”

Hunsley & Bailey (2001): Whither the Rorschach? An analysis of evidence

Several theories have been put forward to account for how Rorschach responses provide data on the inner life and observable behaviors for a person, but none are widely accepted. This lack of consensus makes it easy to dismiss criticisms as irrelevant, or as disrespectful and uninformed. Proponents are frequently mixing the Rorschach method and the CS, but they cannot be equated. The multifaceted nature of the Rorschach can be seen as a potential for confusion, bias and inattention to data, while the proponents see it as a potential for richness and complexity. Future research needs to conduct theoretically informed studies. The stance of dismissing the relationship between Rorschach and self-report is inconsistent with practice, obtained effect sizes, and which studies is used by proponents to argue in favor of the Rorschach. The lack of relationship between the Rorschach and diagnoses might in part be explained by an inappropriate reliance on clinical criteria. Greater consensus is needed on what constitutes the best criteria. Observable data might have some merit. Meta-analyses should be directed toward more specific scales in combination with other measures, and importantly, include dissertations.

The research literature is nonprogrammatic and noncumulative. Promising results are often cited, but not pursued. The five studies cited by Viglione in favor of the clinical utility of the Rorschach give only limited evidence. There is presently no evidence to the clinical utility, which is not the same as evidence for validity. Viglione's argument for S-Con as a useful predictor is not considering the practical consequences of false positives. Even though the clinician may be gratified by Rorschach results, this need not help the patient. A study could be designed where half of the therapists were given Rorschach results, the other half not, then treatment results could be compared. The actual consequences of the Rorschach results for each patient needs further examination.

Weiner replies to criticism with rhetoric, but no evidence. There is a need to establish the scoring accuracy in clinical settings. Scoring accuracy is determined by the complexity of the scoring rules, and the boundaries between “ideal” and “field” reliability cannot be clearly drawn. The CS norms are problematic, as shown by Shaffer, Erdberg and Haroian (1999). The deviancies may explain why pathology is found in the Rorschach and not in other instruments. Although there is some evidence for the global validity of the Rorschach, little evidence is available for specific scales for specific uses. Rorschach data add little or no information compared to other tests.

Proponents and critics agree that the Rorschach must be treated as a test, and therefore meet the psychometric standards for tests. Complete reviews for specific scales are needed to establish their validity, especially their incremental validity. The scoring demands require intensive training and evidence of the clinicians ability to use the method appropriately. More programmatic research is needed to replicate results. Methodological issues must be strictly

considered. There is no scientific basis to support the continued use of the Rorschach in clinical, legal, and occupational settings.

Meyer & Archer (2001): The hard science of Rorschach research: What do we know and where do we go?

Rorschach indices lack face validity, which may be an initial disadvantage. The Rorschach and the MMPI are not correlated, despite scales with similar names. Data shows that distinct methods provide largely unique information, while their intercorrelations are often only in the range of .15 to .30. There is still too little information on the specific virtues and limitations to different tests.

META-ANALYTIC FINDINGS

Parker et al.'s (1988) data was reanalyzed by including any test scale in each study, focused effect sizes were calculated, limiting to effects that addressed the validity of a scale for measuring a construct, distinguishing heteromethod and monomethod coefficients, and correcting any coding mistake encountered. The impact of the changes were negligible compared to the original set. The Rorschach range from .27 to .30, the MMPI from .23 to .28, and the WAIS from .32 to .36. Highly similar ranges were found for the 10% trimmed sample, except the MMPI weighted estimates, which were unstable.

All available meta-analyses taken together (Atkinson, 1986; Garb et al., 1998; Parker et al., 1988) show that the Rorschach hovers around a consistent level of .30. All tests show similar yields when method confounds are corrected.

Global analyses provide diffuse information. Comparisons of global estimates across tests is misleading, since they are not related to the same set of criterion variables. Global results do however provide evidence of validity across a variety of moderators, such as publication year and type, sample sizes, aggregation procedures, etc.

