

McKinzey, R.K. & Campagna, Victoria (April 27, 2002). The Rorschach, Exner's Comprehensive System, Interscorer Agreement, and Death. *WebPsychEmpiricist*. Retrieved (date) from: [http://www.wpe.info/papers\\_table.html](http://www.wpe.info/papers_table.html)

WPE

WebPsychEmpiricist

The Rorschach, Exner's Comprehensive System, Interscorer Agreement, and Death

April 27, 2002

R. K. McKinzey

Oakland, CA

Victoria Campagna

Redwood City, CA

Abstract

Interscorer agreement of Exner's Comprehensive System for the Rorschach was tested. Thirty qualified psychologists volunteered to score a single protocol, which had been used as part of a Death Penalty appeal in a case of rape and murder. 63% of the psychologists made scoring errors. After the errors were corrected, the resulting scorings showed wide variation. Only 10% agreed with the original psychologist's conclusions that both SCZI and DEPI were elevated. The range of scorings and amount of disagreement make Rorschach interpretations easily rebutted in forensic cases.

## The Rorschach, Exner's Comprehensive System, Interscorer Agreement, and Death

April 27, 2002

For a normed test to be accurate, the users of the test must be able to agree on the conversion of raw data to normed scores. That is, the scorers must be able to agree amongst themselves: this is called interscorer agreement, or interscorer reliability. The level of interscorer agreement limits the accuracy of the test, so test authors are required to produce such statistics (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1999). If the scorers can't agree on the scores, of what use is the test?

The Rorschach is a test widely used in practice: McCann (1998) contends that "it has been consistently shown that the Rorschach is used by more than 80% of agencies or practitioners who engage in psychological assessment." (p. 126) The Rorschach was taught in 94% of all APA graduate programs in 1994 and showed consistently high visibility in other years for which data was kept (McCann, 1998).

Interscorer agreement statistics for the Rorschach's Comprehensive System's (CS) scoring method and norms were presented by Exner (1993). However, Wood, Nezworski, and Stejskal (1996) argued that Exner's statistics were inadequate, incomplete, and potentially misleading. Instead of citing how often scorers agreed with each other, Exner had cited how often they agreed with a standard of correctness. Further, Wood et al. noted a previous study (Exner, 1988) that showed a high level of scoring errors.

Because the interscorer reliability of the CS was now in question, three studies were done. One (Acklin, McDowell, Verschell, & Chan, 2000) used 40 valid (i.e.,  $R > 13$ ) protocols, 20 normal, 20 clinical. All 40 protocols were scored by the second author, and the scores compared to a second rater, one of three graduate students trained by the first author. The authors argued that Cohen's kappa and the Intraclass Correlation Coefficient (ICC) were the best statistics to evaluate the scores, and set their acceptable levels at  $> .61$ . They used the Rorschach Interpretive Assistance Program, Version 3.1 (RIAP3) to calculate the various combinations of scores. In the normal sample, 170 score combinations occurred frequently enough to be considered, and 35 (20%) failed, with unacceptable levels of reliability (i.e., below  $.61$ ). In the clinical sample, 174 score combinations occurred frequently enough to be considered, and 16 (9%) failed. As expected, reliability was better in combination scores that were derived from the entire protocol (such as indexes and constellations) rather than from individual responses. Even so, the Schizophrenia Index (SCZI) failed to reach acceptable levels of reliability in either sample.

A second study (Nakata, 1999) also used 40 valid protocols, 20 normal, 20 clinical. All three scorers were graduate students. One rater scored all the protocols; the other two split the samples. No mention was made of RIAP. For the clinical sample, 109 variables occurred frequently enough to be scored, and 40 (37%) had kappas or ICCs below  $.61$ , including the Depression Index (DEPI).

For the normal sample, 107 variables occurred frequently enough to be scored, and 28 (26%) had kappas or ICCs below .61, SCZI amongst them.

A third study (Shaffer, Erdberg, & Haroian, 1999) used 52 records from a normative, nonpatient sample. The two senior authors were raters. No mention was made of RIAP. Of the 109 variables, 90 occurred frequently enough for kappa to be calculated, and 37 (41%) had kappas below .61. The authors grouped 17 variables that included various subscores (e.g., Location, Form Quality), and 8 (47%) had kappas below .61.

While the Rorschach has had no trouble being admitted in courtroom settings (McCann, 1998; Weiner, Exner, & Sciara, 1996), the question of interscorer agreement gains an added urgency when used in forensic settings. For example, a Rorschach protocol judged to reveal a psychosis in a person accused of a capital offense can be used as part of the person's defense: the debate on the test's usefulness becomes part of a life-or-death decision. If the elaborate Rorschach CS scoring rules cannot force two opposing psychologists to agree on the protocol's outcome, the research question greatly affects the practice of using the test.

A unique opportunity to readdress the problem of Rorschach CS interrater agreement appeared during the appellate process of a man condemned to Death Row. A Rorschach was used by a qualified psychologist to bolster a claim of undiagnosed Schizoaffective Disorder. We decided to use the protocol to see if other qualified scorers would agree among themselves or with the scores underlying the interpretations of the original psychologist.

