

## THE RORSCHACH TECHNIQUE: A RE-EVALUATION

ARTHUR R. JENSEN

*University of California*

In the 43 years since Hermann Rorschach published the *Psycho-diagnostik*, his set of ten carefully chosen inkblots has become the most popular of all psychological tests. A recent survey of hospitals, clinics, guidance centers, and the like, indicates that the Rorschach clearly outstrips all its competitors, both in the number of institutions using the test and in the amount of usage (31). Furthermore, the curve depicting the increase in popularity of the Rorschach over the past decade is positively accelerated. On the basis of Sundberg's (31) survey we can safely estimate that, at the very least, the Rorschach is administered to a million persons a year in the United States; it consumes on the average approximately five million clinical man-hours (which is 571 years), at a total cost to the clients of approximately twenty-five million dollars. Thus, in terms of usage the Rorschach is easily the Number One psychological instrument. It has become as closely identified with the clinical psychologist as the stethoscope is with the physician.

The amount of research and publication on the Rorschach is even more impressive. On this count no other test equals it. Over the past decade it has inspired on the average not fewer than three publications per week in the United States alone. The rate of Rorschach publication, also, is positively accelerated. The Rorschach bibliography has already passed 3000.

Of course, it is too much to expect any one person to review and assess in its totality any phenomenon of such fabulous proportions as the Rorschach. The present review focuses attention on the Rorschach literature of the past several years, to determine the degree to which recent research has turned up anything that might in some way alter the negative judgments arrived at by earlier reviewers (e.g. 10).

Much of the early research on the Rorschach has often been criticized for methodological and statistical inadequacy, but this fortunately can no longer be said of the recent research published in the leading psychological journals. There are now a number of methodologically and

statistically sound and sophisticated studies. Even more important, in terms of doing full justice to the Rorschach, is that the good research is now being done by the Rorschachers and projective test experts themselves, often with the full cooperation of their clinical colleagues who are highly experienced in the use of projective techniques. No longer can it be claimed that negative findings are the result of blue-nose methodologists of statistics and experimental psychology applying inappropriate criteria to an instrument for which they have no sympathy, no clinical experience, no intuitive feeling, and no talent.

Detailed reviews of recent Rorschach research have been made by Heiman and Rothney (11) and by Ricciuti (24). A book edited by Rickers-Ovsiankina (25) is probably the most important publication in the field in the past several years and contains excellent discussions of Rorschach research by a number of prominent psychologists in the fields of projective techniques, clinical psychology, and personality research. The reader is also referred to the *Annual Reviews of Psychology* for coverage of the most important contributions; the review by Gleser (9) is especially worthwhile.

#### RORSCHACH TRAINING

The Rorschach is not just another test which the clinician may learn to use by reading a manual. It is a whole culture, the full acquisition of which depends upon intensive tutorial training, a great deal of clinical experience with projective materials, a certain degree of dedicated discipleship, and, perhaps most difficult of all, acclimatization to an atmosphere that is philosophically quite alien to the orientation of modern psychology as it is now represented in the leading American and British universities. In addition, the would-be Rorschacher, if he is to hold his own among the experts, must possess a kind of gift similar to the literary talent of a novelist or biographer, combining a perceptive and intuitive sensitivity to human qualities and the power to express these perceptions in subtle, varied, and complex ways. The Rorschach report of a real expert is, if nothing else, a literary work of art. Indeed, this is the chief criterion of expertness with the Rorschach, for the research has not revealed any significant differences in reliability or validity between beginners and acknowledged masters in the Rorschach technique.

Qualified Rorschachers generally have had at least three semesters, the equivalent of a year and a half, of intensive training in the use of

the Rorschach. The first semester is usually devoted merely to learning how to score the test, while the second and third semesters are devoted to interpretation. As is typical of most textbooks on the Rorschach, there is little or no reference to the research literature in most traditional Rorschach courses. At least one hundred tests must be administered, scored, and interpreted under the close supervision of an expert before the novice is considered sufficiently qualified to be left on his own. Unfortunately, many clinicians, and especially school psychologists, who use the Rorschach in their daily clinical practice are inadequately trained, with the consequence that their reports have a stereotyped, cookbook quality which can add nothing of clinical value to the understanding of the patient and can often be injudiciously misleading or even harmful. It is the reviewer's impression from reading many psychological reports based on the Rorschach, that the acknowledged experts are usually more cautious and wise in their use of the instrument than are clinicians who have had relatively meager training or who are self-taught.

