The Susceptibility of the Rorschach to Malingering: A Critical Review

Glenn G. Perry and Bill N. Kinder
University of South Florida

Empirical studies of malingering on the Rorschach are reviewed. Results from these studies are to this point inconsistent and inconclusive. Although several indices are related to malingering in individual studies, no specific malingering pattern has been found that replicates across studies. A methodological problem specific to this literature is discussed and future research designs are recommended.

There are many instances in which individuals might desire to mangle, or fake a mental illness, in today's society. Criminals wishing to avoid jail sentences, accident victims wishing to sue for damages, and/or veterans seeking substantial government disability payments are examples of individuals who might consider malingering during psychological testing. Often, psychological test batteries include the Rorschach test because it is believed to be resistant to intentional manipulation by the subject (Exner, 1974, 1978, 1986; Fosberg, 1938, 1941, 1943).

In this review, we first briefly examine the process involved in formulating responses to the Rorschach inkblots. Next, we review the published Rorschach malingering literature in an attempt to determine if some reliable response pattern can be detected that would indicate malingering. Finally, we briefly consider the data analysis problems inherent in statistical analysis of the Rorschach test.

THE RORSCHACH RESPONSE PROCESS

Exner (1986) argued that the process of developing a response to the inkblots is a problem-solving task involving decision choices among a range of potential responses. It is likely the response process occurs very quickly and involves at least three phases: (a) visual input, encoding, and rank ordering of the potential
responses; (b) discarding potential answers due to low rankings or through censorship; and (c) selecting the remaining responses by reason of traits, styles, or state influences (Exner, 1986). If Exner's conceptualization is correct, the Rorschach response process is heavily influenced by conscious cognitive processes. Therefore, the possibility exists for volitional conscious alteration of the responses.

MALINGERING ON THE RORSCHACH

The early studies were carried out by Fosberg (1938, 1941, 1943). Fosberg (1938) had 2 subjects take the Rorschach repeatedly under four conditions: (a) standard instructions, (b) instructions to make the best impression, (c) instructions to make the worst impression, and (d) instructions to find specific determinants. Fosberg applied the chi-square technique to the protocols to determine if the four protocols were drawn from the same population (person). Because no significant differences were found, Fosberg concluded,

The Rorschach test does not lend itself to manipulations based upon the subjective personal estimation of character, good, bad, or indifferent (as can so easily be done in the more popular paper and pencil tests) but faithfully traces the more permanent picture. (p. 30)

In 1941, Fosberg again investigated the effect of varied instructions on Rorschach scores. He again used a test-retest design, but with 25 males and 25 females as the subjects. His instructions were: (a) standard instructions; (b) directions to make the best possible impression; (c) directions to make the worst possible impression; and (d) standard instructions. Rather than employing the chi-square technique, he made pairwise correlations across the four conditions. The pairwise correlations were quite high, ranging from the .70s and .80s for contents, to the .80s and .90s for location variables and determinants. Because the correlations between the standard tests and the "best" (.92) and "worst" (.93) tests were high and the correlation between the best and worst tests are still in the .80s, Fosberg concluded, "The attempt at falsifying the Rorschach protocol failed. The conclusion is that the test cannot be faked" (p. 83).

Unfortunately, Fosberg's statistical procedures were unsound. Cronbach (1949) reviewed Fosberg's studies and stated:

In one study he used chi-square to show that the psychograms for each person corresponded. But this statistical test merely showed that the D score in record 1 is nearer to D in record 2 than it is to W, C, or other scores. That is, he showed that the scores are not paired at random. But, since each score has a relatively limited range for all people—i.e., D tends to be large, m tends to be small, etc.—he would
have also obtained a significantly large chi-square if he had applied the same procedure to four records from different persons . . . Fosberg's second study, using correlation technique, is no sounder than the first. That is, pairs of values such as $W_1-W_2$, $D_1-D_2$, etc. were entered in the same correlation chart. As before, the generally greater magnitude of $D$ causes the two sets to correlate, but high correlations would have been obtained if the scores correlated came from two different subjects. (p. 424)

In his 1943 study, Fosberg reported on the methods adopted by subjects in attempting to fake results on the Rorschach during his 1941 study. The distributions of the methods are extremely platykurtic, indicating a diversity of method but little consensus on any method. He suggested that because the subjects were professors and students in psychology and they "could not materially influence the test results, certainly 'test-naive' subjects could not influence the Rorschach test" (p. 120).