#### Method

Participants. Doctoral level psychologists who were practiced in the CS were recruited via e-mail, largely from two Internet mailing lists, PsyLaw-L@crcvms.unl.edu and Rorschach@maelstrom.stjohns.edu. The former is a discussion group sponsored by APA's Division 41, and is dedicated to forensic issues in psychology. The latter is a discussion group dedicated to the use of the Rorschach, including the problems posed by the CS. Entirely via e-mail, volunteers returned a consent form, filled out a demographic questionnaire, agreed to not collaborate, and scored the protocol. All but one returned the results via e-mail.

There was a total of 30 participants.<sup>1</sup> They were 77% male and all Caucasian. All but two were doctoral level, and two were diplomates. Their ages ranged from 30 to 71 ( $M = 48$ ,  $SD = 10$ ). They originally learned the CS from Rorschach workshops (36%), graduate school (53%), or being "self-taught" (10%). Of the 24 who had ever taken a Rorschach workshop, 7 had taken a workshop within the last year, 11 within the last five years, and 5 within the last 8-13 years. Their experience with the CS ranged from 3 to 29 years ( $M = 12$ ,  $SD = 7$ ). The number of times they have used the CS in a forensic setting ranged from 0 to 850 ( $M = 86$ ,  $SD = 176$ ).

---

<sup>1</sup> Two more people withdrew their data after learning of the results, and are not included in any of the analyses.

Materials. The protocol used (Table 1) came from a 42 year old man who is (as of this writing) still alive on his state's Death Row. He was convicted of a rape and murder. He is a high school graduate. He was married, with two children. He has a juvenile arrest record, and has abused marijuana, hallucinogens, alcohol, barbiturates, opium, cocaine, and prescription pain medication. He had a yearlong period of heavy amphetamine use prior to the murder. His physical history includes many skin wounds and infections, a broken right arm, shot wounds in the left hand and right thigh, hepatitis, appendicitis, insecticide poisoning, right knee and hernia surgery, chlorine inhalation, and a stabbing. Documented neurological history includes: a fall with possible loss of consciousness (LOC), being kicked in the right eye, a broken nose, a head cut, and being hit by a car (reportedly with LOC). An EEG was normal. An expanded Halstead-Reitan Battery (HRB) was negative.

The Rorschach and HRB were administered by John Doe, Ph.D.,<sup>2</sup> a graduate of California School of Professional Psychology, a licensed clinical psychologist, and an American Board of Professional Neuropsychology diplomate. He did a post-doctoral fellowship at a well-known state university. He has consulted with the Department of Corrections (DC) of a large state, was a staff psychologist there, and currently teaches assessment at an accredited professional school of clinical psychology. He has been "principal investigator" on four research projects at the state's DC. He is a member of the APA (two divisions), National Academy of Neuropsychology, International Neuropsychological Society, and the Society for Personality Assessment. He has published several times, including a chapter and article on the use of the Rorschach, both co-authored with a well-known Rorschach scholar.

Dr. Doe's CS scoring of the Rorschach is presented in Table 2. The doctor used Exner's 1990 CS program for further scoring, which produced positive SCZI and DEPI scores. Accordingly, the doctor offered a diagnosis of a previously undetected Schizoaffective Disorder, which was used as part of an appellate argument that insufficient efforts were made during the original trial, mandating a new trial.

Table 1 reproduces the protocol as received during discovery. It is in a non-standard format, without response numbering and with the inquiry at the end rather than side-by-side. The doctor's chart was appended to his deposition which made the protocol public record.

Procedure. The scores were collected and entered into the newest version of RIAP, now in its fourth version. Scoring errors as identified by RIAP4™ were sent back to the scorers by e-mail for correction. The resulting corrected scores were compared to each other and to Dr. Doe's.

### Results

Some of the participants commented on the quality of the protocol:

- This was one of the more complex Rorschachs which I have scored, perhaps ever. Very difficult.

---

<sup>2</sup> For privacy, the doctor's identity is shielded.

- The way the protocol was presented, I wasn't sure where one response ended and the other began. This made it difficult to know which inquiry was related to which free association response. Because of this, I would think that the reliability will be artificially low.

- Inquiry could be better.

- The lack of notation on the location sheet interfered with coding. It was, in particular, difficult to code responses to Card III (where was the "spot"?), Card IX (where was the man and his "astroprojection"?), and Card X (where were the crabs? The crabs might have been D1 and, therefore, Populars).

- Response number one looks like THREE DIFFERENT responses to me--- probably an error in administration as this could have been clarified by asking the client--is this all one response or more than one--clearly the fox and 2 winged people are 2 different responses--thought the Wiley coyote could be the same as the first fox. I've scored it as 3 different responses--which is what made the most sense to me. As for VI: This was likely a rejection--I believe that Exner suggests that you do NOT accept "inkblot" for a response--and request another response--so I'm not sure this is a valid response.

- The most difficult portion of the scoring was the fact that it was difficult to tell responses apart from each other. I am used to scoring protocols where it is clear what response is what and exactly which inquiry goes with each initial response (like the workbook). Took me awhile to figure out all the initial responses were together and the inquiries followed.

- Had there been a tape recording of the testing available, I believe many of the scoring decisions would have been more accurate. The present protocol scoring is a 'shot in the dark'. There were few clues to the response locations on the Location Chart.<sup>3</sup>

- It is almost more a guess than an educated scoring, since many things were unclear. First of all, you should number the R. And more details in the location chart should be added.

- I did my best scoring but was unsure about whether some responses were actually separate or not--I'm referring to card III, the "Spot" response, which appears to be part of the "puppy." By the way, nice choice of protocol--a very rich series of responses.

