

A MEANINGFUL PSYCHOMETRIC TEST OR A DECEPTIVE OUIJA BOARD? A CRITICAL ANALYTICAL REVIEW OF THE RORSCHACH INKBLOT TEST

Monika dos Santos, South Africa

(DPhil Psychology, PhD Clinical Psychology Student of Texila American University)

E-mail: monikad@foundation.co.za

ABSTRACT

Controversy has surrounded the Rorschach throughout most of its history, not because it is worthless, but because it has so often been used for the wrong purposes. Psychlit, PEP, PubMed, Google scholar, CAB Abstracts and article references were searched to identify critical commentaries and published globally in English between 1921 and 2012. Findings of the review suggest that the virtues of the Rorschach are modest but genuine.

INTRODUCTION

Remarkable qualities have been ascribed to the Rorschach inkblot test ever since the 1940s, when devotees were fond of comparing its supposed penetrating powers to those of an X-ray (Klopfer, 1940). The test is still held forth as a broad-spectrum measure for a multitude of personality traits and psychological ills, including sense of self-worth, depression, inadequate coping, problem solving deficits, and psychopathy (Exner, 2003; Gacono & Meloy, 1994). Prominent Rorschach advocates have also asserted that it can provide helpful information for identifying individuals who have been abused, forecasting criminal recidivism, and predicting the onset of cancer (Meyer, Finn, Eyde, Kay, Kubiszyn, Moreland, Eisman & Dies, 1998; Viglione, 1999; Kubiszyn, Meyer, Finn & Eyde, 2000).

Such claims, which far outstrip the scientific evidence, have tended to discredit the Rorschach in the eyes of many research-oriented psychologists. As a consequence, some sectors of psychology regard the test as an unfortunate vestige from the discipline's past, only one step removed from tea leaves and crystal balls. But although such dismissals are understandable, they may be too harsh. More than 50 years of research may have confirmed woefully short of the claims made by proponents, nevertheless possess 'validity greater than chance'.

Although the book *What's Wrong With the Rorschach?* identifies the Rorschach's numerous shortcomings, articles such as that by Wood, Nezworski and Garb (2003) (*What's Right with the Rorschach? The Scientific Review of Mental Health Practice*) focus on those aspects of the test with genuine merit and suggest ways in which the Rorschach can fruitfully be used in clinical assessment, research, and therapy (Wood, Nezworski, Lilienfeld & Garb, 2003a). Controversy may have surrounded the Rorschach throughout most of its history not because it is worthless, but because it has so often been used for the wrong purposes.

The following five aspects of the Rorschach assessment, each of which has received considerable attention in the literature, define the basic nature of the instrument: Rorschach assessment is both an objective and a subjective procedure; the Rorschach measures both perceptual and

associational processes, the Rorschach assesses both structural and dynamic aspects of personality functioning, Rorschach testing constitutes multifaceted method of data collection, and the Rorschach assessment rests on a sound psychometric foundation (Weiner, 2003).

Using interpretation of 'ambiguous designs' to assess an individual's personality is an idea that goes back to artists such as Leonardo da Vinci and Botticelli. Interpretation of inkblots was central to a game from the late 19th century. Rorschach's, however, was the first systematic approach of this kind (Groth-Marnat 2003).

It has been suggested that Rorschach's use of inkblots may have been inspired by German doctor Justinus Kerner who, in 1857, had published a popular book of poems, each of which was inspired by an accidental inkblot (Pichot, 1984). French psychologist Alfred Binet had also experimented with inkblots as a creativity test, and, after the turn of the century, psychological experiments where inkblots were utilized multiplied, with aims such as studying imagination and consciousness (Goldstein & Hersen, 2000). book *Psychodiagnostik*, which was to form the basis of the inkblot test (after experimenting with several hundred inkblots, he selected a set of ten for their diagnostic value), but he died the following year (Romesch, 2003). Although he had served as Vice President of the Swiss Psychoanalytic Society, Rorschach had difficulty in publishing the book and it attracted little attention when it first appeared (Alfred, 2009).

In 1927, the newly-founded Hans Huber publishing house purchased Rorschach's book *Psychodiagnostik* from the inventory of Ernst Bircher. Huber has remained the publisher of the test and related book, with Rorschach a registered trademark of Swiss publisher Verlag Hans Huber, Hogrefe AG (*Psychodiagnostics: A Diagnostic Test Based on Perception*, 1998). The work has been described as a densely written piece embedded in dry, scientific terminology (Acklin & Oliveira-Berry, 1996).

After Rorschach's death, the original test scoring system was improved by Samuel Beck, Bruno Klopfer and others. John E. Exner summarized some of these later developments in the comprehensive system, at the same time trying to make the scoring more statistically rigorous. Some systems are based on the psychoanalytic concept of object relations. The Exner system remains very popular in the United States, while in Europe other methods sometimes dominate, such as that described in the textbook by Evald Bohm, which is closer to the original Rorschach system and rooted more deeply in the original psychoanalysis principles (Lang, 1989; Dana, 2000).