A result of Fosberg's findings was a strengthened belief that the Rorschach process was one involving involuntary projection of unconscious personality traits. The Rorschach was believed to be unfakable and only one research project was published for nearly a decade. Benton (1945) reported on clinical data, suggesting that the overall behavior of malingerers appears to be out of keeping with their other test performance and general behavior. He further suggested that a decrease in $P$ and a failure to see any popular ($F$) responses as well as increases in response time and rejections may be characteristic of malingers on the Rorschach. Unfortunately, his conclusions were based on his interpretation of clinical data and did not include any statistical analysis. Therefore, his work attracted little attention and did not appear to stimulate further research activity.

The next studies were published in the 1950s (Carp & Shavzin, 1950; Feldman & Graley, 1954). Carp and Shavzin modified Fosberg's design using 20 male psychology students in a test-retest design counterbalanced on instructions to either give a "good impression" or a "bad impression." When the data were analyzed, using a $t$ test, only the $Z$-score difference was statistically significant. They went on to suggest, however, that the individual profiles clearly showed differences. "But the direction taken by individual subjects was so diverse, among the individual subjects, that they were balanced out in the analysis" (p. 232). Carp and Shavzin then calculated chi squares to determine if the two distributions from each subject were similar and found three of the distributions were significantly different. Carp and Shavzin concluded that some subjects can vary their personality picture as reflected by the Rorschach under instructions to make good or bad impressions.

Feldman and Graley (1954) administered the Rorschach to subjects in group administrations as recommended by Harrower-Erickson (1941). Using a static-group comparison design, one group was given the Rorschach first under standard instructions and retested with instructions to "give the worst possible
impression." A second group was given one administration of the Rorschach with instructions to "give the worst possible impression." They found both "malingering" groups responded with increases in inanimate movement (m), color-form plus pure color (CF + C), form-color (FC), sex, and anatomy (An) and decreases in P when compared to the control group. Their results are of questionable external validity, however, because factor analysis of data derived from group administration of the Rorschach has been shown to measure a different set of underlying constructs than factor analysis of data derived from individual administration of the Rorschach (Shaffer, Duszynski, & Thomas, 1981).

Feldman and Graley asked subjects what strategies were used in the malingering attempt. Strategies reported, in order of frequency, were as follows: "(a) avoiding the normal response; (b) emphasizing sex; (c) mentioning symptoms of maladjustment; (d) using a specific mental disorder as a guide; (e) using an unspecified mental disorder as a guide; (f) stressing aggressive and gory concepts" (p. 332).

In 1967, Easton and Feigenbaum used a test-retest design with a control group. Their experimental group received standard instructions for the first testing and malingering instructions on retest. Their control group received standard instructions on both administrations of the test. This design allowed them to examine the effect for repetition as well as the effect for instructions. They found a decrease in D, P, Obj, and R as a result of the instructions to malingering. They also found, however, an increase in D, H, Ad, and R as well as a decrease in A and P as a result of repetition. Therefore, part of the variance in the results from a test-retest design seems to be a function of the interaction of the repetition effects and the instruction effects. This interaction clouds the interpretation of such designs. This problem appears especially significant when test-retest designs are used with the Rorschach due to the carry-over-effect influence of the Rorschach inquiry.