#### USES OF THE RORSCHACH

The technique has been used with all age levels in clinics, guidance centers, hospitals, schools, and in industry, to assess, diagnose, and describe every aspect of the human personality—cognitive, emotional, and motivational—in both normal and psychiatric subjects. In tabulating the types of interpretive statements made from a single Rorschach protocol (analyzed by Klopfer), Shneidman concluded that the Rorschach concentrates on the areas of affect, diagnosis, quality of perception, ego capacity, personality mechanisms, sexual thought and psychosexual level (29). One is impressed after reading a large number of Rorschach reports that no facet of the human psyche and no aspect of human feeling or behavior is inaccessible to the Rorschach. Certainly it excels all other psychological tests in permitting a richness of personality description that comprehends the entire lexicon of human characteristics. It has even been used to attempt to differentiate children with defective hearing from those with normal hearing (11, p. 75). Its chief use, however, remains that of aiding in the formulation of psychiatric diagnosis and prognosis.

The Rorschach has also been used, with questionable success, as a research tool in the investigation of personality and in anthropological and cross-cultural studies. Its contributions in the personality realm have been evaluated by Gardner and Lois Murphy (25), and Lindzey

(17) has written a comprehensive review of its use in cross cultural research. Neither the Murphys nor Lindzey credit the Rorschach with substantial contributions to research in these fields.

#### ADMINISTRATION AND SCORING

The test materials have not changed in 43 years; they are the same ten, bilaterally symmetrical blots originated by Hermann Rorschach. The Rorschach culture apparently has assumed that these ten blots cannot be improved upon and that they alone are a sufficient foundation for building a science of personality diagnosis. The great orthodoxy and appeal to authority in the Rorschach culture is reflected also in the scoring procedures, which have changed in only minor details from the method originally laid down by Rorschach.

Incidentally, if the color in the five chromatic blots plays as important a role as the Rorschachers claim for it, then note should be taken of the fact that different editions of the blots differ in color, some being more vivid and others more pastel.

The scoring of the subject's responses, which generally number between ten and thirty, is a highly technical procedure requiring many hours of practice before it becomes an easy task. The several different scoring systems currently in use are all basically much alike, and once having learned one it is easy to adopt another. The systems of Rorschach and Binder, Rapaport, and Schafer, Beck, Piotrowski, Hertz, and Klopfer have been systematically compared in the last chapter of the volume edited by Rickers-Ovsiankina (25).

#### RORSCHACH INTERPRETATION

Many elements enter into interpretation. First there are the formal scores, which are generally interpreted in terms of configurations or combinations with other Rorschach scores. Textbooks on interpretation are seldom explicit or precise concerning the quantitative aspects of the Rorschach scores and indices, although the language of the discussion clearly implies quantitative considerations. Reference is made to "a lot of shading responses," "a high M per cent," "long reaction time," "many CF responses," and so on. The exact quantity is rarely specified. Examiners must have had experience with at least 100 protocols before developing some subjective notion of the "norms" of the various scores. There are, however, published norms (e.g., 1, 2, 3), but these are seldom referred to by clinicians, and the leading textbooks on Rorschach inter-



pretation make no use of them. Almost every page of the long-awaited and important book on Rorschach interpretation by Piotrowski (23) contains typical examples of the interpretations connected with various scores. For example: "There is something uncompromising, inflexible, and daring about those subjects who give c'R (dark shading responses). By contrast, the individual with many cR (light shading responses) prefers to sacrifice . . . his important goals of external achievement in order to appear less competitive and assertive to the world. If necessary, he surrenders part of his personality rather than antagonize others" (23, p. 264). These elaborate and subtle interpretations of Rorschach scores are totally unsupported by any kind of research evidence.

But much more than the formal scores enters into the interpretation. The subject's language, the content of his responses, the particular sequence of his responses, his reaction time to each card, the way he handles the cards and turns the cards, every aspect of his behavior during the testing—all are grist for the interpretive mill, which grinds extremely fine. The full flavor of this art can be savoured from a number of published Rorschach reports by masters of the technique. The thinking that enters into the interpretation is clearly delineated by Schafer in his excellent text (28) and in the detailed case analysis presented in the textbook by Philips and Smith (22, pp. 267—312). A highly professional report by Stephanie Dudek, typical of the productions of the most skilled Rorschachers, is to be found in the Appendix of the book by Symonds and Jensen (34, pp. 398—400). It is evident that nothing in the Rorschach protocol or in the subject's behavior during the testing is regarded as "noise" in the system—everything is considered significant and interpretable. And the final report of an expert, in its wealth of detail, its subtlety of personality description, breadth of comprehension, and depth of penetration, can often rival the most elaborate characterizations of Marcel Proust or Henry James.

