

## **THE RORSCHACH: TIME TO THROW OUT THE BATH WATER**

**Edward Aronow, Ph.D.  
Montclair State University  
USA**

First, a little bit about my Rorschach views and myself. Those of you who know my approach to the Rorschach know that in my three books I champion the content, idiographic, and psychodynamic approach to the Rorschach. For a variety of reasons, I consider this information primary in using the technique. However, I also use some traditional scoring though I view this as secondary or supplementary information. The Comprehensive System and other scoring-based approaches, as I currently understand it, take the opposite view, seeing scoring as primary, but also paying attention to the projective use of the instrument. One might therefore say that the difference between my approach and the Comprehensive System approach is then primarily a difference of emphasis.

As many of you may know, in America, the Rorschach has recently come under sustained attacks from an organized group of academics, culminating last year in an article in the respected journal Scientific American, which questioned whether the Rorschach may be just “a waste of ink.” I am not going to go into detail about these critiques of the technique. We should be aware that their Rorschach critiques are finding their way into more and more psychology textbooks in such areas as introductory psychology, abnormal psychology, and tests and measurements. Like many of you, I have dedicated much of my life to the Rorschach Technique, and I am quite sure that it

continues to be an extraordinary instrument, the “king of the projective techniques.” I do believe however, that we could better dialogue with these Rorschach critics. On occasions when I have done so, I have seen some movement in a positive direction.

To get to the main drift of my talk today, I believe that one way in which we play into the hands of Rorschach critics is by inadequately examining the validity data on Rorschach scoring categories. We thus put ourselves in the position of recommending certain interpretations of scoring categories even decades after such inferences have been empirically shown to be incorrect, or insufficient data are available to support hypotheses. That is indefensible with an instrument that has been around for over eighty years, and this certainly cannot be considered a scientific approach. I have been encouraged to see that in America work within the Comprehensive System group has also been moving toward re-examining those variables, which may have questionable validity.

In 1976, my co-author Marvin Reznikoff and I set out to examine the English language validity data on Rorschach content variables. In approximately 300 studies that we surveyed, most content scoring categories for the Rorschach held up very well empirically - this was particularly true for those scoring categories with a psychoanalytic basis. For the last two years, several co-workers and I have set out to do the same for other traditional Rorschach scoring categories, specifically, location, determinants, form-level, populars, and what many have come to call “pathological verbalizations.” This last category was not one proposed by Hermann Rorschach, but has nonetheless become part

of traditional Rorschach lore. In this validity survey that I am now reporting on involving over 600 English language publications, we were to some extent updating and expanding on the work of Dr. George Frank of, who in the 1970's and early 1990's likewise published validity surveys.

With respect to the categories of form level, populars, and pathological verbalizations as seen before with content, the evidence is generally quite favorable. This is consistent with what was found in George Frank's validity surveys. The American and Japanese delegations to this conference, among others, will understand what I mean when I say that this would imply that the Rorschach Technique is generally batting .667 - fully 2/3 of these scoring categories hold up in the empirical research. That is not bad at all when one considers that Hermann Rorschach died very young, not having many decades to refine his vision, as did many other theoreticians such as Freud who tinkered with, changed, and matured in his understanding of his topic until he reached a ripe old age.

First with respect to the category of content scoring. As I have already noted and summarized in my first book, this category holds up very well in the validity research. With respect to the category of form-level, while a number of systems have been proposed for this, across systems, it is possible to state that as one might expect it is a valid measure of what many have called "reality testing," i.e., the ability to perceive the world accurately. Thus, form-level does particularly well in making diagnostic

differentiations between, for example, schizophrenics and non-schizophrenics, psychotics and non-psychotics, chronic and non-chronic schizophrenics, etc.

With respect to the “populars” category – while there is not an extensive literature - again, the evidence is favorable, even across different “popular” lists. Popular responses do indeed appear to measure an individual’s ability to share in the commonality of his culture’s perception, as shown by differences seen between generally normal and abnormal populations.

With regard to the category of “pathological, verbalizations” - first suggested formally by Rapaport and co-authors in 1948 - the research demonstrates that various diagnostic groups - though not just schizophrenic - do indeed show a higher frequency of these indices of thought disorder and disordered thinking. While this area is complex, and involves different types of disordered thinking - the general hypothesis that “pathological verbalizations” are a reflection of disordered thought is generally well supported.

I am going to save the bulk of my talk for the two categories of traditional scoring which do not hold up as well in the validity literature, and which we would do well to further scrutinize before recommending that our students base interpretive statements on them.

First with respect to the “location” category. There is a clear hypothesis what

white space responses may reflect either oppositionalism or hostility. While the validity data are not entirely unidirectional - they do strongly suggest that these hypotheses are not correct with respect to white space responses. With the “W” or whole response- while much more could be done in the way of research, there is some support for the notion that there is a small relationship between “W” responses and intelligence. Unfortunately, there are few studies, which should light on the large or small detail scorings.

