Failures of the Rorschach Technique

JOSEPH ZUBIN, PH.D.
New York Psychiatric Institute, Columbia University

If Herman Rorschach had not been carried off by the infection following a minor operation, he would have celebrated his 68th birthday last November 8th. He would have looked approvingly on our efforts and would have helped make this symposium on failures a success by providing methods for their elimination in the future. For no one was more tentative in his conclusions, more demanding of continual self-survey and revision than the man who lent his name to the major projective technique. "The conclusions drawn," he says, "are to be regarded more as observations (remarks) than as theoretical deductions. The theoretical foundation for the experiment is, for the most part, still incomplete." He even went as far as hoping for "...control experiments taking up each symptom individually, and other psychological methods which might also be used in control research." (24)

It is both a brave and a wise move that the Society for Projective Techniques has undertaken in this symposium and it is to be congratulated on its maturity and integrity. Like the post-mortem in surgery, failures in projective techniques ought to be far more revealing and instructive than the reports of successful cases. We stand to benefit from our mistakes, find out what changes need to be introduced, what hypotheses need to be altered and what expectations need to be amended. Following the custom of the surgeons, only the outstanding failures of the Rorschach technique will be discussed, the outstanding successes being omitted. A cataloguing of one's failures without scanning the successes is trying to even the most mature of souls, but since it is motivated by a search for improvement we can all bear up under it. After analyzing the failures, a hypothesis to explain them will be presented, which may perhaps repay, in part, for the masochistic trend a self-analysis may engender.

One outstanding Rorschach worker has made the following surmise about Rorschach's own reaction to the current scene: "Rorschach today would still recognize the cards. Brilliant though he was, I doubt if he could find time to read the voluminous literature which is well enroute to the thousand mark. I am certain that, if he could, he would be startled that from his little experiment following the 'scoring' of the responses, there emerge invaluable facts relating specifically to the way in which the patient sees his world, approaches and handles it, and of what this world consists. His anxieties and insecurities, his hurts and wishes, his fictions, his needs, his assets and liabilities, his likes and dislikes—all of these and more emerge to be viewed by the examiner. Moreover, the pattern reveals also the meaning of these things to him, the configuration of his personality which thus results, and the motivations of his behavior. It, furthermore, aids in differential diagnosis, particular-
ly between the organic and functional types of illness, and among the affect and content disorders. The expert examiner can also obtain from the response record a practical estimation of such important personality features as intellectual efficiency, emotional maturity and balance, and degree and depth of reality acceptance. Finally, the procedure serves as a guide to therapy and an index of its success or failure." (19)

What validity do these claims have? Let us take a look at the record. The following questions need to be answered: 1) What reliability does the Rorschach technique possess? 2) What validity? 3) What relationship does it bear to the changes brought on by therapy, and 4) to the outcome of therapy, and 5) what light has recent research cast on these problems?

First, what about the reliability of scoring? This question has received but little attention and it is generally taken for granted that scoring has a high degree of reliability. Hertz, (18), however, has stated on the basis of her long experience that “scoring still remains a matter of skill—art, if you will.” Though this statement was made 18 years ago, it still largely holds true today. Ramzy and Pickard (22) found that only after considerable discussion and arbitrary acceptance of certain conventions were they able to obtain consistency in their scoring. It is noteworthy that appeal to text-books only increased their confusion. Even after this confusion, the degree of agreement was only 90% for the location category. Since Beck’s location tabulations were followed, it is surprising that the degree of agreement was not perfect. For determinants of Form and the Movement variety, the agreement dropped to 83%, and for the Color and Shading determinants to 75%. As far as content was concerned, following Beck’s classifications an agreement of 99% was obtained. Ames and her associates report their reliabilities in terms of product moment correlation coefficients: location categories, .92; form and movement categories, .90; color and shading, etc., .80, and content categories, .97 (1). These results are essentially in keeping with those reported by Ramzy and Pickard. Since according to Rorschach, “The actual content of the interpretations comes into consideration only secondarily” the determinants being of major importance, it is clear that the best estimate of agreement for the major scoring categories is only 80%. This is a far cry from the degree of agreement expected of an individual test. The lack of objectivity in scoring is further evidenced by Baughman’s experiment (5) in which Veterans Administration examiners disagreed significantly in scoring on 16 out of 22 scoring categories in records based on a random selection of cases culled from their files. It is clear that failure to provide an objective scoring system is our first failure.