Aside from considerations of reliability and validity, a question must be asked concerning the semantics of the Rorschach report itself. How unambiguously meaningful is the interpretation to a number of different persons reading the final report? Little is factually known about this. It could well be that the Rorschach report is itself projective material for the person to whom the report is referred, serving mainly to bolster his confidence in his own interpretations derived from other sources. The real question is, how much can the report *add* to the psychiatrist's understanding of his patient gained through other means, even assu-

ming it is valid? This we do not know, but the question becomes wholly academic when we take account of the known reliability and validity of Rorschach interpretation.

### RELIABILITY

Few other tests provide so many opportunities for the multiplication of error variance as does the Rorschach. We must consider separately the reliability of scoring and of interpretation, the stability of these in time, the internal consistency of scores, and the effect of the interaction of examiners and subjects.

First, it must be pointed out that most of the traditional Rorschach scores have two strikes against them from a psychometric standpoint. In the typical protocol, most of the scoring categories are used relatively infrequently, so that their reliability is practically indeterminate. For example, the average frequencies of various Rorschach scores in a sample of 28 nonpsychiatric subjects (.4, p. 122) is  $Dd = 1.0$ ,  $S = 0.3$ ,  $M = 2.9$ ,  $k = 0.2$ ,  $K = 0.1$ ,  $FK = 0.6$ ,  $FC = 0.9$ ,  $C = 0.2$ . The only really large frequencies are  $R$  (number of responses)  $= 22.1$ ,  $D$  (large detail)  $= 12.1$ ,  $W$  (whole responses)  $= 8.0$ , and  $F$  (form)  $= 7.5$ . The distribution of these scores are generally very skewed, and the small amount of variation that occurs among the majority of subjects easily falls within the standard error of measurement for most of the scores. By all criteria  $R$  (number of responses) has the highest reliability of any of the scores, and by virtue of this it spuriously inflates the reliability of the various index scores into which it enters, such as  $M\%$ ,  $F\%$ ,  $W\%$ , etc. Most of the combinational scores from the Rorschach, consisting of ratios and differences among the various primary scores, are, of course, even more unsusceptible to a satisfactory demonstration of reliability than are the primary scores.

Another question that is seldom asked is whether the scoring categories themselves have any particular meaning or uniqueness in a psychological sense. That is, are the various movement responses, shading responses, color responses, texture responses, or content of the responses measuring some common factor more or less peculiar to these particular classes of determinants? Factor analyses of the scores indicate that the underlying factors do not coincide at all well with the traditional scoring categories (e.g., 36). Correlations between the various movement responses ( $M$ ,  $FM$ ,  $m$ ) on the Rorschach, Behn-Rorschach, and Levy Movement Cards are in the range from .12 to .41 (21), so that if the

tendency to perceive movement in ambiguous figures is an important and stable characteristic of individuals. as Rorschach theory would have us believe, it is apparent that the Rorschach is unable to demonstrate reliable individual differences in this trait. That is to say, various M responses seem to be highly stimulus-specific. The various color scoring categories have been brought even more seriously into question by experiments using totally achromatic reproductions of the Rorschach blots. In a review of this research Baughman concluded that "... color has little or no effect upon a subject's behavior to the extent that his behavior is represented by the psychogram or similar scoring scales" (5, p. 143). The twenty-five studies of this type reviewed by Baughman lead to the conclusion that "the form or shape of the blot is the only relevant dimension. Certainly color does not appear to affect behavior very much, and if color is ineffective shading seems even less likely to be a significant variable." (5, p. 143). In view of this, how meaningful is an index such as the very important *M: sum C* ratio, which is said to indicate the subject's "experience-type" measured along the dimension of "introversive-extravertive"? The literature on experience-type is reviewed by Singer (25, pp. 223—259), who concludes that after 40 years of the Rorschach nothing yet is known concerning the psychometric or statistical characteristics of the very central experience balance ratio of *M: sum C*.

A word of caution concerning improper estimates of Rorschach reliability: these often consist of reporting the *percentage of agreement* between two or more judges. It should be clear that percentage agreement is not a legitimate measure of reliability and tells us none of the things we want to know when we ask about the reliability of a test. What we want to know is the proportion of variance in the scores that is not error variance. The reliability coefficient tells us this; the percentage agreement does not. The latter measure can often be misleading and should always be discounted as an index of reliability, unless other crucial information is also provided. Take the following fictitious example, in which two judges independently sort a sample of 500 protocols in terms of the presence (+) or absence (—) of indicators of a particular syndrome. Their percentage agreement is 98 per cent — impressively high. When reliability is obtained in the proper way, however, by determining the correlation between the two judges, the reliability coefficient turns out to be only .19.