Specific analyses show that the Rorschach can validly predict a range of criterion variables. Meta-analyses for the Rorschach, MMPI and WAIS show that effect sizes vary as function of the predictor-criterion relationship under consideration. The first conclusion is that there is no reason to for the Rorschach to be singled out for particular criticism or specific praise. The second is that validity is conditional, depending on the selected criterion and predictor.

ISSUES THAT MUST BE ADDRESSED

The locus of effectiveness must be clarified. The Rorschach may be best suited for implicit and underlying personality characteristics. On the other hand, responses are censored, people differ in how much they invest in the task, and many variables are not directly observable. As an example, determinants require more inferential steps to be linked to behavior than cognitive special scores, which are more directly linked to behavior. For depression, the diagnosis is related to behavior that does not show on the Rorschach, such as weight loss, early morning awakening, etc. though disturbances, on the other hand, appear directly in a Rorschach as verbal behavior. Researchers must stop thinking of test scales in terms of the

constructs they are supposed to measure without considering the influence of the measuring method itself.

The present CS normative data is relatively healthy and well functioning. New data is being gathered, and should be explored for influences of gender, age, etc. Smaller cells of protocols would be sufficient, also to be able to rescore them when changes are made in scoring definitions.

The consistency of test administration and inquiry has received little attention. This was demonstrated in the gathering of normative data in Italy, where the initial round of nonstandardized administrations proved deviant from the following standardized administration.

Researchers should make it a priority to collect psychometrically sound retest data sets. Many variables with substantial interpretative emphasis have received little or no attention regarding their validity. Organized and coordinated efforts are needed to accumulate research findings. Test taking styles is an important confounding effect. Unpublished studies (68) are also cited in the WAIS manual (75 in Exner, 1993). All studies must be presented with enough detail to enable evaluation of the method. Studies that are pillars in variable understanding should be peer-reviewed. There is presently no evidence of ethnic influence in the Rorschach, but more data is needed.

Claims cannot be made that a test must prove better than similar tests to justify their existence, only that they are valid. Specifically, the Rorschach alone cannot be subjected to such demands. For applied clinical practice, construct validity is more important than criterion validity, because the goal of assessment is descriptive. Incremental validity ignores the clinical value of convergent conclusions. Still, the Rorschach is time-intensive, and must therefore prove its usefulness despite this handicap. Furthermore, there is a clear link between incremental validity and the locus of effectiveness.

Clinical utility studies need to let the clinicians freely choose the appropriate test. Appropriate criteria must be selected, which mainly address whether the clinicians feel they gain more in understanding and insight.

RECOMMENDATIONS

Sophisticated theorizing. Newer statistical procedure, like item-response theory and taxometric analyses. Systematic analyses of card pull, designed to enable weighting of cards. Examine task investment. Less emphasis on diagnoses as criteria, more on behavior. Examine understudied variables, and define locus of effectiveness. Develop tools that can adequately capture and evaluate the Rorschach as an ideographic method. Collaborative networks to generate systematic, multisite, cross-validated findings.

META-ANALYSES

Parker (1983) analyzed 39 articles published in *Journal of Personality Assessment* in the period 1971 to 1980, and found reliability levels of .83 and validity levels of .45-.50. Commented on later by Garb, Florio and Grove (1998).

Atkinson (1986) analyzed 170 Rorschach and MMPI studies from 1960 to 1980. Effect sizes were in the range of .34 to .58, and the two methods were similar. Conceptually founded studies were more powerful than undirected studies. Commented on later by Garb, Florio and Grove (1998) and Hiller et al. (1999).

Atkinson, Quarrington, Alp and Cyr (1986) analyzed 120 Rorschach articles drawn from every fifth year between 1930 and 1980. Conceptually based studies showed higher effect sizes than other empirical studies. Commented on later by Garb, Florio and Grove (1998) and Hiller et al. (1999).