Rorschach History

In the 1960s, the Rorschach was the most widely used projective test (Chapman & Chapman, 1982). In a national survey in the United States, the Rorschach was ranked eighth among psychological tests used in outpatient mental health facilities (Gacano & Meloy, 1994). It is the second most widely used test by members of the Society for Personality Assessment, and it is requested by psychiatrists in 25% of forensic assessment cases, usually in a battery of tests that often include the MMPI-2 and the MCMI-III (Gacano & Meloy, 1994; Gacano & Evans, 2007). In surveys, the use of Rorschach ranges from a low of 20% by correctional psychologists to a high of 80% by clinical psychologists engaged in assessment services, and McIvor, 2008; Weiner & Greene 2007).

Although the Exner Scoring System (developed since the 1960s) claims to have addressed and often refuted many criticisms of the original testing system with an extensive body of research some researchers continue to raise questions (Exner, 2002). The areas of dispute include the objectivity of testers, inter-rater reliability, the verifiability and general validity of the test, bias

of the test's pathology scales towards greater numbers of responses, the limited number of psychological conditions which it accurately diagnoses, the inability to replicate the test's norms, its use in court-ordered evaluations, and the proliferation of the ten inkblot images, potentially invalidating the test for those who have been exposed to them (Lilienfeld, Wood & Garb, 2001). Exner (1993) contends that the Rorschach makes well-validated contributions in the domain of identification of particular treatment goals, recognising possible obstacles to progress in psychotherapy, selecting appropriate treatment modalities, and monitoring change and progress over time. Therefore, data that are obtained, coded, and presented according to Exner's Comprehensive System produce a reliable set of scores that have empirically significant and meaningful correlations in dynamics of personality functioning (Weiner, 1998). Many theorists believe that emphasis on individualisation in treatment is accomplished through making appropriate judgements about Rorschach results which yield important information regarding an individual's psychological experiences and functioning (Butcher, 1997).

Controversy

Some skeptics consider the Rorschach inkblot test pseudoscience, as several studies suggested that conclusions reached by test administrators since the 1950s were akin to cold reading (Lilienfeld, et al., 2001; Drenth, 2003; Wood, Nezworski, Lilienfeld & Howard, 2003a). In the 1959 edition of *Mental Measurement Yearbook*, Lee Cronbach (former President of the Psychometric Society and American Psychological Association) stated that the test has repeatedly failed as a prediction of practical criteria and that there is nothing in the literature to encourage reliance on Rorschach interpretations (Alexander, 2001). In tests have been administered by hundreds of trained professionals since that time (of a previous review), and while many relationships to personality dynamics and behavior have been hypothesized, the vast majority of these relationships have never been validated empirically, despite the appearance of more than 2,000 publications about the test' (Dawes, 1991:154). A moratorium on its use was called for in 1999 (Garb, 1999). A 2003 report by Wood and colleagues had more mixed views: 'More than 50 years of research have confirmed Lee J. Cronbach's (1970) final verdict: that some Rorschach scores, though falling woefully short of the claims made by proponents, nevertheless possess "validity greater than chance". [...] "Its value as a measure of thought disorder in schizophrenia research is well accepted. It is also used regularly in research on dependency, and, less often, in studies on hostility and anxiety. Furthermore, substantial evidence justifies the use of the Rorschach as a clinical measure of intelligence and thought disorder' (Wood, Nezworski & Garb, 2003: 636).

Illusory and invisible correlations

In the 1960s, research by psychologists Loren and Jean Chapman showed that at least some of the apparent validity of the Rorschach was due to an illusion. At that time, the five signs most often interpreted as diagnostic of homosexuality were 1) buttocks and anuses; 2) feminine clothing; 3) male or female sex organs; 4) human figures without male or female features; and 5) human figures with both male and female features. The Chapmans surveyed 32 experienced testers about their use of the Rorschach to diagnose homosexuality. At this time homosexuality was regarded as a psychopathology, and the Rorschach was the most popular projective test utilized (Chapman & Chapman, 1982; Sutherland, 2007). The testers reported that homosexual men had shown the five signs more frequently than heterosexuals. Despite these beliefs, analysis of the results showed that heterosexual men are just as likely to report these signs, so they are totally ineffective for identifying homosexuals. The five signs did, however, match the guesses students made about which imagery would be associated with homosexuality. Students read

through a stack of cards, each with a Rorschach blot, a sign and a pair of 'conditions' (which might include homosexuality). The information on the cards was fictional, although subjects were told it came from case studies of real patients. The students reported that the five invalid signs were associated with homosexuality, even though the cards had been constructed so there was no association at all. The Chapmans repeated this experiment with another set of cards, in which the association was negative; the five signs were never reported by homosexuals. The students still reported seeing a strong positive correlation (Chapman & Chapman, 1982). These experiments showed that the testers' prejudices could result in them 'seeing' non-existent relationships in the data. The Chapmans called this phenomenon 'illusory correlation' and it has since been demonstrated in many other contexts (Hardman, 2009).