The reliability of the test cannot be judged by split-half methods because of the heterogeneity of the blots, nor can it be judged by test-retest because of the memory factor. Alternate forms are required for this purpose. Eichler (12) reports that he tried to obtain a reliability estimate by correlating the Behn-Rorschach with the Rorschach but obtained such low reliabilities that he concluded the two forms were similar but not parallel. Thus, lack of reliability is the second failure.

The validity of the test will be analyzed from the following points:
of view: 1) subjective, 2) clinical, 3) statistical, and 4) experimental.

Subjective and Clinical

Subjective validation of the testimonial variety in which he who comes to scoff remains to pray, will not be commented on further. This type of evidence is so clearly discountable as utterly unscientific, that it is not even to be counted among our failures.

The clinical method sometimes consists of administering and scoring the test, collecting data on the subjects' subsequent behavior and then going back to the protocol, in which are "found" signs which "unmistakably" foretell such behavior. Unless a cross validation of these signs is undertaken in another study, it is fruitless to accept them as indicative of future behavior, because with sufficient imagination and exertion of effort through trial and error, pseudosignificant signs can be found in any test. Unfortunately, such cross validation is rarely encountered.

Blind analysis is one of the spectacular aspects of the Rorschach technique and has probably been the most important factor in the acceptance of the Rorschach. One would wish that this method could be made more explicit and more public, and that the enthusiastic proponents of this method were as ready to publish their failures as their successes. Until this method becomes more open to public scrutiny, it has to be placed in the doubtful category and counted neither as a success nor as a failure.

The matching technique is another way of demonstrating validity. Unfortunately, there are many inadvertent and tangential characteristics in this method, not germane to validity, which may influence the outcome. Successful matching is frequently effected on the basis of minor details or coincidences, rather than essential equivalence. Heterogeneity of matches also makes the task too easy. Determination of the precise ground on which successful pairing is made is virtually impossible. Furthermore, most of the results indicate only that the matching is better than chance, an insufficient criterion for validity. Cronbach (9) has recently devised a trenchant methodology for freeing the matching methods from these defects, but it is quite intricate, and the one application which he made yielded no success in the matching process. Thus, the clinical evidence for validity cannot be accepted scientifically, even though it is impressive. Our failure to provide more cogent evidence for clinical validity must be regarded as our third failure.

Statistical Studies

In view of Rorschach's original purpose of devising a diagnostic test for mental disorders—especially for schizophrenia—we shall first review studies that deal with the diagnostic efficacy of the test. There are very few studies in which clinical scoring and interpretation disagree markedly with clinical diagnosis when the study is conducted in the same clinic and when the two diagnosticians have had considerable experience in working together. This is certainly highly in favor of the test, and the only question a carping critic might ask is: is there a tendency for collusion to occur in such cases, since only a few authors, such as Benjamin and Ebaugh (4), point out the care they took to avoid collusion. When we examine the relationship between the individual Rorschach scores and diagnosis, a totally different picture emerges. Guilford (16) found in 3 successive samples of
about 50 neurotic patients, who were given the Rorschach in orthodox fashion, that no significant differences could be detected between their performance and that of a large normative group of cadets. Wittenborn and Holzberg (32) found zero correlation between Rorschach factors and diagnosis in 199 successive admissions. Cox (9) found only 5 scores out of a total of 45 scores differential between normal and neurotic children, and of these 5, 3 were in the content categories and only 2 in the determinant categories. These are only samples of the well-known failure of the individual Rorschach scoring categories to relate to diagnosis, which is in marked contrast to the success of the global evaluation claimed by clinical workers generally. This must be regarded as the fourth failure.

In prognosis too, the recent review by Windle (30) leaves one with very little faith in the efficacy of the prediction attributable to the Rorschach. The only successful prognostic elements seem to be based on content rather than formal factors as shown by McCall (21). This is the fifth failure.