Parker, Hanson and Hunsley (1988) analyzed 411 studies of the Rorschach, the MMPI and the WAIS from 1970 to 1981. The three methods showed similar reliability (.74-.87), while the WAIS exceeded the others in validity levels (.62 for WAIS, while .46 and .41 for the MMPI and Rorschach, respectively, for convergent validity). The Rorschach was lower than the other instruments on unknown validity. Reanalyzed by Garb, Florio and Grove (1998) and Meyer and Archer (2001), and commented on by Hiller et al. (1999).

Meyer (1997a) analyzed interscorer reliability in 16 articles published in *Journal of Personality Assessment* from 1992 to 1995 as a reply to Wood et al.'s (1996a) criticism. Fixed chance agreement estimates were entered into kappa calculations, combined with observed percentage agreement in the studies at hand. The traditional Cohen's kappa mean was .86 (.72-.96), the fixed chance agreement kappa was .91 (.87-.96), while percentage agreement was .92 (.74-1.0). Percentage agreement did not seem problematic when agreement was substantial.

Meyer and Handler (1997, with correction in Meyer & Handler, 2000) analyzed 18 studies using the Rorschach Prognostic Rating Scale (RPRS). They found an uncorrected correlation between RPRS and outcome of .44, and a corrected correlation of .56. They concluded that they were "not aware of any other personality scale that uniformly demonstrates such a high predictive validity."

Garb, Florio and Grove (1998) criticized several of the previous analyses. Parker (1983) did not separate reliability and validity in the analyses, Atkinson, Carrington, Alp and Cyr (1986) did not analyze magnitude of effects, and Atkinson (1986) did not aggregate estimates across studies. Parker, Hunsley and Hanson (1988) did not pool the results according to the type of statistics, some studies shared assessment instrument variance with the construct measures which resulted in inflated effect sizes, and small effect sizes were apparently interpreted as a test being valid. A reanalysis of Parker, Hanson and Hunsley's data indicated that MMPI was more valid than the Rorschach (.48 versus .29 for confirmatory studies), contrary to Parker et

al.'s conclusions. Commented on by Parker, Hanson and Hunsley (1999), and in turn by Garb, Florio and Grove (1999).

West (1998) reported high effect sizes for various tests in connection with the detection of child abuse. Characterized as flawed by Garb, Wood and Nezworski (2000).

Bornstein (1999) analyzed 51 studies of dependency published between 1950 and 1997. Validity coefficients for projective tests, among them the Rorschach Oral Dependency Scale, were generally larger than those for objective tests.

Hiller, Rosenthal, Bornstein, Berry and Brunell-Neuleib (1999) analyzed Rorschach and MMPI studies. They criticized Atkinson (1986) for using effect sizes that can only assume positive values. Parker, Hanson and Hunsley (1988) were criticized for selection bias and the inclusion of "course" statistics. Garb, Florio and Grove's (1998) reanalysis suffered from the same limitations as the original study by including effect sizes which can only be positive. 30 randomly selected studies of the Rorschach method and the MMPI published from 1977 to 1997, including a few unpublished studies, were analyzed. Overall effect sizes were in the range of .26 to .29 for both instruments. Somewhat higher effects were found for MMPI when self-report inventories were used as criteria, and higher effects were found for the Rorschach when objective measures were used. Commented on by Garb (1999) and by Rosenthal, Hiller, Bornstein and Berry (2001).

Garb, Wood and Nezworski (2000) showed that West systematically excluded nonsignificant results, and a reanalysis showed that projective tests should not be used to detect child abuse.

Rosenthal, Hiller, Bornstein and Berry (2001) reanalyzed their own data (Hiller et al., 1999) according to the criticism from Garb (1999), but essentially the corrections did not have any impact on the results.

Meyer and Archer (2001) reanalyzed Parker et al.'s (1988) data, and found minimal changes from the original analysis. The Rorschach ranged from .27 to .30, the MMPI from .23 to .28, and the WAIS from .32 to .36.