A related phenomenon called 'invisible correlation' applies when people fail to see a strong association between two events because it does not match their expectations. This was also found in clinicians' interpretations of the Rorschach. Homosexual men are more likely to see a monster on Card IV or a part-animal, part-human figure in Card V. Almost all of the experienced clinicians in the Chapmans' survey missed these valid signs (Chapman & Chapman, 1982). The Chapmans ran an experiment with fake Rorschach responses in which these valid signs were always associated with homosexuality. The subjects missed these perfect associations and instead reported that invalid signs, such as buttocks or feminine clothing, were better indicators (Hardman, 2009).

In 1992, the psychologist Stuart Sutherland argued that these artificial experiments are easier than the real-world use of the Rorschach, and hence they probably underestimated the errors that testers were susceptible to. He described the continuing popularity of the Rorschach after the Chapmans' research as a 'glaring example of irrationality among psychologists' (Sutherland, 2007).

Tester Projection

Some critics argue that the testing psychologist must also project onto the patterns. A possible example sometimes attributed to the psychologist's subjective judgement is that subject's response fits with how the blot actually looks. Superficially this might be considered a subjective judgment, depending on how the examiner has internalised the categories involved. But with the Exner system of scoring, much of the subjectivity is eliminated or reduced by use of frequency tables that indicate how often a particular response is given by the population in general (Exner, 2002). Another example is that the response 'bra' was considered a 'sex' response by male psychologists, but a 'clothing' response by female (Wood, Nezworski & Garb, 2003). In Exner's system, however, such a response is always coded as 'clothing' unless there is a clear sexual reference in the response (Exner, 2002).

Third parties could be used to avoid this problem, but the Rorschach's inter-rater reliability has been questioned. That is, in some studies the scores obtained by two independent scorers do not match with great consistence (Wood, Nezworski & Garb, 2003). This conclusion was challenged in studies using large samples reported in 2002 (Meyer, Hilsenroth, Baxter, Exner, Fowler, Piers & Resnick, 2002).

Cultural differences

Comparing North American Exner normative data with data from European and South American subjects showed marked differences in some features, some of which impact important variables, while others (such as the average number of responses) coincide (Dana, 1999). For instance,

texture response is typically zero in European subjects (if interpreted as a need for closeness, in accordance with the system, a European would seem to express it only when it reaches the level of a craving for closeness), and there are fewer 'good form' responses, to the point where schizophrenia may be suspected if data were correlated to the North American norms (Dana, 1999). Form is also often the only determinant expressed by European subject; while colour is less frequent than in American subjects, colour-form responses are comparatively frequent in opposition to form-colour responses; since the latter tend to be interpreted as indicators of a defensive attitude in processing affect, this difference could stem from a higher value attributed to spontaneous expression of emotions (Dana, 1999). Cultures will exhibit different 'common' objects (French subjects often identify a chameleon in card VIII, which is normally classed as an 'unusual' response, as opposed to other animals like cats and dogs; in Scandinavia, 'Christmas elves' (nisser) is a popular response for card II, and 'musical instrument' on card VI is popular for Japanese people), and different languages will exhibit semantic differences in naming the same object (the figure of card IV is often called a troll by Scandinavians and an ogre by French people) (Weiner, 2003; Dana, 1999). Many of Exner's 'popular' responses (those given by at least one third of the North American sample used) seem to be universally popular, as shown by samples in Europe, Japan and South America, while specifically card IX's 'human' response, the crab or spider in card X and one of either the butterfly or the bat in card I appear to be characteristic of North America (Dana, 1999; Weiner, 2003).

Form quality, popular content responses and locations are the only coded variables in the Exner systems that are based on frequency of occurrence, and thus immediately subject to cultural influences; therefore, cultural-dependent interpretation of test data may not necessarily need to extend beyond these components (Weiner, 2003).

The cited language differences mean that it's imperative for the test to be administered in the subject's native language or a very well mastered second language, and, conversely, the examiner should master the language used in the test. Test responses should also not be translated into another language prior to analysis except possibly by a clinician mastering both languages, for example, a bow tie is a frequent response for the center detail of card III, but since the equivalent term in French translates to 'butterfly tie', an examiner not appreciating this language nuance may code the response differently from what is expected (Weiner, 2003).

Validity

When interpreted as a projective test, results are poorly verifiable. The Exner system of scoring (the 'Comprehensive System') is meant to address this, and has all but displaced many earlier (and less consistent) scoring systems. It makes heavy use of what factor (shading, color, outline, etc.) of the inkblot leads to each of the tested person's comments. system, latitude remained in the actual interpretation, and the clinician's write-up of the test record is still partly subjective (Goldman, 2000). Reber (1985:652) comments '.. there is essentially no evidence whatsoever that the test has even a shred of validity.'