Another decade followed without further published research. When research on malingering resumed in the late 1970s and the early 1980s, there was a shift in research design, moving away from the test-retest design. Test-retest designs are seen in only one study during the 1970s and 1980s (Seamons, Howell, Carlisle, & Roe, 1981). Seamons et al. used a counterbalanced test-retest design with four cells of 12 prison inmates from each of the following diagnostic categories: nonschizophrenic, latent schizophrenic, residual schizophrenic, and psychotic schizophrenic. During the initial testing they administered the Rorschach to half of each group with instructions to "appear as if you are a normal, well adjusted individual," and to the other half with instructions to "appear as if you are mentally ill and psychotic." During the second testing, each subject received the alternate instructions. The protocols were scored in accordance with the Exner (1974) scoring system and an analysis of variance (ANOVA) was performed on the 48 variables scored. Seamons et al. found that
when the subjects were asked to appear mentally ill, their protocols showed a decrease in P and an increase in Bi, ep, INCOM, FABCOM, and dramatic responses (defined as themes of depression, sex, blood, gore, confusion, mutilation, hatred, fighting, and decapitation). No significant changes were noted in the ratios, percentages, and derivations. Expert judges were able to differentiate correctly between those who were asked to appear normal or asked to fake psychosis. Seamons et al. suggested that when the X + %, F + %, and L are in the normal range with a high number of dramatic, blood, texture, shading, vista, nonhuman movement, or inappropriate combinations, it may be indicative of an attempt to appear mentally ill.

More recent studies have generally used multicell designs with control groups rather than test-retest designs. The change to multicell designs with control groups began with Bash (1978). Bash compared diagnosed malingerers with hallucinating schizophrenics, nonhallucinating schizophrenics, and non-psychotic mental patients. She found that the malingerers could be differentiated from all other groups only on the following scores: rejection of cards, failure to report easy populars, and low F%. Bash concluded that malingerers are not a special group because they cannot be differentiated on many other variables.

The final four studies (Albert, Fox, & Kahn, 1980; Meisner, 1984; Mittman, 1983; Overton, 1984) trained subjects in various aspects of psychopathology and then asked them to fake the psychopathology during psychological testing. The assumption in these studies is that subjects could not easily fake a condition they did not understand. This lack of knowledge of the role to be played could have led to very weak manipulations in previous research, because past subjects were asked to fake a condition when they could have had little or no knowledge about the condition to be faked.

Albert et al. (1980) and Mittman (1983) evaluated the degree to which expert judges could detect malingered. Albert et al. established four cells of six protocols each from the following groups: psychotic inpatients, uninformed fakers, role-informed fakers, and normals with standard instructions. These protocols were randomly assigned to packets containing one protocol of each condition. The packets were mailed to 261 Fellows of the Society for Personality Assessment with a request that a diagnosis be developed, that the judge reports his or her degree of certainty in the diagnosis, that the judge evaluates the possibility of malingering, and that the protocol be rated on eight dimensions of pathology. Usable sets of responses were obtained from 46 of the packets. Albert et al. used a chi-square technique and found that (a) the uninformed fakers were diagnosed psychotic as often as the actual psychotics, (b) the informed fakers were diagnosed psychotic at a higher rate than the actual psychotics, and (c) judges were equally certain of the diagnoses across experimental groups.

Mittman (1983) also used judges to interpret Rorschach protocols. She asked 90 clinicians, well trained in the interpretation and scoring of the Rorschach, to
judge a packet of five randomly assigned protocols. The protocols were taken from groups of inpatient depressives, inpatient schizophrenics, uninformed fakers (who were asked to simulate schizophrenia but were uninformed about the role), informed fakers (asked to simulate schizophrenia and informed about the role), and normal controls with standard instructions. Mittman found that the judges diagnosed the uninformed fakers as schizophrenic significantly less often than the actual schizophrenics. Judges, however, were fooled by the protocols of the informed fakers who were well versed in the role of schizophrenia. Mittman concurred with Albert et al. (1980) and concluded that the Rorschach could be susceptible to malingering if respondents were well informed about their role.

These two studies have demonstrated that “experts" cannot always discriminate between actual psychotic profiles and role-sophisticated malingering profiles. Because role-informed malingerers have been judged to be psychotic more frequently than actual psychotics, the validity of the Rorschach protocol must be questioned if the protocol arises from a situation in which malingering for secondary gain could be an issue. To detect this malingering, research turned again to seeking discriminatory patterns in the structural data, but this time informed fakers were included in the research designs (Meiser, 1984; Overton, 1984).