Mihura and Meyer (2002) reported preliminary data from an analysis of interjudge reliability for interpretational judgements in the Rorschach, MMPI, WAIS and TAT. The Rorschach and the MMPI showed the highest levels (.65 and .64, respectively) while the WAIS and TAT were lower (.54 and .46).

Grønnerød (in press) analyzed all 23 published temporal stability studies for the Rorschach method. Stability was high, with overall levels of .68 to .72, and the variables ranged from .40 to .92. Predicted levels based on the regression analysis were .81 and higher.

Grønnerød (2002) analyzed all 30 published studies of assessment of change following psychotherapy with the Rorschach method. Overall effect sizes ranged from .23 to .29, different variables from .02 to .52. Predicted levels based on the regression analysis varied considerably. The Rorschach method was found to be sensitive to changes, and was not inferior to other instruments.

REFERENCES

- Abraham, P. P., Lepisto, B. L., Lewis, M. G., Schultz, L., & et al. (1994). An outcome study: Changes in Rorschach variables of adolescents in residential treatment. *Journal of Personality Assessment*, 62(3), 505-514.
- Acklin, M. W. (1999). Behavioral science foundations of the Rorschach Test: Research and clinical applications. *Assessment*, 6(4), 319-326.
- Acklin, M. W., McDowell, C. J., & Orndoff, S. (1992). Statistical power and the Rorschach: 1975-1991. *Journal of Personality Assessment*, 59(2), 366-379.
- Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment*, 74(1), 15-47.
- Adair, H. E., & Wagner, E. E. (1992). Stability of unusual verbalizations on the Rorschach for outpatients with schizophrenia. *Journal of Clinical Psychology*, 48(2), 250-256.
- Archer, R. P. (1999). Introduction to a special section: Perspectives on the Rorschach. *Assessment*, 6(4), 307-310.
- Archer, R. P., & Gordon, R. A. (1998). MMPI and Rorschach indices of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment*, 52, 276-287.
- Archer, R. P., & Krishnamurthy, R. (1993). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment*, 61(2), 277-293.
- Archer, R. P., & Krishnamurthy, R. (1997). MMPI and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment*, 69, 517-533.
- Aronow, E. (2001). CS norms, psychometrics, and possibilities for the Rorschach technique. *Clinical Psychology: Science and Practice*, 8(3), 383-385.
- Atkinson, L. (1986). The comparative validities of the Rorschach and the MMPI: A meta-analysis. *Canadian Psychology*, 27, 238-247.
- Atkinson, L., Quarrington, B., Alp, I. E., & Cyr, J. J. (1986). Rorschach validity: An empirical approach to the literature. *Journal of Clinical Psychology*, 42(2), 360-362.
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, 11(1), 48-57.
- Bornstein, R. F. (2001). Clinical utility of the Rorschach Inkblot method: Reframing the debate. *Journal of Personality Assessment*, 77(1), 39-47.
- Burns, B., & Viglione, D. J., Jr. (1996). The Rorschach Human Experience Variable, interpersonal relatedness, and object representation in nonpatients. *Psychological Assessment*, 8(1), 92-99.
- Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment*, 11(3), 297-302.
- Erdberg, P. (1999, July 19-24). An international symposium on Rorschach nonpatient data: Findings from around the world. Paper presented at the XVI International Congress of Rorschach and Projective Methods, Amsterdam, the Netherlands.