Nevertheless, there is substantial research indicating the utility of the measure for a few scores. Several scores correlate well with general intelligence. Interestingly, one such scale is R, the total number of responses; this reveals the questionable side-effect that more intelligent people tend to be elevated on many pathology scales, since many scales do not correct for high R: if a subject gives twice as many responses overall, it is more likely that some of these will seem 'pathological'. Also correlated with intelligence are the scales for Organisational Activity,

Complexity, Form Quality, and Human Figure response (Wood, Nezworski & Garb, 2003). The same source reports that validity has also been shown for detecting such conditions as schizophrenia and other psychotic disorders; thought disorders; and personality disorders (including borderline personality disorder). There is some evidence that the Deviant Verbalisations scale relates to bipolar disorder (Khadiji, Wetzler & Wilson, 1997). Wood et al. (2003) conclude that 'otherwise, the Comprehensive System doesn't appear to bear a consistent relationship to psychological disorders or symptoms, personality characteristics, potential for violence, or such health problems as cancer (cancer is mentioned because a small minority of Rorschach enthusiasts have claimed the test can predict cancer) (Wood, Nezworski & Garb, 2003; Graves, Thomas & Mead, 1991).

Reliability

It is also thought that the test's reliability can depend substantially on details of the testing procedure, such as where the tester and subject are seated, any introductory words, verbal and nonverbal responses to subjects' questions or comments, and how responses are recorded. Exner has published detailed instructions, but Wood et al. (2003) cites many court cases where these had not been followed. Similarly, the procedures for coding responses are fairly well specified but extremely time-consuming leaving them very subject to the author's style and the publisher to the quality of the instructions (such as was noted with one of Bohm's textbooks in the 1950s as well as clinic workers (which would include examiners) being 1975).

United States Courts have challenged the Rorschach as well. Jones v Apfel (1997) stated (quoting from Attorney's Textbook of Medicine) that Rorschach 'results do not meet the requirements of standardization, reliability, or validity of clinical diagnostic tests, and interpretation thus is often controversial' (Gacono & Evans, 2007:83). In State ex rel H.H. (1999) where under cross examination Dr Bogacki stated under oath 'many psychologists do not believe much in the validity or effectiveness of the Rorschach test' and US v Battle (2001) ruled that the Rorschach 'does not have an objective scoring system' (Gacono & Evans, 2007:83).

Population Norms

Another controversial aspect of the test is its statistical norms. Exner's system was thought to possess normative scores for various populations. But, beginning in the mid-1990s others began to try to replicate or update these norms and failed. In particular, discrepancies seemed to focus on indices measuring narcissism, disordered thinking, and discomfort in close relationships (Lillienfeld, Wood & Garb, 2000). Lillienfeld et al. (2000), who are critical of the Rorschach, have stated that this proves that the Rorschach tends to 'overpathologise normals'. Although Rorschach proponents, such as Hibbard (2000), suggest that high rates of pathology detected by the Rorschach accurately reflect increasing psychopathology in society, the Rorschach also identifies half of all test-takers as possessing 'distorted thinking', a false positive rate unexplained by current research (Radford, 2009).

The accusation of 'over-pathologising' has also been considered by Meyer, Erdberg and Shaffer (2007). They presented an international collaborative study of 4 704 Rorschach protocols, obtained in 21 different samples, across 17 different countries, with only 2% showing significant elevations on the index of perceptual and thinking disorder, 12% elevated on indices of depression and hyper-vigilance and 13% elevated on persistent stress overload—all in line with expected frequencies among nonpatient populations.

Albert Binet considered including inkblots in his famous intelligence test (Zubin, Eron & Schumer, 1965). Although he eventually abandoned the idea, his original intuition turned out to be correct. As research has shown, several Rorschach variables are correlated with intelligence test scores (for reviews, see Meyer, 1992; Wood, Krishnamurthy & Archer, 2003). The highest correlations, which range from .30 to .40, have been found for Developmental Quality and Organizational Activity, scores that measure the degree to which responses synthesize diverse parts of a blot into a unified image. Lambda and the closely related F%, which reflect a tendency to give responses based on colour and shading rather than form alone, also appear to correlate above .30 with intelligence test scores. Somewhat lower (.20 to .30) are the correlations for Form Quality, Human Movement responses, and R (the total number of responses given to the blots). However, the best Rorschach indicator of intelligence is to be found not among these scores but in the vocabulary that the respondent uses to describe the blots (Davis, 1961; Hauser, 1963; Trier, 1958). For example, Thomas Trier of the University of California at Berkeley asked clinicians to read a group of Rorschach protocols and identify the seven most sophisticated words used by each respondent. Then, by consulting a commonly available word book, he estimated the average grade level of these words for each respondent. This simple Rorschach-based estimate of vocabulary level correlated .77 with intelligence test scores.