In short, one would have to conclude that the usual interpretations about locations are either incorrect, reflective of weak relationships – or not yet explored by research – thus generally not usable. Students should be made aware of this fact.

However, perhaps the most crucial examination of validity data concerns the “determinants” category, the heart of traditional Rorschach scoring. A number of hypotheses have traditionally been associated with the determinants categories, including such assertions as that use of shading reflects anxiety, the use of color is related to affect and emotionality, achromatic color responses can be related to depression, and human movement responses are related to many positive aspects of functioning such as creativity and intelligence.

While again with so many studies the results are not wholly unidirectional - with the exception of the human movement response - the general summary of relevant validity studies paints a dismal picture. The determinant hypotheses are usually quite

unsupported by the empirical literature. In volumes 1 of his 1993 text, The Rorschach: A Comprehensive System,” John Exner has likewise stated that the determinants do not have “exacting correlations with behavior or with personality characteristics”, i.e. with external criteria. While the M response is related to many “positive” aspects of functioning - the human content response is similarly related to such criteria. It is thus yet unproven that the human movement response represents more than just the fact that a person was seen, not in itself a determinant category.

Thus, to sum up of the six traditional scoring categories surveyed, four hold up well - hardly a stinging indictment of the validity of the Rorschach. However, two categories show at the very least questionable validity. It is therefore incumbent on us to either demonstrate their validity, even at this late date - or cease using them.

I would also like to address the issue of how this could have happened – that after more than 80 years we could still be using invalid categories of scoring such as determinants.

Three possibilities have come to my mind in this connection. At least to some degree, the Rorschach has become immune to the validity data. When George Frank returned in the early 1990’s to update his validity surveys – he marveled at how little impact his validity reviews had had on the field. After all these decades we have not dropped even one of Hermann Rorschach’s main scoring categories. This suggests that

either Dr. Rorschach was an incredible genius – or that there are factors preventing us from responding to the dictates of the research literature.

First, many Rorschach hypotheses have truly superb “face validity” – that is to say, they look valid, more a matter of public relations than anything else. For example, the white space response reverses figure and ground – so why shouldn’t the subject who gives this response be considered oppositional? Similarly, colors have long been associated with certain emotional responses. That is the reason that advertisements for sports cars usually show them in the color red. Again, black in many cultures is often associated with mourning – so why shouldn’t a black color response be associated with depression? The problem with this line of reasoning is that, as psychometricians know, “face validity” is not a real psychometric validity at all. To be valid, a test-based measure must demonstrate relationships with relevant external criteria.

It seems to me that a second reason for our failure to drop long-standing scoring categories is our unconscious desire for an infallible, God-like guide to show us the way, as we viewed our parents in early childhood. Many may well view Hermann Rorschach and others who have taken up his mantle this way. I believe that this is part of the human condition, and I have seen it in other areas of psychology. Followers of psychoanalysis may well view those who may disagree with certain tenets of Sigmund Freud not only as wrong but also as heretics. I think that we must guard against any tendency of the Rorschach field to become some kind of religious cult as well.

A third reason that I see for our continuing to use invalid indices is that they have aspects, which actually are valid in some way, shape, or form. Thus, as I have mentioned, the human movement response does in fact have external correlates, although I suspect these correlates may be at least in part an artifact. Another example is the ratio of pure color responses to color responses that are more or less form dominated. There is no question but that individuals with deficient ego functioning give percepts that are more amorphous. However, this is likely true across the spectrum, and not just with color responses. It remains to further research to tease out the valid components of suspect general categories.

If in fact the determinant categories are not valid – does not have external correlates – this may well open the door to a more time-effective Rorschach than is currently the case. With a brief inquiry which ascertains location (necessary for the scoring of certain clearly valid categories) this could well cut in half the time required for the Rorschach as part of a test battery. In a field of psychology in America that I am very familiar with – psychology in the schools – the Rorschach has become largely extinct, mostly because of the time requirements. One might then have the choice of giving the briefer Rorschach – or a longer Rorschach depending on the circumstances. If one had the time for a lengthier Rorschach – I personally would favor using the extra time for associations, as one does with dreams, and I have suggested in my books. Other more traditional Rorschachers might feel otherwise about the “long” version.

As I have asserted in my books – I favor a collaborative relationship with patients together using their associations to Rorschach percepts to help figure out their dynamics, and to break logjams in the psychotherapy. I suspect that the Rorschach in coming decades will gradually become more of a psychotherapy technique, and less just evaluation. Nonetheless, when the Rorschach is “scored” in traditional fashion, we can no longer continue business as usual: we can only use indices, which have external validity.