Three participants pronounced the protocol unscorable altogether (the first note is from the second of the two diplomats):

- This protocol, strictly speaking, is unscorable via Exner's method. The CS has standardized administration procedures, and this record doesn't appear to follow them. For example, I can't tell if the response to card I is two responses or one, but it sure looks like it should be two (fox, 2 winged people), and probably 3 (Wylie coyote). I can tell you that if a trainee gave me this protocol, I'd be taking them back to the basics and have them read the administration procedures again. Also, the queries don't seem to address the gestalt nor key words with any consistency. You should get another protocol that is CS in data collection before doing a study on scoring reliability. GIGO<sup>4</sup> you know.

- I have spent a fair amount of time with your Rorschach. Some of it was a bit of frustration. You see, I've taught Rorschach for several years and do quite a few of them. I mention that to point out that this protocol, strictly speaking, is unscorable via Exner's method in its present form. If I could ask a few questions of the person performing the inquiries, I could probably clean this up so I could score them. For instance, the first percept is either a very strange fabulized combination or is actually two or three percepts. However, this is not clarified via the inquiry. Therefore, fundamental material has to be either inferred or assumed.

<sup>3</sup> This diplomate nevertheless provided a scored, interpretable (i.e.,  $R > 13$ ) protocol.

<sup>4</sup> Garbage In, Garbage Out

•This protocol was not collected or recorded properly. It is difficult to tell where one percept begins and another ends. Each percept should be numbered. The way it looks here, there are fewer than 14 responses, thus it isn't a valid protocol and shouldn't have been accepted as one. It doesn't qualify for scoring. Also, I believe that when you do the inquiry, you are to read back the whole response, or the vast majority of it, not just repeat certain words. You haven't adequately encouraged description in the inquiry. Thus, scoring would not be accurate. Better luck next time.

Thus, 10% of the participants pronounced the protocol to be unscorable. Although the protocol was in non-standard format, the inquiry was sufficient for 27 (90%) of the participants to offer scorings in the same ranges as Dr. Doe's (Table 3).

Consistent with previous research (Guarnaccia, Dill, Sabatino, & Southwick, 2001), RIAP4™ proved to be quite necessary. Of the 30 participants, 27 offered scorings. Of that 27, 19 (70%) made one or more scoring errors<sup>5</sup> that were identified by RIAP4™. Two participants made seven errors apiece, one participant made five errors, two participants made four errors, two participants made three errors, five participants made two errors, and seven participants made one error. The most common error was to omit or assign an incorrect z score (a total of 28 errors made by 16 scorers). Other errors made by more than one scorer included scoring incompatible determinants, misscoring "Popular" for inappropriate content, and assigning form quality scores to formless determinants. One participant refused to correct his error.

The scorers with a Psy.D. made more errors than the scorers with a Ph.D., who in turn made more errors than the scorers without any doctorate. The scorers who had attended at least one Rorschach Workshop made more errors than the scorers who had never attended a Workshop. None of these differences were statistically significant.

The number of responses scored varied widely. Of the 27 participants who thought the protocol was scorable, 6 (22%) scored 13 responses and 1 (4%) scored 12 responses, making the protocol uninterpretable.

Table 3 presents the constellation distributions of the remaining 20 participants compared to each other and to Dr. Doe. Of the 30 participants, 3 (10%) agreed that this protocol supported Dr. Doe's finding of elevated SCZI and DEPI constellations. Of the 30 participants, 6 (20%) agreed that SCZI (but not DEPI) could be scored as elevated; 4 (13%) thought that DEPI (but not SCZI) could be scored as elevated; 2 (7%) thought HVI could be scored as elevated; and 1 (3%) thought S-CON could be scored as elevated. None scored positive OBS or CDI<sup>6</sup> constellations. Dr. Doe's scores are not outliers, as they are within the ranges produced by the other scorers.

---

<sup>5</sup> Some participants scored Level 1 Special Scores without using a number (e.g., DV instead of DV1). This scoring practice is allowed by previous Rorschach scoring programs, but not by RIAP4™. Such scores were not counted as scoring errors.

<sup>6</sup> This score would seem to be a false negative.

Table 4 presents the scores' distributions among the 20 interpretable scorings. Of the 76 variables,<sup>7</sup> only the 7 (9%) unused content scorings<sup>8</sup> achieved unanimity. Only two of Dr. Doe's scores were outside the distributions. Table 5 presents the distributions of interpretable variables compared to their cut scores. As with the constellations, there was considerable disagreement among the scorers on the number of responses (14-16), EB Style, level of D or Adjusted D, levels of form quality, and color ratios.

Table 6 presents a variety of variables using specific cutoffs, each indicating a specific interpretation. Some agreement is found here, with 8 (31%) of the 26 showing unanimity. Again, Dr. Doe's scorings are within the distributions.

### Discussion

Acklin et al. (2000), Nakata (1999), and Shaffer et al. (1999) found less than perfect interscorer reliability rates for Exner's Rorschach CS under laboratory conditions (i.e., high quality protocols, and similarly trained scorer pairs). Such imperfection, it seems, leads to a dramatic amount of disagreement when a lower quality protocol and many totally independent scorers are used. However, if the scorers can't agree on the scores, of what use is the test?

We used a single protocol provided in a life-or-death forensic case. The protocol was administered by an experienced and well-qualified psychologist. It was sent to 30 independent, equally well-qualified and experienced psychologists, asking them to score it. We did so looking for scoring agreement.