Although such results demonstrate that Rorschach responses can be used to estimate intelligence, modern standardised intelligence tests are definitely superior for the purpose (Davis, 1961). However, when intelligence testing is impossible, for example with an uncooperative child, inkblots may provide an acceptable substitute. The use of Rorschach-based vocabulary as an index of intelligence has been virtually ignored in the assessment literature since the 1950s, so that standardised procedures and norms are unavailable. With some scientific groundwork, however, the Rorschach might well be put on a solid footing as a rough intelligence measure, to be pulled out of the psychologist's briefcase under pressing circumstances.

There is abundant evidence that two kinds of Rorschach scores are related to psychotic disorders. First, as Hermann Rorschach (1921,1964) noted, the inkblot responses of patients with schizophrenia often exhibit poor form quality (Rieman, 1953; Sherman, 1952; see reviews by Frank, 1990; Goldfried, Stricker & Weiner, 1971). That is, the images reported by these patients often do not 'fit' the shape of the blots. Form quality is also poor among many patients with bipolar disorder (Frank, 1990).

Second, as David Rapaport and his colleagues (1946) first noted in their famous book *Diagnostic Psychological Testing*, the Rorschach can be used to identify thought disorder, the disorganised cognition and peculiarities of language exhibited by many patients with schizophrenia. Several scoring methods have been developed to measure thought disorder on the Rorschach (for reviews, see Aronow & Reznikoff, 1976; Goldfried, Stricker & Weiner, 1971; Kleiger, 1999), the most prominent being the Thought Disorder Index (Johnston & Holzman, 1979; Solovay, Shenton, Gasperetti, Coleman, Kestnbaum, Carpenter & Holzman, 1986), the TETRAUT of the Logical Rorschach (Wagner, 2001), and the Weighted Sum (WSum6) of the Comprehensive System for the Rorschach (Exner, 2003). The Comprehensive System's Schizophrenia Index (revised recently as the Perceptual Thinking Index) combines scores for thought disorder and form quality (Exner, 2003). Research has shown that all these scores are related to schizophrenia (Greaves, 2000; Johnston & Holzman, 1979; Jorgensen, Andersen & Dam, 2000; Kleiger, 1999; Wagner, 1998; 2001). Many patients with schizotypal personality disorder and bipolar disorder in the manic phase also apparently exhibit thought disorder on the Rorschach (Coleman, Levy, Lenzenweger & Holzman, 1996; Singer & Brabender, 1993).

The Rorschach—particularly the Thought Disorder Index—has proven useful to researchers who examine genetic and familial patterns of schizophrenia (e.g., Knight & Silverstein, 1998; Lenzenweger, 1998). These various scales are also potentially useful in clinical settings, although it is unclear whether Rorschach indices of thought disorder are necessary if a clinician has already had an opportunity to observe a patient's thinking and language during an interview (for example, see Whitehead, 1985).

indexes of John Exner's (2003) Comprehensive System, currently the most popular method for scoring and interpreting the Rorschach. Exner's indexes (e.g., the SCZI, WSum6, Level 2 scores, and Conventional Form) presently have only limited clinical usefulness because their published norms appear to be seriously in error (Wood, Nezworski, Garb & Lilienfeld, 2001a, 2001b; see Shaffer, Erdberg & Haroian, 1999; also see Exner 2001, Meyer, 2001). Clinicians who rely on the Comprehensive System and its norms are likely to significantly overdiagnose thought disorder and psychotic symptoms.

Applications

The test is also controversial because of its common use in court-ordered evaluations. This controversy stems, in part, from the limitations of the Rorschach, with no additional data, in making official diagnoses from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 1994). Irving B. Weiner (co-developer with John Exner of the Comprehensive system) has stated that the Rorschach is a measure of personality functioning, and that it provides information concerning aspects of personality structure and dynamics that make people the kind of people they are. Sometimes such information about personality characteristics is helpful in arriving at a differential diagnosis, if the alternative diagnoses being considered have been well conceptualised with respect to specific or defining personality characteristics' (Weiner, 1999). In the vast majority of cases, anyway, the Rorschach test wasn't singled out but used as one of several in a battery of tests, and despite the criticism of usage of the Rorschach in the courts, out of 8,000 cases in which forensic psychologists used Rorschach-based testimony, the appropriateness of the instrument was challenged only six times, and the testimony was ruled inadmissible in only one of those cases (Gacono & Evans, 2007; Gacono & Kaser-Boyd, 2007; Weiner & Greene, 2007). One study has found that use of the test in courts has increased by three times in the decade between 1996 and 2005, compared to the previous fifty years (Gacono & Evans, 2007). Others however have found that its usage by forensic psychologists has decreased (Garb, Wood, Lilienfeld & Nezworski, 2005).

Several Rorschach scores have repeatedly demonstrated their validity in research. The Elizur Anxiety and Hostility scales, which are based on the emotional content of patients' responses, have a well-demonstrated relationship to anxious and hostile behaviors (Aronow & Reznikoff, 1976; Goldfried et al., 1971). The Rorschach Oral Dependency scale (ROD), based on responses that involve eating, mouths, or other 'oral' imagery, appears to be a valid measure of normal variations in dependency (Bornstein, 1996), although it has been less successful as a measure of pathological dependency (Bornstein, Hilsenroth & Padawer, 2000; see also Garb, Wood, Nezworski, Grove & Stejskal, 2001).