		Judge A		Total
Judge B	+	1	4	5
	—	4	491	495
Total		5	495	500

The writer has presented a detailed discussion of the reliability of Rorschach scores elsewhere (15, pp. 11—17), and a more recent consideration of the whole reliability problem has been presented by Holzberg (25). Some of the conclusions may be summarized briefly.

*Scoring reliability* per se has been determined very seldom. The few instances reported in the literature constitute the highest reliabilities to be found for any aspect of the Rorschach. Reliability of scoring depends to a large extent upon the degree of similarity of the training of the scorers and has been reported as ranging from .64 to .91.

*Split-half reliability* has always been frowned upon by Rorschachians as inappropriate. Nevertheless, split-half estimates have yielded comparatively high reliabilities, ranging in one study (35) from .33 (F 1 %) to .91 (R), with an average reliability coefficient of .54 (corrected by the Spearman-Brown formula). In another study (13) an odd-even split of the cards for 100 subjects yielded an average reliability for 20 Rorschach scores of .83, with a range from .67 to .97.

*Test-retest reliability* ranges from about .10 to about .90, depending largely upon the test-retest interval and the particular score. For a two-weeks interval the reliabilities of various scores range between .60 and .80 (15, pp. 12—13). The most extensive determination of retest reliability is that of Epstein *et al.*, (7), who gave the Rorschach to 16 college students a total of ten times over a period of five weeks. The average reliabilities for various response categories ranged from .29 to .56.

*Parallel forms reliability* has been determined by use of the Behn-Rorschach, a set of similar blots which seem to meet all the psychometric criteria for qualifying as an equivalent form of the Rorschach. For 35 scoring categories the means and standard deviations of the Behn and the Rorschach do not differ significantly in normal and psychiatric populations and the two forms seem to correlate as highly with each other as each correlates with itself. The correlations for various scores range from about zero to .86, with a mean around .60.



*Examiner and situational influences* have been increasingly recognized in recent research as significant contributors to the variance of Rorschach scores (e.g., 4, 8, 12, 20, 27). The subject-examiner interaction is certainly one of the most important aspects of the test. The effect of the setting in which the test is taken and the fact that different examiners consistently elicit different amounts of various scored determinants from subjects should make it imperative that future Rorschach studies be based upon a representative sampling of examiners as well as of subjects.

*Reliability of interpretation* is, of course, the most important matter of all. It may be stated as a general principle that the most crucial reliability is that of the end-product of the test, which, in the case of the Rorschach, usually consists of a verbal description of personality characteristics based on a global evaluation of all aspects of the subject's protocol. Contrary to the usual claim of Rorschachers that this global interpretation is more reliable or more valid than any of the elements on which it is based, such as the scores and the various derived combinations and indices, a systematic search of the literature has not turned up a single instance where the overall interpretation was more reliable than the separate elements entering into it. Rorschach textbooks have not presented any evidence of satisfactory reliability of the final product of the test and the reviewer has not been able to find any such evidence in the research literature.

Here are some typical examples of what has been found. Lisansky (18) had six highly qualified Rorschachers rate 40 subjects on ten personality items which they agreed could be confidently assessed from the Rorschach protocol. To make the experiment similar to clinical conditions the Rorschachers were provided also with an abstract of each patient's history. The degree of agreement between the judges was measured by the phi coefficient, which averaged .33. Six other clinicians rated the same traits on the basis of the case history abstracts alone, with an average phi of .31, which is not significantly different from the reliability of the clinicians who were aided by the Rorschach. The interesting point is that the ten rated personality items were specially selected as being the kinds of questions which the Rorschach, and not particularly the case history, is supposed to be able to answer.

Korner and Westwood (16) had three clinical psychologists, qualified in the use of the Rorschach, sort the protocols of 96 college freshmen

into three categories for level of personality adjustment. The average correlation among the three judges was .31.

Datel and Gengerelli (6) found that when 27 Rorschachers were required to match personality interpretations written by each other on the basis of the protocols of six subjects (presented for matching in sets of six), there were more mismatches than correct matchings. Of the total of 324 discrete matchings, 148 were correct and 176 were incorrect. Despite the fact that the subjects from whom the protocols were obtained differed greatly from one another in nosology, etc., the average reliability for the individual clinicians was not significantly greater than zero.