Meiser (1984) used five trained examiners, blind to the experimental conditions of the study, to administer the Rorschach to 58 nondepressed university students. Immediately prior to taking the Rorschach, 29 of the subjects were trained in the clinical symptoms of depression and asked to fake depression. The remaining 29 subjects received standard instructions. Meisner analyzed the variables by first evaluating the distribution of the variable and then using a two-factor ANOVA, a Mann-Whitney U test, or a Fischer Exact Test. Meisner found that when the subjects attempted to fake depression their protocols showed a reduction in R and an increase in blood content and morbid special scores. No determinants were significantly affected by attempts to fake depression. Meisner concluded that the susceptibility of the Rorschach indices of depression to faking was markedly limited.

Overton (1984) randomly assigned 10 male and 20 female undergraduate psychology students to three groups. Subjects were administered the Rorschach by examiners who were blind to the experimental conditions. The control group received standard instructions, the noninformed group received malingering instructions, and the informed group was given specific instructions on how to mangle in the role of a schizophrenic. Data were analyzed using multivariate analyses of variance (MANOVAs) and ANOVAs. Although Overton found multiple treatment effects, he suggested that reliable patterns of malingering could not be found. The data indicated that the noninformed fakers appeared closer to psychotic than the informed fakers who produced protocols too pronounced to be believed and whose behavioral presentation was inconsistent
with psychosis. Overton suggested that malingering could best be identified in the context of extratest behavior, history, and possible motivation of the subject.

Pettigrew, Tuma, Pickering, and Whelton (1983) used a new multiple-choice group administration Rorschach instrument to differentiate students asked to simulate psychosis from students with normal instructions and civilly committed psychotics as well as forensic inpatient psychotics. Their stimuli were color reproductions of standard inkblots in reduced size on a single page. Five responses were preselected for each blot with four multiple-choice alternatives designated for each response forming a 50-question objective test. The alternative answers for each question varied form and wording from good form, nonbizarre wording to poor form, bizarre wording. Pettigrew et al. found that simulators gave more good-form, bizarre-wording answers and suggested that simulators are “exposed by their own perceptual accuracy” (p. 468). Although this is the only design to date that has controlled for variance in R, the uniqueness of the design suggests that the results are difficult to generalize to the general Rorschach malinger literature.

The Rorschach Workshops (1987) Alumni Newsletter reported that their 5-year project on malingering is nearing completion. With respect to the Schizophrenia Index (SCZI), their early results suggest that

between 20% and 25% of subjects who have considerable information about schizophrenia can obtain values of 4 for the SCZI. Almost none obtain values of 5 for the SCZI . . . 5 of 15 nonpsychotic psychiatric inpatients who have contact with schizophrenics were able to produce a record in which the SCZI has a value of 4 but none in the sample produced a record in which the SCZI has a value of 5. (p. 11)

Considered as a whole, the literature suggests that Rorschach scores can be altered by conditions external to the subject and/or volitionally by the subject. At this time, however, the results are extremely inconsistent and no reliable pattern of responses in a protocol that would indicate the presence of malingering has been identified. A few patterns have been suggested, but to date these have not been replicated. Perhaps a significant proportion of this heterogeneity, and failure to replicate, is due to unsound design and/or data analysis procedures.

RORSCHACH DATA ANALYSIS PROBLEMS

When research is conducted using the Rorschach, research design and data analysis must take into account the variability of the number of responses (R) between subjects and the dependence on R of many determinants, indices,
ratios, and so on. This appears particularly critical in Rorschach malingering research because the malingering or fake bad subjects have been shown to respond by reducing \( R \) in their protocols. The primary problem created by differences in \( R \) is that the various Rorschach indices are not experimentally independent—the total record is collected at once, and productivity influences nearly all the scores. Therefore, if group scores are examined without controlling for differences in \( R \), the group differences found could be a function of differences in \( R \) rather than a function of the experimental condition.

One of the strengths of the Rorschach is the lack of structure, which allows the subject to structure the experience to meet his or her idiosyncratic needs. This freedom to create a distinctive response set results in an individualistic protocol, giving the examiner a unique glimpse of a subject's cognitive processes. This type of information is superb when the interest is in an idiographic examination of the individual, but complicates the investigation when attempting to look at nomothetic data from a group of individuals when not only each group member has a different number of responses, but also, each group mean differs in responsivity.