Rorschach signs identified by Piotrowski (1937) differentiate what used to be called 'organic' from 'functional' brain disorders (Goldfried et al., 1971). For instance, Piotrowski found that many patients with 'organic' brain disorders take a long time to react to the blots and often give repetitious responses. Finally, Klopfer's Rorschach Prognostic Rating Scale (RPRS) has a well-demonstrated relationship to treatment outcomes, for example, patients who report imagery involving animals or humans in movement receive higher scores on Klopfer's scale and have

somewhat better outcomes in psychotherapy (Meyer & Handler, 1997).

Despite their respectable performance in research, these Rorschach scores are currently unsuitable for clinical applications. Most important, they lack adequate norms and involve elaborate scoring procedures that many clinicians may find impractical. In addition, some of these scores (e.g., the RPRS and Elizur scales) were validated using administration or scoring procedures from Beck and Klopfer that are now obsolete. Thus, these Rorschach scores are far more attractive as research instruments than as clinical tools.

Psychotherapy

Aronow and Reznikoff (1976) have long argued that the Rorschach, though arguably a failure as a psychometric test, has considerable value as an adjunct technique in psychotherapy. These authors approach the patient's responses to the blots analogously to dream interpretation, asking the patient 'What does this image make you think of?' or 'What does it bring to mind?' Such an approach seems compatible with some forms of psychotherapy and the Rorschach as a psychotherapeutic technique has not yet been demonstrated. Furthermore, therapists who use the test to generate symbolic interpretations must beware of the potential influence of confirmation bias, and should actively seek evidence that disconfirms their Rorschach interpretations, as well as evidence that confirms them (Nisbett & Ross, 1980).

Measurement of change in psychotherapy

The importance of measuring change through psychotherapy is essential to assess the effectiveness of particular psychotherapeutic strategies and interventions. Accurate measurement of change assists in the development of further strategies and techniques in service of improved psychotherapy. Weiner and Exner (1991) evaluated patients in both long-term and short-term psychotherapy using the Rorschach Inkblot Test (Comprehensive System). Their findings demonstrated the effectiveness of long-term therapy, and the validity of the Rorschach in assessing the effects and changes in therapy.

Weiner and Exner (1991) also identified a number of indicators for successful psychotherapy. These were the patient's ability to manage stress more adequately, to deal with problematic situations in a specific coping style, and to be more aware of their experiences, to be involved in positive self-examination, and to be more comfortable in interpersonal situations. Improvement and change in psychological functioning through therapy can therefore be reflected in these psychological activities.

In the study of Weiner and Exner (1991), part of the indices that could indicate psychotherapeutic change included specific variables. These affect variables are related to underlying personality structures that influence the use and expression of affect in individuals. These Comprehensive System indices are D, Adjusted D, Lambda, Affect Ratio, Shading responses, EA, Texture responses, EB and Colour responses.

Weiner and Exner (1991) found a general improvement when measuring change in psychotherapy subjects, after one year of treatment. The general improvement in functioning was consistent with identified indicators for successful psychotherapy (e.g. the ability to manage stress more adequately) in the functioning of psychotherapy subjects. Generally, the depressed, enjoying experiences and modulating affect more effectively, being more realistic as opposed to escapist, and improved interpersonal relationships. The measured changes perpetuated over the period of testing, and up to four years after treatment commenced. After one year of therapy there were areas of functioning that indicated little change, as indicated by Rorschach tested

variables (Adjusted D, D, and EA). According to Weiner and Exner (1991) this indicated the subjects as experiencing subjectively felt distress. This relates the finding of Weiner and Exner (1991) in a significant increase of the Form Dimension (FD) responses after one year of therapy, and up to two years after therapy commenced, which indicates increased self-examination during this period of therapy. The results of Weiner and Exner's study (1991) demonstrates change in six areas of personality functioning, namely, stress management, a conventional and consistent manner of dealing with experiences, being more capable of taking enjoyment from emotional experiences and modulating affect, more effective ideation, being less preoccupied with themselves, and having and desiring better interpersonal relationships.

The Weiner and Exner (1991) study demonstrates the beneficial effects of psychotherapy through improved functioning of the subjects in the six identified dimensions of personality, as previously mentioned. The long-term patients displayed greater beneficial personality changes in comparison to the short-term psychotherapy patients. These changes were improvement in the frequency of loose and arbitrary thinking, excessive intellectualisation, excessive self-focusing, and the lack of expecting close and interpersonal relationships. The accuracy of the Rorschach is demonstrated through the research of Weiner and Exner (1991:464), as they state, 'the successful measurement of these expected measurement by Rorschach variables lends construct validity to their use for this and related purposes'. This lends validity to changes through psychotherapy, as assessed by the Rorschach. As stated by Weiner and Exner (1991:464), changes seen in psychotherapy through Rorschach assessments may not be expected unless '(a) psychotherapy makes a difference and (b) the Rorschach can validity measure this difference.' This demonstration of the accuracy of Rorschach assessment of change through therapy in the Weiner and Exner (1991) study gives validity to Rorschach assessment in this case study of change in psychotherapy.