The most careful and methodologically sophisticated study of Rorschach reliability and validity has been carried out by two leading projective test experts, Little and Shneidman (19). The editors of the *Journal of Projective Techniques* chose twelve distinguished Rorschach experts—all are eminent teachers and writers in this field—to participate in the study. Rorschach protocols were obtained from 12 patients, three each from the psychotic, neurotic, psychosomatic, and psychiatrically normal diagnostic categories. The Rorschach judges were each provided with one protocol from each of the four categories and asked to perform the following interpretive tasks: assign diagnostic labels, rate the subject for personality adjustment (on a scale from 0 to 8), answer 100 True-False factual items taken from the case histories of the subjects, answer 117 True-False personality items typical of those contained in psychological reports, and perform a Q-sort of 76 items typical of the kinds of statements made in Rorschach interpretations. The reliability estimate of the diagnostic labeling consisted of having four other judges rate degree of similarity of diagnosis among pairs of the Rorschach judges on a six-point scale (0—5). The mean rating among all the Rorschach judges was 2.50, which led the authors to conclude that “diagnostic labels based upon blind analyses of protocols may be quite wide of the mark and the present analysis indicates that the judges may not be even shooting at the same target” (19, p. 11). The method of treating the ratings of maladjustment makes it difficult to obtain an estimate of inter-rater reliability, but it is interesting that the non-psychiatric patients were rated as considerably more pathological on the basis of their Rorschachs (as well as on three other clinical tests of personality) than when they were rated solely on the basis of anamnestic data. (This tendency for Rorschach interpretations to be

excessively biased toward the pathological has been well-known from earlier studies; a good illustration of the tendency may be found in the Rorschach analyses of 28 non-psychiatric subjects reported in great detail by Symonds and Jensen (4, 119—169.) The True-False factual and personality items were correlated with outside criteria and therefore will be discussed in the section on validity. The Q-sort yielded the most easily interpretable index of inter-judge reliability. The correlations between the judges' Q-sorts for the 12 patients range from  $-.13$  to  $.64$ , with a mean of  $.31$ . It is instructive to note that when the Q-sorts of each set of four subjects rated by the same judge are inter-correlated, the mean correlation is  $.27$ , which is not significantly different from the *inter-judge* reliability of  $.31$ . In other words, at least as much of the variance in Rorschach interpretations is attributable to differences among the interpreters as to differences among the subjects. Little and Shneidman concluded, "Test interpreters tend to make their interpretations in a stereotyped manner independent of the subject."

How well did each interpreter agree with himself? To find out, the investigators had the judges perform the same interpretive tasks on the same protocols just ten days later and intercorrelated the ratings of the first occasion with those of the second. Only those results which can be reported in terms of a correlation coefficient are reported here. For the factual True-False items the average correlation is  $.74$ ; for the Q-sorts the correlations range from  $.26$  to  $.81$ , with a mean of  $.61$ .

Silverman (30) carried out a somewhat more detailed study of Rorschach reliability and validity, using the Q-sort. The judges were selected in terms of amount of training and clinical experience with projective techniques, including the Rorschach. There were 10 noted projective test experts, 10 clinicians with 5 to 8 years of experience in projective testing, and 10 clinicians with fewer than three years of experience in projective testing. The Rorschach, Thematic Apperception Test, House-Tree-Person test, and the Most Unpleasant Concept test, were obtained from 10 adult males undergoing psychotherapy. There were six separate Q-sorts for different areas of interpretation. The 180 Q-sort items were typical of the statements found in Rorschach and projective reports. The reliabilities, as estimated from the correlations among the Q-sorts, were: Defenses =  $.27$ , Motivating Needs and Affects =  $.25$ , Character Traits =  $.44$ , Diagnosis and Symptoms =  $.44$ , Interpersonal Behavior =  $.21$ . The overall reliability was  $.34$ . The degree of reliability was unrelated to the amount of experience of the judges:

there was no higher agreement among the most experienced clinicians than among the least experienced.

One recent study (14) strongly stacked the cards in favor of maximizing the reliability by selecting seven clinicians who had very similar orientations toward the use and interpretation of psychological tests and ten subjects who were very heterogeneous in pathology. The clinicians' task was to rank 10 psychological needs as to their relative importance for each of the ten subjects. The inter-rater reliability was .12. (When the same task was performed with the TAT and a Sentence Completion test, the reliabilities were .14 and .30, respectively.)

### VALIDITY

Considering the reliability of the Rorschach, its poor validity would seem to be a foregone conclusion. However, though it is axiomatic in psychometric theory that the validity of a test cannot be higher than the square-root of its reliability, it has often been claimed that the Rorschach (as well as other projective tests) is exempt from this general rule. Therefore a study of the evidence for the validity of the Rorschach might be worth while.

Guilford succinctly reviewed the status of Rorschach validity up to 1959 and came to the following conclusions: "In spite of the widespread popularity and use of the Rorschach ink blots, the reliabilities of scores tend to be relatively low, and validities, although quite varied, are generally near zero. This statement regarding validity applies to use of the instrument in discriminating pathological from normal individuals, for diagnosis of more particular pathologies such as anxiety, for indicating degrees of maladjustment in the general population, and for predicting academic and vocational success" (10, p. 313).