Since Windle's article was published, two additional bits of evidence of failure of the Rorschach in the prognostic sphere have appeared. In Barron's study (2) in which the Rorschach, together with several other tests including the MMPI were given to both patients and therapists before the beginning of therapy, while the MMPI predicted outcome significantly, the Rorschach, despite all attempts ranging from the global to the atomic, failed to do so. Rogers, Knauss, and Hammond (23) report a similar experience. As long as other tests failed to predict outcome one might have attributed the failure to the heterogeneous nature of the patient group—to an admixture of early and chronic cases, for example. When other tests succeed where the Rorschach fails, one can either conclude that the Rorschach is unsuited to prediction, or that basic personality which the Rorschach claims to measure is unrelated to the type of therapy involved.

The differentiation of organic from functional conditions has also had a checkered history of success and failure: success when viewed retrospectively but failure when the results of retrospective analysis were applied to a new sample. The latest in the series is the study by Dörken and Kral (11), who after demolishing the signs of previous workers propose a new set of their own, which will probably in turn be demolished by the next worker. The vitality of this search for signs despite so many failures can only command the awe and respect of the onlooker. Whether it will finally succeed only time can tell; meanwhile it must be placed in the doubtful category.

The success that clinicians have had in global evaluation of mental patients has not been duplicated in the global evaluations of normals. Here the series of failures is truly appalling. The story is too long to review. Some of the recent examples are—Grant, Ives, and Ranzoni (15), who found zero correlation between Rorschach evaluations and case history evaluations of adjustments in 18-year-old normals. The more specific adjustment signs provided by Helen Davidson (10), fared a little better, yielding correlations from .25 to .56. The failure of the Rorschach to serve as a predictor of success in the screening programs of the armed forces, in the screening of clinical psychology students and students of psychiatry, is too well known to warrant fur-
ther comment. I shall quote only from one of these:

"It was regarded as very important that the Rorschach test should be given full opportunity to show what it had to offer in a personnel-selection setting. It was recognized that neither time nor personnel requirements for the routine administration and use of this test were consistent with the mass testing required... Yet the test was administered experimentally to several hundred students individually according to the prescribed procedures by members of the Rorschach Institute who were serving in one of the psychological units. Two methods of group administration were also tried, the Harrower-Erickson and our own version.

"The results were almost entirely negative. From the individual administration of the test, neither the 25 indicators taken separately or collectively nor the intuitive prediction of the examiner based upon the data he had from the administration of the test gave significant indications of validity against the pass-fail criterion. There were two samples, one of nearly 300 and the other of nearly 200. The Harrower-Erickson group-administration form also gave no evidence of being valid for pilot selection. The AAF group-administration form when scored for the number of most popular responses showed a coefficient of .24, based upon a sample of more than 600 students." (16). The inability to differentiate between normals is the sixth failure.

Special studies aimed at evaluating intelligence by means of the Rorschach also usually come-a-cropper. Wittenborn (35) compared the extreme groups selected on the basis of college-entrance examinations on 18 scores of the Rorschach and failed to find any significant relationship. (This is even worse than chance, because by chance you might have expected about one of these 18 scores to show significance on the .05 level.) The relationship of Movement to intelligence has been investigated by Tucker (28), in 108 neurotics, who found a very low correlation—.26. Wilson (31) made a more extensive study of a large college population and used Movement, Form level, Whole, Responses, Z (organization), diversity of content, and a new specially-designed-variable designated as 'specification'—and found zero correlation with intelligence.

The studies in creative ability conducted in our own laboratory on creative vs. non-creative writers, mathematical statisticians, and high school students have failed to reveal any differences on Rorschach performance and even tests especially designed to elicit Movement have failed (34). This is the seventh failure.