We found discord, and at a disconcerting level. Our well-qualified, experienced scorers disagreed on whether this death penalty protocol was scorable, interpretable, schizophrenic (i.e., SCZI was elevated),<sup>9</sup> or depressed (i.e., DEPI was elevated). Some of them made errors in scoring, and some argued with the scoring program. Some disagreed with each other and with the original, equally qualified scorer. Some (and only some) complained the protocol was of poor quality, then scored it anyway.

An obvious criticism of this study is its use of a protocol that was, in the opinion of some of our scorers, poorly administered. Even if this were so (and surely the highly credentialed psychologist who administered it would disagree with that statement) the fact that the inquiry generated so much criticism appears to be ammunition for the epithet GIGO. However, if the original psychologist, who has such impeccable CS qualifications, can produce such a protocol, there is at least a reasonable expectation that other psychologists in research and practice will make

---

<sup>7</sup> The selection of variables for Tables 4, 5 and 6 was intended to be as inclusive as possible, and was modeled on other recent Rorschach studies that have reported descriptive statistics for a large number of CS variables (e.g., Acklin et al., 2000)

<sup>8</sup> That is, none of the protocol's responses could be characterized as being that type of content.

<sup>9</sup> The issue of whether an elevated SCZI or DEPI score allows a diagnosis of schizophrenia or depression is beyond the scope of this article.

similar or (given the high level of credentials of the current original psychologist) even more errors. The point is, this protocol should, by all rational standards, have been of the highest quality. That so many agreed that it is not, is worrisome.

From our data, we offer one conclusion about practice, many questions, and one recommendation about research. We conclude that the practice of using the Rorschach in a forensic context is no longer supportable, not because the Rorschach will be kept out of court (McKinzey & Ziegler, 1999) but because it is so easily rebutted. Consider the problems of surrebuttal<sup>10</sup> when these rebuttal arguments are made:

1. The rebuttal psychologist announces the protocol in question to be so badly administered that it is unscorable (as 10% of our participants did), and any opinions derived discounted.
2. The rebuttal psychologist announces RIAP4™ has discovered scoring errors which, when corrected, change the interpretation.
3. The rebuttal psychologist announces the protocol was scored incorrectly (due to poor inquiry?), which changes the interpretation.
4. The rebuttal psychologist combines the responses so that R is less than 14, making the protocol uninterpretable.
5. The rebuttal psychologist declines to score the protocol altogether, choosing instead to use dramatic language that denounces the Rorschach in general as useless, due to poor interscorer agreement (see Table 5), botched normative system (Hamel, Shaffer, & Erdberg, 2000; Shaffer et al., 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001a),<sup>11</sup> and 81% false positive rate (Mittman, 1983).
6. The rebuttal psychologist denounces the Rorschach literature as hopelessly uninterpretable, citing the multiple problems of inaccurate norms, problematic scoring system, and lack of attention to scoring errors.
7. The rebuttal psychologist cites the lack of any method to detect response bias (i.e., faking good or bad) (Exner & Weiner, 1991; McCann, 1998).
8. The rebuttal psychologist agrees with the scoring, but, using the same scores, disagrees with the interpretation (Mittman, 1983).
9. The rebuttal psychologist declines to testify, and instead crafts a “learned treatise attack,” a series of cross-examination questions that features the above issues.

After wrestling with the data, we are left with many questions, some rhetorical, some heuristic, some practical. Was this protocol (which had come with such impressive credentials) so

---

<sup>10</sup> That is, what will the administering psychologist argue if given the chance to return to the courtroom and reply to a rebuttal psychologist’s arguments?

<sup>11</sup> For a robust discussion, see the commentaries following Wood et al. (2001a): (Aronow, 2001; Exner, 2001; Hunsley & Giulio, 2001; Meyer, 2001; Widiger, 2001; Wood, Nezworski, Garb, & Lilienfeld, 2001b).



completely badly done that it is unrepresentative? If so, why was there no consensus on its scorability? What will the administering psychologist offer as a defense? How many other clinical (field) and research protocols are of the same debatable quality? What does that question mean for the use and interpretation of the Rorschach and its literature?

Are the rules of sufficient inquiry and administration debatably inadequate? Did our participants score an unscorable protocol simply because of demand characteristics?<sup>12</sup> Is such misscoring happening in clinical, forensic, and research settings? Why didn't they check their scoring more carefully? Would simply numbering the responses have increased agreement? How about putting the responses and inquiry side by side? Would creating new standards of training and certification help? Should similar research be done with objective test raw protocols?

Did our results arise from using many independent scorers? Can anything be concluded from the use of a single protocol? Were 30 psychologists naïve about the usefulness of the design when they volunteered? Does Footnote 1 imply anything? What happened to the Rorschach Workshops reliability study done by Robert Dies, Ph.D.? Do our results imply anything about the CS coding system itself? Do our results explain the findings of Wood et al (2001a) better than the hypotheses argued by Meyer (2001)? Do they have implications for the ongoing CS renorming project referenced by Meyer? Which of these many questions are rhetorical, which are heuristic, and which are practical?