Paradoxically, although the Rorschach is held in disrepute by many research psychologists, it has perhaps achieved its greatest successes as a research tool. Its value as a measure of thought disorder in schizophrenia research is well accepted. It is also used regularly in research on dependency, and, less often, in studies on hostility and anxiety. Furthermore, substantial evidence justifies the use of the Rorschach as a clinical measure of intelligence and thought disorder. Although clinicians should normally rely on well-established tests such as the Wechsler Adult Intelligence Scale-Third Edition (Wechsler, 1997) to measure intelligence, and on clinical interviews to assess thought disorder, there may be times when the Rorschach can usefully supplement these 'front-line' methods. In addition, the Rorschach may be useful as an exploratory technique in some forms of insight-oriented psychotherapy.

The virtues of the Rorschach are modest but genuine. If, over its long history, the test had been promoted solely for the uses identified here, it probably would have been less popular among psychologists, but also far less controversial. It remains to be seen whether clinical psychologists of the future can learn to accept the limitations of the Rorschach while respecting its strengths, otherwise, it will continue to be promoted for purposes for which it has no usefulness and will inevitably be a flashpoint for controversy.

REFERENCES

1. Acklin M. W. & Oliveira-Berry J. (1996). Return to the source: Rorschach's Psychodiagnostics. *Journal of Personality Assessment* 67: 427-433.
2. Alexander, M. (2001). Lee Cronbach, pioneer in educational psychology dead at 85.

- Stanford Report (Stanford University School of Education). Accessed 28 December 2011 from <http://ed.stanford.edu/suse/news-bureau/displayRecord.php?tablename=press&id=12>
3. Alfred, R. (2009). April 2, 1922: Rorschach Dies, Leaving a Blot on His Name at wired.com Accessed on 28 December 2011 at http://www.wired.com/science/discoveries/news/2009/04/dayintech_0402
 - a. disorders (4th ed.). Washington, DC.
 4. Aronow, E. & Reznikoff, M. (1976). Rorschach content interpretation. New York: Grune & Stratton. p. 7.
 5. Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency Scale: 1967–1995. *Psychological Assessment*, 8, 200–205.
 6. Bornstein, R. F., Hilsenroth, M. J. & Padawer, J. R. (2000). Interpersonal dependency and personality pathology: Variations in Rorschach Oral Dependency scores across Axis II disorders. *Journal of Personality Assessment*, 75, 478–491.
 7. Buros, O.K. (1975). *Personality tests and reviews: including an index to the mental measurements yearbooks, Volume 1*. Gryphon Press.
 8. Butcher J.A. (1992). Introduction to the special section on assessment in psychological treatment: a necessary step for effective intervention. *Psychological Assessment*, 9, 331–333.
 9. Chapman, L.J. & Chapman, J. (1982). Test results are what you think they are. In Kahneman, D, Slovic, P. & Tversky, A. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press. pp. 238–248.
 10. Coleman, M. J., Levy, D. L, Lenzenweger, M. F. & Holzman, P. S. (1996). Thought disorder, perceptual aberrations, and schizotypy. *Journal of Abnormal Psychology*, 105, 469–473.
 11. Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
 - Dana, R.H. (1999). *Handbook of Cross-Cultural and Multicultural Personality Assessment* (Personality and Clinical Psychology Series).
 12. Dawes, M.R. (1991). Giving up Cherished Ideas: The Rorschach Ink Blot Test. *Institute for Psychological Therapies Journal* 3, 4. Accessed 30 December 2011 from http://www.ipt-forensics.com/journal/volume3/j3_4_5.htm
 13. Drenth, P.J.D. (2003). Growing Anti-intellectualism in Europe: A Menace to Science. Annual Report 2003. ALLEA (All European Academies). Accessed 4 January 2012 from <http://www.allea.org/Pages/ALL/4/881.pdf#page=61> Volume 1. Hoboken, NJ: John Wiley & Sons.
 14. Exner, J. E. (2003). *The Rorschach: A comprehensive system. Basic foundations and principles of interpretation* (4th ed.). New York: Wiley.
 15. Frank, G. (1990). Research on the clinical usefulness of the Rorschach: 1. The diagnosis of schizophrenia. *Perceptual and Motor Skills*, 71, 573–578.
 16. Garb, H.N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. *Assessment*, 6,4: 313–8.
 17. Garb, H.N., Wood, J.M., Lilienfeld, S.O. & Nezworski, M.T. (2005). Roots of the Rorschach controversy. *Clinical Psychological Review*, 25 (1): 97–118.
 18. Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M. & Stejskal, W. J. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment*, 13, 433–448.