But the story of the use of the Rorschach with normals is not entirely a hopeless one. Sen (26), in England, one of Cyril Burt's students, applied the Rorschach to 100 Indian students who had lived together for at least two years. Scoring by means of Beck's scoring system, the correlations with personality evaluations by their colleagues were non-significant. However, when scored for content a la Burt, the correlations ranged from .57 to .66. When matching was resorted to, a global method, both scoring methods yielded a high degree of success: .85 for Beck's system, and even higher for Burt's system. Interestingly enough, however, when a factor analysis was performed on the Beck Scores and on the Burt Scores, the results of both analyses are equally trenchant in their relationship between the derived factor scores and personality. This is a
general finding in the studies in which the raw Rorschach scores failed to relate to diagnosis or personality. When factor analysis is resorted to, a rotation of the factors usually permits certain striking correlations with behavior to emerge.

We might stop here a moment to differentiate between content analysis as used by Burt and in our own laboratory and the content category as used by Rorschach. In the Rorschach there are really 3 types of classifications: Location, Determinants and Content. The Location and Determinant categories are usually spoken of as the Formal Categories, and the Content Categories are those which simply classify the percepts as animal, vegetable or mineral—so to speak—that is, the category of objects it belongs to. It might be better to contrast in the Rorschach the perceptual factors—or structural factors with the content categories. There are left reaction time, popular responses, and confabulations, contaminations, etc.—which are neither determinants nor locations and hence could be classified with content. It is the non-perceptual part of the Rorschach performance—the thought content—which we designate as the content aspects of Rorschach performance.

Examples of the scales used for content analysis of the Rorschach protocols are: 1) Formal content a la Rorschach 2) Dynamic content—(a) degree of evaluation included in response as judged by qualifying adjectives, (b) degree of dehumanization, (c) ascendance-submission in concepts portrayed (slaves-versus-kings, for example), (d) definiteness of concept (e) abstractness (f) dynamic qualities—alive or dead, static or moving, (g) distance in time

1 Whether reaction time belongs here is debatable.

and space (h) self-reference, (i) perseveration, (j) elaboration (k) blot vs. concept dominance, (l) interpretive attitude, etc. (33). Further evidence for the success of this type of analysis is found in Elizur (15) and in Watkins and Stauffacher (29) in this country and Sandler (25) in England.

Elizur found that an analysis of content in relation to hostility yielded significant correlations with ratings of hostility. Sandler, working with Rorschach's content categories, (and not with the type of content analysis being discussed here) made a factor analysis of the content scores of 50 psychiatric patients at Maudsley Hospital, ranging over 8 types of mental disorder. He emerged with 4 factors and determined the psychological meaning of each factor by its correlation with the personality evaluation made by psychiatric interview and case history methods. These were drawn from three levels—previous personality, general background data, and present symptoms. The productivity factor, R, for example, was highly related to previous productivity in life, to chronicity of symptoms and to a schizo-affective picture at time of hospitalization. The Anatomy factor—internal anatomical objects vs. external objects as another example, was related to an insecure, withdrawn, "previous" personality picture, bad physical health and an emotional deflated state for the "present symptoms." The remaining factors were analyzed in similar fashion.

Watkins and Stauffacher (29) provided a series of indexes of "deviant verbalizations" based on the content of the protocols and found that such indicators had a reliability of .77 between two raters, and that these indexes distinguished normals from neurotics and the latter from psychotics.
Factor analysis, when applied to either the orthodox scoring categories or to the content scales, emerges with factors like the following: fluency, generalizing ability, emotionality, imagination, extraversion-introversion, neurotic tendencies. Apparently, what the Rorschach expert does intuitively in evaluating the records of normals and neurotics can be obtained objectively by factor analysis. But it should be noted that direct statistical manipulation of the original Rorschach scoring categories does not lead to significant results unless they are distilled either through the mind of the expert or the hopper of factor analysis.

As for the relation of the Rorschach to changes accompanying psychotherapy, the results are in doubt. One study claims positive findings (21a) and three show negative results—Lord, (20) Carr (6) and Hamlin and Albee (17). The latter found that Muench's indicators of improvement did not hold up when groups exhibiting different levels of adjustment were compared; thus negating the one positive study mentioned.

Barry, Blyth, and Albrecht (3) compared test and retest data on the Rorschach with pooled judgment of patients at a Veterans Administration Mental Hygiene Clinic. Changes in ratings of adjustment level failed to correlate with changes on the Rorschach.