We recommend that all of these questions be answered with more data, not debate. The next logical research step is to provide a number of protocols of varying quality and complexity to many independent scorers, rather than just pairs of independent scorers, in hopes that our results do not replicate. The data for such a study has already been collected (Guarnaccia et al., 2001), but interrater reliability figures were not reported, since the 33 independent scorers made so many errors. "The error rates were of such magnitude that one cannot but wonder whether there exists a serious problem in the field use of the CS in Rorschach testing." (p.473).

---

<sup>12</sup> One participant suggested that the 27 psychologists who scored the test were "compliant personalities" who were simply "trying to be cooperative."

## References

- Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment, 74*(1), 15-47.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Aronow, E. (2001). CS norms, psychometrics, and possibilities for the Rorschach technique. *Clinical Psychology: Science and Practice, 8*(3), 383-385.
- Exner, J. E. (1988). Scoring issues. *Alumni Newsletter, 4-8*.
- Exner, J. E. (1993). *The Rorschach: A comprehensive system. Vol. I: Basic foundations (3rd ed.)*. New York: Wiley.
- Exner, J. E. (2001). A comment on "The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach". *Clinical Psychology: Science and Practice, 8*(3), 386-388.
- Exner, J. E., & Weiner, I. B. (1991). *The Rorschach: A comprehensive system. Vol. II: Interpretation (2nd ed.)*. New York: Wiley.
- Guarnaccia, V., Dill, C. A., Sabatino, S., & Southwick, S. (2001). Scoring accuracy using the Comprehensive System for the Rorschach. *Journal of Personality Assessment, 77*(3), 464-474.
- Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment, 75*(2), 280-294.
- Hunsley, J., & Giulio, G. D. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice, 8*(3), 378-382.
- McCann, J. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment, 70*(1), 125-144.
- McKinzey, R. K., & Ziegler, T. (1999). Challenging a flexible neuropsychological battery under Kelly/Frye: A case study. *Behavioral Sciences & the Law, 17*, 543-551.
- Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science and Practice, 8*(3), 389-396.
- Mittman, B. L. (1983). *Judges' ability to diagnose schizophrenia on the Rorschach: The effect of malingering*. Unpublished Ph.D. dissertation, Long Island University.
- Nakata, L. M. (1999). *Interrater reliability and the Comprehensive System for the Rorschach: Clinical and non-clinical protocols*. Unpublished Ph.D. dissertation, Pacific Graduate School of Psychology, Palo Alto.

- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS--R, and MMPI-2. *Journal of Personality Assessment, 73*(2), 305-316.
- Weiner, I. B., Exner, J. E., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment, 67*(2), 422-424.
- Widiger, T. (2001). The best and the worst of us? *Clinical Psychology: Science and Practice, 8*(3), 374-377.
- Wood, J., Nezworski, M. T., & Stejskal, W. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*, 3-10.
- Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001a). The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice, 8*(3), 350-373.
- Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001b). Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations. *Clinical Psychology: Science and Practice, 8*(3), 397-402.

Table 1

The protocol

I. IT LL A FOX. (MOST PEOPLE SEE MORE THAN ONE THING...) IT LL A FOX BUT -- THIS SOUNDS STUPID--ACTUALLY LOOK RIGHT HERE. IT LL THESE ARE HANDS LIKE MAYBE THERE ARE 2 WINGED PEOPLE DANCING. IT LL WYLIE COYOTE.

(Fox?) Here are the ears. This is the whole outline of a face and his eyes. It just ll a fox face.

(Hands?) There are the hands and there are the people with their heads thrown back and the wings here (People?) Just they are together but they got their hands raised. They don't have their hands raised, they are together. (Winged?) Here. These ll wings.

(Wylie Coyote?) Just the overall look. If you have ever watched that cartoon how they show him, how gets him how he does his face and this is the way he looks. His ears are hanging and these are the way he looks like "Oh darn, I did it again." (?) The way the ears are so droopy.

II. IT LL 2 DANCING BEARS. THAT'S FUNNY. IT LL 2 DANCING BEARS THAT LOST THEIR HEADS. THAT IS ALL I REALLY SEE. JUST 2 DANCING BEARS THAT DON'T HAVE HEADS.

(2 dancing bears?) Here is the feet. Big heavy butt. And these are paws together but they don't have heads. Their paws are touching. This is the right and this is the left. (Heavy butt?) The way it is hanging down. Just the color, the black. The way they have their paws together and neck together (Neck together?) Just the way the paws are together.

III. THIS LL THE PROGRAM I WATCHED LAST NIGHT--2 NATIVE GIRLS POUNDING ON A DRUM. BUT IT LL A PUPPY TOO. THAT LL A SPLOT. THERE ARE EARS. BUT IT STILL LL NATIVE GIRLS FIRST. IT DOES LL A PUPPY FACE. I JUST NOTICED THE OTHER COLOR.

(2 native girls?) I saw this program last night as soon as I saw this that is what I saw because they have breasts and the drum and that is the way they were. One on each side and because of the color (?) There is no color just the way they had their skirt it looks just like them (Help me...) They are both Black and that stood out immediately as soon as I looked at it. (Drum?) Just the way their arms are down and the picture I had in my mind

(Puppy?) I didn't see it at first because the color is light. These are the ears and nostrils and outside of it. (Color light?) This color is light. When I see color...I see red (Touches D3) but to me it is danger and I know this must be red because I can see it. But I don't see color the way others do. (Help me...) This is the puppy.

(Spot?) It ll it, but these don't belong (D2). This is it (Circles DD99) (Puppy?) The ears. And it is red and black.