- Exner, J. E., Jr. (1980). But it's only an inkblot. *Journal of Personality Assessment*, 44, 563-577.
- Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations* (3rd ed.). New York: John Wiley & Sons.
- Exner, J. E., Jr. (1996). A comment on "The comprehensive system for the Rorschach: A critical examination." *Psychological Science*, 7(1), 11-13.
- Exner, J. E., Jr. (2001). A comment on "The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach". *Clinical Psychology: Science and Practice*, 8(3), 386-396.
- Exner, J. E., Jr., & Andronikof-Sanglade, A. (1992). Rorschach changes following brief and short-term therapy. *Journal of Personality Assessment*, 59(1), 59-71.
- Exner, J. E., Jr., Armbruster, G. L., & Viglione, D. J. (1978). The temporal stability of some Rorschach features. *Journal of Personality Assessment*, 42(5), 474-482.
- Gacono, C. B., Loving, J., J. L., & Bodholt, R. B. (2001). The Rorschach and psychopathy: Toward a more accurate understanding of the research findings. *Journal of Personality Assessment*, 77(1), 16-38.
- Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Gacono, C. B., Meloy, J. R., & Berg, J. L. (1992). Object relations, defensive operations, and affective states in narcissistic, borderline, and antisocial personality disorder. *J Pers Assess*, 59(1), 32-49.
- Ganellen, R. J. (1996). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: A review. *Journal of Personality Assessment*, 67(2), 219-243.
- Ganellen, R. J. (2001). Weighting the evidence the Rorschach's validity: A response to Wood et al. (1999). *Journal of Personality Assessment*, 77(1), 1-15.
- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4(6), 641-655.
- Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. *Assessment*, 6(4), 313-317.
- 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(5), 402-404.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1999). The Rorschach controversy: Reply to Parker, Hunsley, and Hanson. *Psychological Science*, 10(3), 293-294.
- Garb, H. N., Wood, J. M., & Nezworski, M. T. (2000). Projective techniques and the detection of child sexual abuse. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, 5(2), 161-168.
- Garb, H. N., Wood, J. M., Nezworski, M. T., & Grove, W. M. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment*, 13(4), 423-448.
- Garfield, S. L. (2000). "The Rorschach test in clinical diagnosis": A brief commentary. *Journal of Clinical Psychology*, 56(3), 431-434.
- Grønnerød, C. (2002). Assessment of change following psychotherapy with the Rorschach method: A meta-analytic review. Submitted manuscript.
- Grønnerød, C. (in press). Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment*.

- Haller, N., & Exner, J. E., Jr. (1985). The reliability of Rorschach variables for inpatients presenting symptoms of depression and/or helplessness. *Journal of Personality Assessment*, 49, 516-521.
- 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(3), 278-296.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11(3), 266-277.
- Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach? An analysis of the evidence. *Psychological Assessment*, 13(4), 472-485.
- Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice*, 8(3), 378-382.
- Lerner, P. M. (2000). A nonreviewer's comment: On the Rorschach and baseball. *Journal of Clinical Psychology*, 56(3), 439.
- 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.
- McDowell, C., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interrater reliability: Conceptual and empirical foundations. *Journal of Personality Assessment*, 66(2), 308-320.
- Meyer, G. J. (1992). The Rorschach's factor structure: A contemporary investigation and historical review. *Journal of Personality Assessment*, 59(1), 117-136.
- Meyer, G. J. (1993). The impact of response frequency on 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(4), 480-489.
- Meyer, G. J. (1997b). Thinking clearly about reliability. More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment*, 9, 495-498.
- Meyer, G. J. (1999). Introduction to the first Special Section in the Special Series on the utility of the Rorschach for clinical assessment. *Psychological Assessment*, 11(3), 235-239.
- Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment*, 75, 46-81.
- Meyer, G. J. (2001a). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science and Practice*, 8(3), 389-396.
- Meyer, G. J. (2001b). Introduction to the final Special Section in the Special Series on the utility of the Rorschach for clinical assessment. *Psychological Assessment*, 13(4), 419-422.
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment*, 13(4), 486-502.
- 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), 1-38.
- Meyer, G. J., & Handler, L. (2000). Correction to Meyer and Handler (1997). *Journal of Personality Assessment*, 74(3), 504-506.