The recent topectomy study (7) offered an opportunity for testing what effect the lowering of anxiety induced by the operation might have on Rorschach performance. Neither orthodox scoring nor anxiety indicators (with the single exception of reaction time) succeeded in demonstrating any changes in the Rorschach performance of the patients, although other psychological tests showed such changes.

Psychometric scaling, however, did reveal certain changes and also provided a prognostic indicator. Three pairs of patients were selected, each pair consisting of one individual who decreased in anxiety and one who increased in anxiety after operation. The judgment of loss and gain in anxiety was based on psychological interviewing by means of anchored scaling devices and on the judgment of the psychiatrist. Only those patients in whom the two criteria concurred were selected. The results indicated that perception of movement of whatever variety, regardless of whether it was accompanied by empathy, correlated positively with anxiety, rising when the anxiety level rose and dropping when the anxiety level fell. The degree of tentativeness or insecurity in giving responses also correlated positively with anxiety. The following variables showed only a unilateral relationship to anxiety levels, declining with a decline in anxiety but showing no corresponding rise with rise in anxiety level: sensitivity to chiaroscuro, anatomical responses, perception of animate objects, perception of objects with texture, and degree of self-reference. The following variables also showed a unilateral but negative relationship with anxiety, showing increases as anxiety fell—accuracy of form perception and degree of congruity of the response. The statistical significance of these differences could be readily established since each patient could be analyzed as a separate sample and the significance of the difference for each patient determined. Only the variables that showed consistent changes from patient to patient were reported.

The fact that the classical Rorschach scoring is not sensitive to changes induced by somatotherapy is an old story. Lord (20) reports a
similar finding in psychotherapy. Perhaps the Rorschach test reflects only basic personality structure. Someone has suggested that the goal of therapy is to arrest diseases into defects and then teach the patient to accept these defects. If therapy consists in nothing more than the acceptance of one's disabilities, no change in fundamental personality is to be expected.

Experimental Studies
Perhaps the most important question that the experimentalist would like to answer about the Rorschach technique is: what is the stimulus, what role does it play, and whether present scoring of stimulus qualities such as Color, Form, Shading and perhaps Movement have definite stimulus correlates. One way of answering this question is to alter the stimulus characteristic to see whether the responses will change correspondingly.

The most revealing study of the stimulus properties of the Rorschach is a still unpublished study by Baughman (5). He set as his goal to differentiate as far as possible between that part of the response which inheres in the stimulus and that which inheres in the responder himself. Since the characteristics of the stimulus are more readily manipulated, he devised a series of modifications of the Rorschach plates so as to reveal the potency of a given part of the stimulus for evoking characteristic responses. He started off with the standard card and eliminated first the hue factor through photographing the standard series in black and white on panchromatic film, retaining all the nuances of the shading or differences in brightness and all the other characteristics of the blot. Then, he removed the shading by making line drawings of the more striking contours within the blot and the periphery. The third modification consisted of blotting out the inside details by making the entire inside of the blot black but leaving the islands of pure white, yielding a silhouette effect. The final modification consisted simply of the periphery or outline of the blot.

While it is difficult to draw up a correspondence between the altered appearance of the card and the specific Rorschach determinant which is most prominently present or absent, the following tentative suggestions can be made. The cards in which only the periphery was present would tend to accentuate whole responses and form responses. The cards with the inside details would tend to accentuate detail responses and perhaps organization, (Beck's Z). The silhouette cards would tend to accentuate form responses and perhaps tend to suppress white space responses. The achromatic cards and the complete original set are too well-known to require further discussion. The modified cards as well as the original set were administered to a group of 100 veterans, hospitalized for neuroses and character disorders. Each of the five series of blots was given to a group of 20 patients selected randomly which was equated with the other groups on IQ and educational level. Beck's scoring system was employed.