IV. BIG FOOT. THAT IS WHAT IT LL. LIKE SOMEBODY YOU ARE TAKING A PICTURE OF BIG FOOT FROM THE GROUND UP, LOOKING UP. I DON'T REALLY

SEE ANYTHING ELSE BECAUSE IN A WAY IT IS SMEARED. NO I DON'T SEE ANYTHING ELSE. IT JUST LL BIG FOOT. WHEN I SPILLED INK IT DON'T LOOK HALF AS INTERESTING AS THIS. IT JUST LL BIG FOOT STANDING IN FRONT OF A TREE.

(Big foot?) It is like you are below--laying down--and taking a picture. Here is the bottom of his feet and his arms and head--leaning down--and this is the tree. It is like he is standing in front of the tree. (Standing in front?) Just I can't see the back of it (?) Just the bottom of his feet are so big and his head is tiny and it is getting smaller as you go up. (Smeared?) It just is sloppy. IDK it just does. The color--the black ink. (Tree?) IDK it just does.

V. A BUTTERFLY OR A BAT. IT LOOKS MORE LIKE A BAT. I LIKE TO DRAW BUTTERFLIES BUT THIS LL A BAT. IT DOESNT HAVE THE ROUND WINGS OF A NICE BUTTERFLY. THAT IS THE ONLY THING I SEE.

(Butterfly?) Or a bat. It depends. More like a bat (?) It is the look and the ears and the wings as if it is flopping

VI. THAT IS A FUNNY ONE. I DON'T SEE ANYTHING BUT AN INK SPLOT. I DON'T SEE MUCH DESIGN WITH THIS ONE. THIS ONE LL ITS A INK SPOT. I DON'T SEE ANYTHING ELSE--NO SHAPE OR NOTHING THAT STANDS OUT. JUST A INK SPLOT A BIG INK SPLOT.

(Ink splot?) Just because of the way it is run everywhere (Run everywhere?) See it is lighter and darker in the center and light and dark again and then there is bands of lighter. (Run?) The way it is run--because I have done this myself.

VII. THIS LL 2 WOMEN DANCING WITH THEIR HEADS COCKEYED. THAT IS WHAT IT LL. SOMEHOW THEY GOT THEIR HEADS SCREWED AROUND BACKWARDS. HEAD. EAR. THAT IS ALL I REALLY SEE. I JUST DON'T SEE ANYTHING ELSE IN THIS ONE.

(2 women dancing?) There is one. Here is the hand and body and skirt. It ll she has a pony tail and it is flying in the air. And here is the face (Skirt?) Just the shape that it would be in and that is where it would be (Heads cockeyed?) Just it is twisted around.

VIII. THERE ARE COLORS. THAT LL AN ANIMAL. THIS ONE LL IT SHOULD BE TURNED THIS WAY (TURNS CARD UPSIDE) IT LL IT IS PICTURE THAT A MIRROR IMAGE HAS BEEN MADE OF. WHAT IT LL IS AN ANIMAL. IT LL IT IS LIKE AN ANIMAL STEPPING FROM THE ROCKS ACROSS THE STREAM. SEE THE ROCKS AND THE ANIMAL STEPPED ACROSS AND THAT IS THE WATER WITH THE STREAM BELOW IT. BUT IT GOT AN UGLY MUG. THAT IS WHAT THAT LL. I JUST DON'T SEE ANYTHING ELSE. IT NEEDS TOO BE TURNED THIS WAY. THAT IS IT.

(colors?) I can just see that there are different colors--different shades. Here is the dark and then a lighter color. But when I looked at it again I see the animal.

(Animal?) This is the way it looks. There is a rock and a rock and a tree and this is the mirror image from the water showing on it. It ll the shore and the animal crossing to get across the lake.

(Water?) The mirror image and the way this ll rocks set up against a creek. And it is so clear it is just the mirror image like you see in the water (Clear?) It is white. See the white around here? You can see the image but it is white around it and the image standing out (Standing out?) There is nothing behind it--there is nothing behind it and it jumps out at you.

(You mentioned an ugly mug?) The face for the animal ll a cross between an armadillo and an anteater and there are no such animal because I draw animals and I know there ain't anything like this.

IX. IT LL A MESS. I DON'T SEE ANY SHAPE. THAT IS WEIRD I DO SEE A SHAPE IN THE BACKGROUND RIGHT HERE. SEE THIS YERE? THAT LL A PROGRAM I SAW THE OTHER NIGHT "SIGHTINGS" THAT SHOWED A MAN AS HE WAS WAKING-- THEY CALLED IT ASTROPROJECTION--LIKE HIS SPIRIT IS MOVING OUT OF HIM. THAT IS WHAT IT LL. THIS MAN STANDING IN THE BACKGROUND LIKE FADED OF SOME COLOR LIKE A SPIRIT MOVING OUT OF ITS BODY AND THAT IS SPLOTCHES IN FRONT OF IT. THIS IN THE BACKGROUND THAT GETS MY ATTENTION. THAT IS WEIRD. WHEN YOU LOOK AT IT CLOSELY YOU CAN SEE IT. SEE IT? THE SPLOTCHES ARE JUST SPLOTCHES.