The most recent comprehensive review and discussion of Rorschach validity is the chapter by Harris (25, pp. 380-439). It is the most thoughtful and objective article on this subject the present writer has encountered. From his extensive survey, Harris concluded: "By the canons of test analysis, the Rorschach technique as a whole has been shown at present to have neither satisfactory validity nor invalidity" (p. 436). Predicting the future of Rorschach research, Harris states, "There is very little concrete basis for making an optimistic prediction that a review of studies of validity, in which the ten Rorschach cards have served as the sole instrument of investigation, will be any different



25 years from now than they were when reviewed in 1954 by Ainsworth . . ." (p. 424)<sup>1</sup>.

What, specifically, have the most recent studies found?

First, a distinction must be made between experimental and clinical types of validation studies. In experimental studies, particular Rorschach scores (often scores that have been specially derived for the particular study) are in some way tested for their correlation with some non-Rorschach criterion. The criterion may or may not be of clinical relevance. Clinical validation studies, on the other hand, involve a more global use of the Rorschach protocol, typical of its use in clinical practice, with the aim of testing the correlation of the Rorschach with various clinically relevant criteria. Older reviews of Rorschach validity are based predominantly on the experimental type of study. Recent research has concentrated more on the clinical validity of the instrument as it is typically used by clinical psychologists. Many of the experimental type of studies have been reviewed by Zubin (37). The fact that some of these studies have reported validity coefficients which, when significant at all, are generally in the range of .20 to .40, cannot be interpreted as supporting the clinical usefulness of the test. Aside from the fact that validity in this range is practically useless for individual assessment, the validated "scores" are often not those used by the clinician or they are used in a different way. Even when the scores do happen to be those that enter into the clinical interpretation of the protocol, such as the M per cent, clinicians seldom heed the experimental findings. It is easy to find statements in current clinical reports that a subject is "creative" on the basis of a high M per cent in his protocol, despite the well-known failure of this relationship to be borne out in research studies which are seemingly ideal for capturing it (e.g., 26).

Beck's z and g scores, characterized as an "organizational factor", are derived scores which have gained popularity in clinical use. These scores are a systematic weighted combination of Rorschach attributes claimed to be indicative of intelligence and efficiency of intellectual functioning. The g score does have some validity, showing correlations with psychometrically measured intelligence in the range of .20 to .25 (25, p. 36).

Another special scoring method has been devised by Holt (25), to measure degree of adaptive versus maladaptive regressive tendencies.

---

<sup>1</sup> Ainsworth, Mary D., Problems of validation. In: Klopfer, B. et al. *Developments in the Rorschach technique*. Vol. I. Yonkers: World Book Co., 1954.

When this index was correlated with 55 items of various behavioral and personality test criteria, 20 of the correlations were significant beyond the .10 level. The mean of the correlations significant beyond the .10 level was .59. Crossvalidation of such studies generally loses many of the formerly significant correlations, and no such correlations should be accepted without evidence of cross-validation. For example, Holt states concerning the validity of his regression score, "The correlation coefficients are not impressively large, for the most part not even being highly significant, but they are in the right directions. A word of caution, however: Incomplete but largely negative preliminary results from a group of college girls of the same age (as the college men on whom the original correlations were obtained) suggest that these correlations may not hold up in different samples, but may, in some as yet unknown way, be specific to unknown parameters of the present group of college boys" (25, p. 314).

It seems safe to conclude that experimental studies of particular Rorschach attributes have been able to show statistically significant correlations with other psychological criteria. These correlations have been generally rather low (i.e., between .20 and .40), only rarely exceeding .50, and most such correlations have not stood the test of cross-validation.

How valid is the Rorschach when it is used as a clinical instrument by acknowledged experts? Three recent studies, which have taken care to avoid the criticism that the obtained validity coefficients do not represent the validity of the Rorschach when used by experts, are instructive.

The study by Little and Shneidman (19), which has already been described in the section on Reliability, used 12 Rorschach experts, who were selected by the editors of the *Journal of Projective Techniques* and whose names are given in the Appendix of the published monograph. The Rorschach protocols were obtained from 12 patients equally divided among the psychiatrically normal, neurotic, psychosomatic, and psychotic categories. The various criteria against which validation was attempted were obtained from the pooled judgments of 23 psychiatrists and one psychologist on the basis of a comprehensive psychiatric case history on each patient, obtained by one psychiatrist in 4 to 8 interviews of 1 to 3 hours duration. On a True-False questionnaire of 117 personality items typical of those in Rorschach reports, the correlation between the Rorschach judges and the anamnestic judges ranged from —.26 to