When he compared the records of the various groups, which had been administered variants of the original stimulus cards, he found, instead of the theoretically expected changes, a considerable degree of constancy in the responses. Virtually all of the significant differences were attributable to differences in stimuli which were simply objectively necessary for the occurrence of a given category. Surprising responses were not found distributed in all the various blot except for a case of the peripheral shading. Color and contour are not important in detail category. Apparently shading is necessary before changes in resulting patterns are produced.

Another analysis of the cases made by submitting the responses to an analysis of variance, indicating that the movement is independent of the content. Categories showing no slight differences from score were: R, W, Form T/R, A, FM, Total of content, and 8-9-10%. The study summarizing these findings is: "The severest assault upon the stimulus is necessary before changes in resulting patterns are produced."

Joseph Zubin
currence of a given category of responses. For example, Detail responses were found uniformly distributed in all the variant forms of the blots except for a drop in the case of the peripheral form series, when the stimulus for D is eliminated. Apparently color and shading are not important for the Detail category. Surprisingly, the M response occurred with significantly higher frequency in the silhouette version, indicating that perception of movement is independent of shading. Categories showing but slight differences from series to series were: R, W, Form level, P, T/R, A, FM, Total time, Diversity of content, and B-9-10%. Baughman aptly summarizes these findings: "The severest assault upon the stimulus is necessary before significant changes in resulting performances are produced."

Another analysis of the data was made by submitting the protocols including the reaction time data and the responses to experienced Rorschach workers to see whether color and shading shock patterns occurred only in the appropriate series. As a result of this investigation, it is reported quite conclusively that the time latency and response patterns supposedly typical of color shock occur with the same frequency whether color is present or absent. The same was found to hold true of shading and shading shock. A very substantial question is thus raised as to the wisdom of continued use of the shock indicators. Further evidence against the concept of "color shock" is provided by Siipola (27). In a specially ingenious experiment, she concluded that the affinity that color bears to emotion is based on a misunderstanding. Color shock for example is not due to color itself, but to the incongruity between the color and contour of a given blot area. The conflict engendered in the observer will take different paths depending upon the personality of the subject. While Siipola's experiment is not itself conclusive, no statistical verification being given, her explanation of color shock is ingenious, to say the least.8

Many of the moot questions in scoring could be answered by Baughman's research. Thus, "Bat" to Card 1 practically disappears as a response when "shading" and "black" are removed, while "Butterfly" occurs equally frequently in all the modified presentations. Similarly, "Map" occurs only when shading is present. Colored areas which yield anatomy responses practically cease to do so when color is eliminated. Color is of little importance in the response "Bat" to Card II—D32, in "Monkeys" in Card III—D3 and in "Bow-tie" and "Ribbon" to the same card. "Rejection" in this study was much more prominent when Color and Shading were absent, indicating that Form rather than Color or Shading is the primary source of rejection. This study needs to be repeated on groups other than the neurotic, with other scoring systems, and with other types of modification.

The second question deals with the effect of alteration of the state of the subject by means of drugs or hypnosis or shock. These experiments have not yielded very much because of the inexactness with which the psychological correlates of these induced states are known.

The third question deals with the effect of alteration of the circumstances surrounding the test. Prestige suggestions as to the importance of certain types of responses will alter the distribution of the responses in the direction of the prestige. Similarly, social situa-

---

8See paper of Bellak in this symposium.
tion, induced anxiety, etc. have been tried out. Some of the changes expected by Rorschach workers were validated, others not. The important conclusion to be drawn is that standard conditions are required and that Rorschach performance is not as insensitive to external conditions as some workers have claimed.

To summarize our findings thus far, the following facts are seen to emerge from our survey:

(1) Rorschach scoring and sign evaluation has an a priori basis which is not always validated by experimentally contrived techniques such as alteration of stimulus, alteration of state of the organism, etc.

(2) Globally, the Rorschach is an apparent success when the Rorschach diagnostician and the clinical diagnostician work closely together.

(3) Atomistically it is an apparent failure.

(4) Content evaluation whether done globally or atomistically is a success.

(5) Factor analysis of atomistic scores consisting of the usual combination of perceptual and content factors, or of content alone, correlate with personality.