(Mess?) IDK. This ll a man right here. IDK what the color is but that is different colors. When I first saw it I saw those, but then I saw the man standing there. Here is the man arms and shoulders and head and if you look higher it looks like someone who can take his spirit out of his body. It ll his spirit is raising up. And there is shading or another color that is real light that gives it that affect (Spirit raising up?) Just it isn't a mirror image but it is like there is more to him, like it is coming out of him. Because I felt that-like the spirit has moved out of me. If a think about it, it is the way I feel. It is junny. (?) Because he is smaller than the rest but you can see it here. The arms and body and legs and feet but this other color is over him so it ll he is behind these other colors (?) These colors--they are over him and kinda blot him out.

X. THAT LL A MUSTACHE. AND THAT IS YELLOW--SEE IT IS A FUNNY FACE. IT LL A GOOFY FACE BUT IT ALSO LL--THESE LL CRABS AND THESE LL OTHER KINDS OF MONSTERS--MORE CRABS. BUT THERE IS A FACE WITH A BUNCH OF MONSTERS AROUND IT. THAT IS WHAT I SEE. HIS FACE WITH A BUNCH OF CREEPY MONSTERS.

(Moustache?) It ll an old goofy mustache--handlebar (?) This ll the mustache and nose and eyes but it is yellow around the eyes, like grease paint made up to be a clown (Funny face?) Because of

the hair and the grease paint makes it ll made up for a party (Crabs?) This is the face and the hair and eyes and mustache and the colors around it are like a man got a lot on his mind (?) He has a mustache and he looks worried because of the yellow. The way this color is on here gives it a worried look and the eyebrows are down like he has his face worried and the monsters are at his head--coming at his head.

Table 2

Dr. Doe's Original Scoring

Card	No	Loc	#	Determinants	(2)	Contents	Pop	Z	Special Scores
I	1	WSo	1	Fo		Ad		3.5	
	2	W+	1	Mau	2	H		4.0	INC, COP
	3	Wo	1	Fo		(Ad)		1.0	
II	4	D+	6	FMp.FC'o	2	A, Sx	P	3.0	MOR
III	5	D+	1	Ma.C'Fo	2	H,Id,Sx,Cg	P	3.0	PER, COP
	6	Ddo	99	FY.CF.Mp.C'Fu		Ad, Hx			INC
IV	7	W+	1	FV.Mpo		(H), Bt	P	4.0	
V	8	Wo	1	FMao		A	P	1.0	DR
VI	9	Wv	1	Y		Id			
VII	10	W+	1	Mau	2	H, Cg	P	2.5	PER, INC, COP
VIII	11	Wv	1	Y		Id			
	12	W+	1	Fr.FMa.FC'o		A, Na	P	4.5	PER, MOR
IX	13	W+	1	Mp.FV.CF.FY.Fr-		H, (H)		5.5	FAB2, MOR, PER
X	14	W+	1	CF.Mp-		Hd,A,(H)	P	5.5	FAB2

Tables continue.



Table 3

Constellation Scorings of 20 Scorers and Dr. Doe

	Scorers in present study				Dr. Doe		
	Protocols			Min	Mdn	Max	Positive or
	With Positive Constellation (%)						Negative
SCZI	9 (45)	1	3	5	Positive	4	
DEPI	7 (35)	3	4	6	Positive	5	
CDI	0 (0)	0	2	3	Negative	1	
S-CON	1 (5)	4	5	8	Negative	6	
HVI	2 (10)	2	3	6	Negative	4	
OBS	0 (0)	0	0	1	Negative	0	

  

Scorers	Number (%)	Dr. Doe
SCZI POSITIVE, DEPI POSITIVE	3 (15)	SCZI POSITIVE, DEPI POSITIVE
SCZI POSITIVE, DEPI NEGATIVE	6 (30)	
SCZI NEGATIVE, DEPI POSITIVE	4 (20)	
SCZI NEGATIVE, DEPI NEGATIVE	7 (35)	

Tables continue.

Table 4

Agreement among 20 scorers and Dr. Doe for Rorschach CS variables

Variable	Min	25 <sup>th</sup> %ile	Mdn	75 <sup>th</sup> %ile	Max	Dr. Doe
R	14	14	15	15	16	14
W	10	11	11	12	12	11
D	1	1	2	2	4	2
Dd	0	1	2	2	3	1
Space	0	1	2	2	4	1
M	3	4	6	7	8	7
FM	1	3	4	4	5	3
m	0	0	0	1	2	0
FC	0	1	2	2	4	0
CF+C	0	2	2	3	4	3
Pure C	0	1	1	2	3	0
WSumC	1.0	2.8	4.0	4.5	5.5	3.0
SumC'	0	3	4	5	5	4
SumT	0	0	0	0	1	0
SumV	0	0	1	1	2	2
SumY	1	2	3	3	4	4
Sum Shading	2	2	3	4	6	6
Fr + rF	1	1	1	1	1	2
Pairs	2	4	4	4	5	4
3r+(2)/R	0.31	0.42	0.47	0.50	0.57	0.71
Lambda	0.07	0.11	0.15	0.21	0.36	0.17
EA	6.0	7.8	9.0	11.2	12.5	10.0
es	9	10	11	12	16	13
D Score	-3	-1	0	0	0	-1
AdjD	-1	0	0	0	1	0
a (active)	5	6	7	8	8	5
p (passive)	0	2	4	4	5	5
M <sup>a</sup>	2	4	4	5	6	3
M <sup>p</sup>	0	2	2	3	4	4
Intellect	0	0	2	4	6	0
Zf	9	11	11	11	12	11
Zd	-4.5	0.2	2.2	3.0	6.5	3.0

Tables continue.