.74, with a mean of .37. With a True-False questionnaire of 100 factual items which could be verified from the case history, the Rorschach correlations ranged from  $-.12$  to  $.42$ , with a mean of  $.14$ . The correlations between a Q-sort of personality items obtained from the Rorschach judges and from the anamnestic judges ranged from  $-.10$  to  $.47$ , with a mean of  $.17$ . This validity coefficient becomes  $.21$  when corrected for attenuation of the criterion. But as compared with the other psychological tests used in the Little and Shneidman study (the Make-A-Picture-Story test, the Thematic Apperception Test, and the Minnesota Multiphasic Personality Inventory), the Rorschach is not much worse. The MMPI, for example, which made a consistently better showing than any of the projective techniques, had an overall Q-sort validity of  $.33$  (corrected for attenuation).

The study by Silverman (30), described in the section on reliability, compared Q-sorts of projective test experts with Q-sorts performed by the therapists of 10 adult males after 35 hours of psychotherapy. There were six-Q-sorts made up of typical Rorschach report items covering the areas of Defenses, Motivating Needs and Affects, Character Traits, Diagnosis and Symptoms, Interpersonal Behavior, and Infancy and Childhood Perceptions of Parental Figures. The validity coefficients for these areas range from  $.12$  to  $.50$ , with a mean of  $.29$ .

On the basis of a preliminary study (33) in which the Rorschach protocol of a patient in psychotherapy was sent to 12 Rorschach experts for independent interpretations, the one expert with the largest percentage of "hits" in agreement with the psychotherapist's knowledge of the patient was selected to perform Rorschach analyses of 28 non-psychiatric subjects who were also assessed by interviews and other tests. A detailed account of the Rorschach analyses is presented by Symonds and Jensen (34, pp. 119—169). The Rorschach expert was asked to rank the subjects for overall personality adjustment on the basis of her analysis of the Rorschach protocols. As the criterion two psychologists performed the same task from anamnestic data and from direct impressions gained in several hours of interview with each subject. The correlation between the Rorschach ratings and the criterion, corrected for attenuation, was  $.34$ . One could argue that the criterion itself had little validity, but this points up one of the crucial problems of Rorschach interpretation: are the test interpreter and the person to whom the interpretation is addressed both speaking the same language? If not, of what value is the Rorschach report? Most psychiatrists receiving psychological reports

based in whole or in part on the Rorschach, it should be remembered, have not been trained in Rorschach interpretation.

The use of the Rorschach in vocational psychology has been reviewed by Super and Crites (32), who conclude that ". . . too little is now known to justify its use in practical counseling or personnel work" (p. 575). Similarly, Ricciuti has concluded a recent review of this subject as follows: "The practical usefulness of projective techniques in predicting educational or industrial criteria continues to be small" (24, p. 74).

#### SUMMARY

Recent research on the Rorschach has not brought forth any substantial evidence that would alter the negative evaluations of earlier reviewers. If anything, recent studies add support to the conclusion that the Rorschach as a clinical instrument has too inadequate reliability and too meagre validity, even in the hands of the most expert, to justify any claims for its practical usefulness. The strong bias toward pathology in Rorschach reports on non-psychiatric subjects can lead to harmful consequences in non-psychiatric settings, such as in schools and in industry. Even in cases where harm might not result, one must weight the scant validity of the test against the fact that of all psychological assessment techniques it is the most time consuming and requires by far the most extensive training of its practitioners. Many psychologists who have looked into the matter are agreed that the 40 years of massive effort which has been lavished on the Rorschach technique has proven unfruitful, at least so far as the development of a useful psychological test is concerned.

Until proponents of the Rorschach can produce evidence which substantially contradicts this verdict—and thus far such evidence is conspicuously lacking in the Rorschach textbooks—it seems not unreasonable to recommend that the Rorschach be altogether abandoned in clinical practice, and that students of clinical psychology not be required to waste their time learning the technique.

The question of why the Rorschach still has so many devotees and continues to be so widely used is quite another problem and is beyond the scope of this review. A satisfactory explanation of the whole amazing phenomenon is a task for future historians of psychology and will probably have to wait upon greater knowledge of the psychology of credulity than we now possess. Meanwhile, the rate of scientific progress in clinical psychology might well be measured by the speed and thoroughness with which it gets over the Rorschach.

#### REFERENCES

1. Ames, Louise B. *et al.*, *Child Rorschach responses: developmental trends from two to ten years*. New York: Hoeber, 1952.
2. ———, *Rorschach responses in old age*. New York: Hoeber, 1954.
3. ———, *Adolescent Rorschach responses: developmental trends from ten to sixteen years*. New York: Hoeber, 1959.