What kind of a hypothesis, what kind of a model, could satisfy the above conditions? That is the scientific question before us. Before answering, let us examine Rorschach’s technique as an experiment. Experiments must have as their minimal requirements—subject, experimenter, apparatus or stimulus of some kind, a well-defined task, directions for the task, acceptance by the subject of the task, a response made by the subject and recorded automatically or by the experimenter. But that is not all—the most important part is still missing—the hypothesis.

What is the hypothesis underlying the Rorschach experiment—Rorschach never stated it explicitly, but it can be stated as follows:

(1) We perceive in the artificial Rorschach space in the same way we perceive in real space.

(2) The way we perceive in real space is determined by our personality.

Both of these assumptions are impossible to test at this time because we do not know how perception takes place in real space, nor how it takes place in Rorschach space. Gibson (14) has laid the foundations for the experimental determination of these two processes but we still have a long way to go before we can experiment with them. The relation between perception and personality must await the solution of the first two problems. What can be done meantime, and how can we explain the five facts which I have listed previously?

One hypothesis that suggests itself and which I humbly think merits consideration requires a shift of emphasis from the perceptual to the content aspects of the Rorschach. It is true that Rorschach veered away from the content analysis of ink-blot which was so popular with the psychologists of his day and espoused the formal aspects. He states “The content of the interpretations... offers little indication as to the content of the psyche.” But he may have been wrong, or, may have defined content too narrowly. If, we define content as the essential elements of the protocol, and regard it as one would regard any other interview material, and analyze the content, the mystery is solved. Once the perceptual scoring is eliminated, and instead a content analysis of the verbal productions of the subject is made according to such categories as: compulsive thinking,
disorganized thinking, or creative thinking; poverty of ideas or fluency; confabulation or clarity; rigidity or flexibility; contamination or its opposite; perplexity or straightforwardness; rejection or compliance, etc. etc., it will be discovered that such characteristics reveal themselves in the Rorschach the way they reveal themselves in the psychiatric interview. To be sure, the Rorschach interview is a standard interview and may lead to results which the free psychiatric interview can not lead to. But it is still an interview—an interview behind the veil of ink-blots.

This would explain why content of Rorschach protocols is related to personality, whether evaluated globally or in isolated scales, while formal Rorschach factors fail to relate to personality. This would also explain why factor analyses of both formal as well as content factors relate to personality. In the course of the analysis, the content factors affecting the formal scores are teased out—viz.—the kind of mental content which serves to reduce R, disorganize F, disembodv C or Sh, or prevent good M from arising in the mental patient and, mutatis mutandis, the kind of mental content which increases productivity and good responses in the normal, reveal themselves in the rotated factors. If this hypothesis be true, we should turn away from the indirect expression of mental content through determinants and location, and begin building scales for analyzing the content of the verbal productions directly. Such a beginning has been made by several workers and if we spend but ten per cent of the harnessed energy behind the Rorschach wheel to studying the interview basis of the Rorschach, we may bring nearer the day when the contradictions that now exist within the Rorschach field are resolved.

New developments in the interview itself are fast turning it into a scientific tool, and since the interview, in the last analysis, is the basis for personality evaluation, no test today can rise above it. If we obtain objective criteria via the interview for the classification and evaluation of personality, perhaps such criteria may serve as a basis for the validation of tests. But without an anchored interview, we float aimlessly in the sea of personality without compass or rudder.

Summary: This review of the failures of the Rorschach technique has found the following outstanding relationships:

(1) Global evaluations of the Rorschach seem to work when the Rorschach worker and the clinician work closely together.

(2) Atomistic evaluation, as well as global, of the content of the Rorschach protocols (as distinct from the perceptual scoring) seem to work.

(3) Atomistic analysis of the perceptual factors is a failure.

(4) Factor analysis of atomistic scores of both the perceptual as well as the content variety, seem to work.

The best hypothesis to explain these four facts is that the Rorschach is an interview and that its correct evaluation, like the correct evaluation of any interview, is dependent upon its content. If we provide scales for analyzing its content, we shall be well on the way towards clarifying many of the present day contradictions and obtain a better perspective on the evaluation of personality.

References


Received May 27, 1953.