19. Gacono, C. B. & Evans F.B. (2007). *The Handbook of Forensic Rorschach Assessment*. New York: Routledge
20. Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Erlbaum.
21. Goldfried, M. R., Stricker, G. & Weiner, I. B. (1971). *Rorschach handbook of clinical and research applications*. Englewood Cliffs, NJ: Prentice-Hall.
22. Goldman, H.H. (2000). *Review of General Psychiatry*. Maryland: McGraw-Hill.
23. Goldstein, G & Hersen, M(eds). (2000). *Handbook of psychological assessment*. Amsterdam: Pergamon Press. p. 437.
24. Greaves, A. R. (2000). *A validation of Wagner's Rorschach autism classification system*. Unpublished doctoral dissertation, Forest Institute of Professional Psychology.
25. Groth-Marnat, G. (2003). *Handbook of Psychological Assessment*. Canada: John Wiley & Sons, Inc.
26. Hardman , D. (2009). *Judgement and Decision making: Psychological perspectives*. West Sussex: BPS Blackwell.
27. Hauser, R. J. (1963). *The validity of the formal and linguistic aspects of the Rorschach in predicting intelligence*. (Doctoral dissertation, New York University, 1962). *Dissertation Abstracts International*, 24, 833.
28. Johnston, M. H. & Holzman, P. S. (1979). *Assessing schizophrenic thinking*. San Francisco: Jossey-Bass.
29. Jorgensen, K., Andersen, T. J. & Dam, H. (2000). *The diagnostic efficiency of the Rorschach depression index and the schizophrenia index: A review*. *Assessment*, 7, 259–280.
30. Khadivi, A., Wetzler, S. & Wilson, A. (1997). *Manic indexes on the Rorschach*. *Journal of Personality Assessment*, 69, 365–375. Klopfer, B. (1940). *Personality aspects revealed by the Rorschach method*. *Rorschach Research Exchange*, 4, 26–29.
31. Knight, R. A. & Silverstein, S. M. (1998). *The role of cognitive psychology in guiding research on cognitive deficits in schizophrenia: A process-oriented approach*. In M. F. Lenzenweger & R. H. Dworkin (Eds.), *Origins and development of schizophrenia*. *Advances in experimental psychopathology* (pp. 247–295).
32. Kubiszyn, T. W., Meyer, G. J., Finn, S. E. & Eyde, L. D. (2000). *Empirical support for psychological assessment in clinical health care settings*. *Professional Psychology: Research and Practice*, 31, 119–130.
33. Lang, M. (1989). *Psicologia clinica*. Milano: F. Angeli. ("Nonostante il Sistema Comprensivo di J.E. Exner rappresenti ai nostri giorni il Metodo Rorschach più diffuso a livello mondiale, in Italia è ancora non molto utilizzato. Although J. E. Exner's Comprehensive Systems nowadays represents the most widely adopted method worldwide, it is not yet very widespread in Italy.").
34. Lilienfeld, S.O., Wood, J.M. & Garb, H.N. (2001). *What's wrong with this picture?* *Scientific American*.
35. Lowrey, L.G. (1946). *American journal of orthopsychiatry*. *American Orthopsychiatric Association*, 16, p. 732.
36. Meyer, G. J. (1992). *The Rorschach's factor structure: A contemporary investigation and historical review*. *Journal of Personality Assessment*, 59, 117–136.
37. Meyer, G. J. (2001). *Evidence to correct misperceptions about Rorschach norms*. *Clinical Psychology: Science and Practice*, 8, 389–396.

38. Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69, 1–38.
39. Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C. & Resnick J. (2002). An examination of interrater reliability for scoring the Rorschach comprehensive system in eight data sets. *Journal of Personality Assessment*, 78 (2): 219–274.
40. Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall. Pichot, P. (1984). Centenary of the birth of Hermann Rorschach. (Rosenzweig, S. & Schriber, E. Trans.). *Journal of Personality Assessment* 48: 591–596.
41. Radford, B. (2009). Rorschach Test: Discredited But Still Controversial. Accessed on 26 December 2011 from <http://www.livescience.com/9695-rorschach-test-discredited-controversial.html>
42. Rapaport, D., Gill, M. & Schafer, R. (1946). *Diagnostic psychological testing*. Vol. II. Chicago: Year Book Publishers.
43. Raynor, P. & McIvor, G. (2008). *Developments in Social Work Offenders (Research Highlights in Social Work)*. London: Jessica Kingsley Publishers. p. 138.
44. Reber, A.S. (1985). *Penguin Dictionary of Psychology*. Penguin Books.
45. Romesh.V. (2003). *Textbook Of Statistics, Psychology & Education*. Anmol Publications.
46. Rorschach, H. (1927). *Rorschach Test – Psychodiagnostic Plates*. Hogrefe.
47. Routledge Davis, H. S. (1961). Judgments of intellectual level from various features of the Rorschach including vocabulary. *Journal of Projective Techniques*, 25, 155–157.