Variable	Min	25 <sup>th</sup> %ile	Mdn	75 <sup>th</sup> %ile	Max	Dr. Doe
Blends	5	6	8	8	9	7
Afr	0.27	0.36	0.38	0.42	0.60	0.40
Populars	2	5	5	6	7	7
X+%	0.33	0.48	0.53	0.60	0.64	0.50
F+%	0.50	0.50	0.71	1.00	1.00	1.00
X-%	0.13	0.17	0.20	0.27	0.40	0.14
Xu%	0.00	0.00	0.13	0.17	0.33	0.21
S-%	0.00	0.00	0.00	0.21	1.00	0.00
Isolate/R	0.00	0.13	0.16	0.20	0.25	0.21
H	2	3	4	4	5	4
(H)	0	1	2	3	4	3
Hd	0	0	1	1	2	1
(Hd)	0	0	0	0	2	0
Hx	0	0	0	1	2	1
A	2	3	4	4	5	4
(A)	0	0	0	1	2	0
Ad	1	2	2	4	4	2
(Ad)	0	0	1	1	2	1
An	0	0	0	0	0	0
Art	0	0	0	2	4	0
Ay	0	0	0	0	1	0
Bl	0	0	0	0	0	0
Bt	0	1	1	1	1	1
Cg	0	1	2	2	2	2
Cl	0	0	0	0	1	0
Ex	0	0	0	0	0	0
Fi	0	0	0	0	0	0
Food	0	0	0	0	0	0
Ge	0	0	0	0	0	0
Hh	0	0	0	0	2	0
Ls	0	0	0	1	1	0
Na	0	0	1	1	1	1
Sc	0	0	0	1	2	0
Sx	0	0	1	1	3	2

Tables continue.

Variable	Min	25 <sup>th</sup> %ile	Mdn	75 <sup>th</sup> %ile	Max	Dr. Doe
Idio	0	1	2	4	5	3
Sum 6 Sp.Sc.	1	6	7	10	12	6
WSum6	2	13	26	32	47	23
AB	0	0	0	1	2	0
AG	0	0	0	0	1	0
COP	0	0	2	3	4	3
CP	0	0	0	0	0	0
MOR	0	1	1	2	3	3
PER	0	2	3	4	6	4
PSV	0	0	0	0	1	0

Tables continue.

Table 5

Agreement among 20 scorers and Dr. Doe on 5 interpretable protocol level indexes

Scorers			Dr. Doe's scoring
Index	Number	Percentage	
<b>No. of Responses</b>			
R = 14	6	30	R = 14
R = 15	12	60	
R = 16	2	10	
<b>EB Style</b>			
Introversive	11	55	EB Style = Introversive
Super introversive	2	10	
Ambitent	8	40	
Extratensive	1	5	
Super extratensive	0	0	
<b>D scores</b>			
D score > 0	0	0	D Score = -1
D score = 0	14	70	
D score = -1	4	20	
D score < -1	2	10	
Adj D score > 0	2	10	Adj. D = 0
Adj D score = 0	15	75	
Adj D score = -1	3	15	
Adj D score < -1	0	0	
<b>Form Quality</b>			
X+% > .89	0	0	X+% = 0.50
X+% < .70	20	100	
X+% < .61	16	80	
X+% < .50	5	25	
F+% < .70	10	50	F+% = 1.00
Xu% > .20	2	10	Xu% = 0.21
X-% > .15	15	75	X-% = 0.14
X-% > .20	9	45	
X-% > .30	1	5	
<b>FC:CF+C ratio</b>			
FC > (CF+C) + 2	0	0	FC = 0; CF+C = 3
FC > (CF+C) + 1	3	15	(i.e. CF + C > FC + 2)
FC = (CF+C)	5	25	
(CF+C) > FC + 1	6	30	
(CF+C) > FC + 2	1	5	

Table 6

Agreement of 20 scorers and Dr. Doe on 26 interpreted variables as applied to CS cutoffs

Scorers			Dr. Doe's scoring
Index	Number	Percentage	
Lambda > .99	0	0	Lambda = 0.17
Dd > 3	0	0	Dd = 1
S > 2	5	25	S = 1
SumT = 0	19	95	SumT = 0
SumT > 1	0	0	
3r+(2)/R < .33	1	5	3r+(2)/R = 0.71
3r+(2)/R > .44	14	70	
Fr + rF > 0	20	100	Fr + rF = 2
PureC > 0	17	85	PureC = 0
PureC > 1	6	30	
Afr < .40	10	50	Afr = .40
Afr < .50	16	80	
(FM+m) < Sum Shading	5	25	FM = 3; m = 0; Sum Shad. = 6
(2AB+Art+Ay) > 5	2	10	2AB+Art+Ay = 0
Populars < 4	2	10	Populars = 7
Populars > 7	0	0	
COP = 0	6	30	COP = 3
COP > 2	6	30	
AG = 0	17	85	AG = 0
AG > 2	0	0	
MOR > 2	1	5	MOR = 3
Level 2 Sp.Sc. > 0	13	65	Level 2 Sp.Sc. = 2
Sum6 Sp.Sc. > 6	15	75	Sum6 Sp.Sc. = 6
Pure H < 2	0	0	Pure H = 4
p > a + 1	0	0	p = 5; a = 5
Mp > Ma	1	5	Mp = 4; Ma = 3