The very beauty of the Rorschach, its lack of structure allowing perceptive clinicians a glimpse at the individual's idiographic thought processes, now becomes a hindrance. Differences in \( R \) must be controlled to the maximal extent possible to investigate nomothetic Rorschach data. Without such control, differences in the magnitude of the Rorschach variables in question can be confounded, resulting in interpretation difficulties. Cronbach (1949) stated,

Where groups differ in total number of responses, this factor must be held constant before other differences can be soundly interpreted. Three devices for doing this are: rescoring a fixed number of responses on all papers, constructing subgroups equated on the number of responses, and analyzing profiles of normalized scores (pattern tabulation). (p. 426)

A fourth method for controlling for the differences in \( R \) is statistical control through partialing or residualizing (Exner, Viglione, & Gillespie, 1984; Mason, Cohen, & Exner, 1985; Shaffer et al., 1981).

As early as 1949 Cronbach suggested that because the errors and poor choices of statistical procedures are so widespread, few of the conclusions for statistical studies of the Rorschach can be trusted. Almost 40 years later, these same criticisms appear to apply, at least with respect to the malingering data reviewed in this article. It is apparent that when conducting a Rorschach study using nomothetic data, research groups must be equated on \( R \) prior to any significance tests. No research to date of malingering on the Rorschach has controlled for differences in \( R \) prior to submitting the scores to various statistical tests. This suggests that the resulting findings from the entire literature on malingering on the Rorschach must be questioned because any variability found may have been
a factor of the variable's dependency on R rather than the experimental conditions. The findings of Albert et al. (1980) and Mittman (1983) are exempt from this scrutiny because their findings were based on decisions made by judges who reviewed idiosyncratic protocols, and they did not attempt to analyze nomothetic data by collapsing the internal scores within the protocols across groups.

**SUMMARY**

Perhaps the finding most prominent in the literature is that when subjects were asked to mangle or fake psychosis on the Rorschach, they responded with protocols exhibiting a reduction in R. Because R was uncontrolled in further analyses, findings indicating reductions in other determinants must be questioned. Researchers have reported that the protocols of informed malingerers display a reduction in P, F, M, and D responses as well as a reduction in the F + % and X + %. Each of these indices has been shown to be positively correlated with R, which clouds the interpretation of these findings.

Increases in determinants or indices, in a reduced R protocol, are easier to interpret as the results of the experimental intervention. This is especially true of those determinants and indices that are positively correlated with R. Increases have been shown in the protocols of informed malingerers for Bl, INCOM, FABCOM, bizarre content, CF + C, and in the number of rejections. Ongoing studies suggest that informed malingerers can also increase the number of SCZI responses. As yet, however, no reliable pattern of responding has been found to be specifically related to malingerers across studies.

Future research on Rorschach malingerers needs to control for the number of responses through either experimental design or appropriate statistical control methods. Second, the measurement scale and the frequency distributions of the variables in the study should be carefully evaluated to ensure that the proper statistical tests are chosen and that the assumptions of these statistical tests are not violated. Third, the experimental group (malingerers) should be trained in the psychopathology to be role played. Fourth, no study of malingerers has utilized an independent manipulation check to determine if malingerers are indeed attempting to carry out their instructions to “fake bad.” The inclusion of the MMPI validity scales might provide valuable data in this regard (Graham, 1987).

Finally, it is important to consider that individuals actually attempting to mangle in a clinical setting are operating from a motivational set quite different from that of college undergraduates instructed to mangle or fake the Rorschach in an experimental setting. Additionally, the question remains as to whether individuals consciously planning to mangle or fake the Rorschach can actually “learn” the types of behaviors or response styles thought to be indicative
of serious emotional disturbance without specific instructions. The early data from the Rorschach Workshops (1987) suggested that this is a distinct possibility, at least for some individuals, and this possibility clearly merits further investigation.

REFERENCES


Bill Kinder
University of South Florida
Department of Psychology
Tampa, FL 33620

Received May 10, 1988
Revised August 18, 1988