- Mihura, J., & Meyer, G. J. (2002, September 9-14). Interjudge reliability for interpretive judgment made on psychological tests: a meta-analysis. Paper presented at the XVII International Congress of Rorschach and Projective Methods, Rome, Italy.
- Ornberg, B., & Zalewski, C. (1994). Assessment of adolescents with the Rorschach: A critical review. *Assessment*, 1(2), 209-217.
- Parker, K. C. H. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment*, 47(3), 227-231.
- 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(3), 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(3), 291-292.
- Perry, W., McDougall, A., & Viglione, D. (1995). A five-year follow-up on the temporal stability of the Ego Impairment Index. *Journal of Personality Assessment*, 64(1), 112-118.
- 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(3), 487-501.
- Rosenthal, R., Hiller, J. B., Bornstein, R. F., & Berry, D. T. R. (2001). Meta-analytic methods, the Rorschach, and the MMPI. *Psychological Assessment*, 13(4), 449-451.
- Schwartz, N. S., Mebane, D. L., & Malony, H. N. (1990). Effects of alternate modes of administration on Rorschach performance of deaf adults. *Journal of Personality Assessment*, 54(3-4), 671-683.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, 73, 305-316.
- Stricker, G., & Gold, J. R. (1999). The Rorschach: Toward a nomothetically based, idiographically applicable configurational model. *Psychological Assessment*, 11(3), 240-250.
- Stricker, G., & Healey, B. J. (1990). Projective assessment of object relations: A review of the empirical literature. *Psychological Assessment*, 2(3), 219-230.
- Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment*, 11(3), 251-265.
- Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment*, 13(4), 452-471.
- Weiner, I. B. (1996). Some Observations on the Validity of the Rorschach Inkblot Method. *Journal of Personality Assessment*, 8, 206-213.
- Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment*, 6(4), 327-339.
- Weiner, I. B. (2000). Using the Rorschach properly in practice and research. *Journal of Clinical Psychology*, 56(3), 435-438.
- Weiner, I. B. (2001). Advancing the science of psychological assessment: The Rorschach Inkblot Method as exemplar. *Psychological Assessment*, 13(4), 423-432.
- Weiner, I. B., & Exner, J. E., Jr. (1991). Rorschach changes in long-term and short-term psychotherapy. *Journal of Personality Assessment*, 56(3), 453-465.

- West, M. M. (1998). Meta-analysis of studies assessing the efficacy of projective techniques in discriminating child sexual abuse. *Child Abuse and Neglect*, 22(11), 1151-1166.
- Widiger, T. A. (2001). The best and the worst of us? *Clinical Psychology: Science and Practice*, 8(3), 374-377.
- Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment*, 6(4), 341-351.
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000a). Limitations of the Rorschach as a diagnostic tool: A reply to Garfield (2000), Lerner (2000), and Weiner (2000). *Journal of Clinical Psychology*, 56(3), 441-448.
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000b). The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology*, 56(3), 395-430.
- Wood, J. M., Lilienfeld, S. O., Nezworski, M. T., & Garb, H. N. (2001). Coming to grips with negative evidence for the Comprehensive System for the Rorschach: A comment on Gacono, Loving and Bodholt; Ganellen; and Bornstein. *Journal of Personality Assessment*, 77(1), 49-70.
- Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001a). The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice*, 8(3), 350-373.
- Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001b). Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations. *Clinical Psychology: Science and Practice*, 8(3), 397-402.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996a). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, 7(1), 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(1), 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(4), 490-494.
- Wood, J. M., Nezworski, M. T., Stejskal, W. J., & Garven, S. (2001). Advancing scientific discourse in the controversy surrounding the Comprehensive System for the Rorschach: A rejoinder to Meyer (2000). *Journal of Personality Assessment*, 76(3), 369-378.
- Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999a). Erratum for "Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996)". *Assessment*, 6(3), 305.
- Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999b). Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). *Assessment*, 6(2), 115-129.