4. Baughman, E. E., Rorschach scores as a function of examiner difference. *J. proj. Techn.* 1951, 15, 243—249.
5. ———, The role of the stimulus in Rorschach responses. *Psychol. Bull.* 1958, 55, 121—147.
6. Dattel, W. E., and Gengerelli, J. A., Reliability of Rorschach interpretations. *J. proj. Tech.* 1955, 19, 372—382.
7. Epstein, S., Nelson, Jane V., and Tanofsky, R., Responses to inkblots as measures of individual differences. *J. consult. Psychol.* 1957, 21, 211—215.
8. Gibby, R. G., Miller, D. R., and Walker, E. L., The examiner's influence on the Rorschach protocol. *J. consult. Psychol.* 1953, 17, 425—428.
9. Gleser, Goldine G., Projective methodologies. *Ann. Rev. Psychol.* 1963, 14, 391—422.
10. Guilford, J. P., *Personality*. New York: McGraw-Hill, 1959.
11. Heiman, R. A., and Rothney, J. W. M., Development and applications of projective techniques. *Rev. educ. Res.* 1959, 29, 73—83.
12. Henry, Edith M., and Rotter, J. B., Situational influences on Rorschach responses. *J. consult. Psychol.* 1956, 20, 457—462.
13. Hertz, Marguerite R., The reliability of the Rorschach inkblot test. *J. appl. Psychol.* 1934, 18, 461—477.
14. Howard, K. L., The convergent and discriminant validation of ipsative ratings from three projective techniques. *J. clin. Psychol.* 1962, 18, 183—188.
15. Jensen, A. R., The reliability of projective techniques: Review of the literature. *Acta Psychol.* 1959, 16, 3—31.
16. Korner, I. N., and Westwood, D., Inter-rater agreement in judging student adjustment from projective tests. *J. clin. Psychol.* 1955, 11, 167—170.
17. Lindzey, G., *Projective techniques and cross-cultural research*. New York: Appleton-Century-Crofts, 1961.
18. Lisansky, E. S., The inter-examiner reliability of the Rorschach test. *J. proj. Tech.* 1956, 20, 310—317.
19. Little, K. B., and Shneidman, E. S., Congruencies among interpretations of psychological test and anamnestic data. *Psychol. Monogr.* 1959, 73, No. 6 (Whole No. 476).
20. Lord, Edith, Experimentally induced variations in Rorschach performance. *Psychol. Monogr.* 1950, 54, No. 316.
21. Okarski, J. F., Consistency of projective movement responses. *Psychol. Monogr.* 1958, 72, No. 6 (Whole No. 459).
22. Phillips, L., and Smith, J. G., *Rorschach interpretation: advanced technique*. New York: Grune & Stratton, 1953.
23. Piotrowski, Z. A., *Perceptanalysis*. New York: Macmillan, 1957.
24. Ricciuti, H., Development and application of projective tests of personality. *Rev. educ. Res.* 1962, 32, 64—77.
25. Rickers-Ovsiankina, Maria A. (Ed.), *Rorschach psychology*. New York: Wiley & Sons Inc., 1960.

26. Roe, Anne, Painting and personality. *Rorschach Res. Exch.* 1946, 10, 86—100.
27. Sanders, R., and Cleveland, S. E., The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.* 1953, 17, 34—50.
28. Schafer, R., *Psychoanalytic interpretation in Rorschach testing: theory and application*. New York: Grune & Stratton, 1954.
29. Shneidman, E. S., In Klopfer (Ed.), *Developments in the Rorschach technique*. Yonkers: World Book Co., 1956, pp. 595—642.
30. Silverman, L. H., A Q-sort study of the validity of evaluations made from projective techniques. *Psychol. Monogr.* 1959, 73, No. 7 (Whole No. 477).
31. Sundberg, N. D., The practice of psychological testing in clinical services in the United States. *Amer. Psychol.* 1961, 16, 79—83.
32. Super, D. E., and Crites, J. O., *Appraising vocational fitness by means of psychological tests*. (Revised edition) New York: Harper & Bros., 1962.
33. Symonds, P. M., A contribution to our knowledge of the validity of the Rorschach. *J. proj. Tech.* 1955, 19, 152—162.
34. ———, and Jensen, A. R., *From adolescent to adult*. New York: Columbia Univ. Press, 1961.
35. Vernon, P. E., The Rorschach inkblot test. *Brit. J. med. Psychol.* 1933, 13, 89—118 & 179—200.
36. Wittenborn, J. R., A factor analysis of Rorschach scoring categories. *J. consult. Psychol.* 1950, 14, 261—267.
37. Zubin, J., Failures of the Rorschach technique. *J. proj. Tech.* 1954, 18, 303—315.