



## Roots of the Rorschach controversy<sup>☆</sup>

Howard N. Garb<sup>a,\*</sup>, James M. Wood<sup>b</sup>, Scott O. Lilienfeld<sup>c</sup>, M. Teresa Nezworski<sup>d</sup>

<sup>a</sup>*Psychology Research Service, Wilford Hall Medical Center, Lackland Air Force Base (AFB),  
59 MDOS/MMCPR, 2200 Bergquist Drive, Suite 1, San Antonio, TX 78236-5300, United States*

<sup>b</sup>*University of Texas at El Paso, United States*

<sup>c</sup>*Emory University, United States*

<sup>d</sup>*University of Texas at Dallas, United States*

Received 14 January 2004; accepted 27 September 2004

---

### Abstract

The controversy surrounding the Rorschach is updated, and an analysis of its dynamics is offered. Results on normative data and validity are reviewed, followed by a summary of, and rebuttal to, arguments made by Rorschach advocates. We argue that the current controversy can be traced, at least in part, to two unwarranted beliefs. First is the belief that informal impressions and popularity provide dependable evidence for evaluating validity. Second is the belief that Rorschach scores with low individual validity are likely to yield much higher levels of validity if they are interpreted in combination with each other, or with other sources of information, by experts. After presenting historical background information, we show how several arguments made recently in defense of the test reflect these two beliefs, even though they are contradicted by research findings. We conclude that a variety of other divisive conflicts in clinical psychology are related to the inappropriate weight placed on informal and unsystematic impressions relative to systematic research.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Assessment; Methodology; Judgment; Rorschach

---

<sup>☆</sup> The views expressed in this article are those of the authors and are not the official policy of the Department of Defense or the United States Air Force.

\* Corresponding author.

*E-mail address:* [howard.garb@lackland.af.mil](mailto:howard.garb@lackland.af.mil) (H.N. Garb).

In recent years, critics have voiced numerous concerns about the Comprehensive System (CS; Exner, 1991, 1993, 2001b), the most popular system for administering, scoring, and interpreting the Rorschach Inkblot Test. According to critics, the CS norms are in error and can make relatively normal individuals appear to have serious psychopathology (e.g., Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b; see also Hamel, Shaffer, & Erdberg, 2000; Shaffer, Erdberg, & Haroian, 1999). Questions have also been raised about the validity of many commonly used CS indices (e.g., Hunsley & Bailey, 1999; Lilienfeld, Wood, & Garb, 2000). Arguments have been made that Rorschach-based expert testimony should not be admissible in court (Grove & Barden, 1999; Grove, Barden, Garb, & Lilienfeld, 2002; cf., Ritzler, Erard, & Pettigrew, 2002a, 2002b), and recent surveys indicate that the use of the test is declining among forensic psychologists (Lally, 2003; Tolman & Mullendore, 2003).

Much has been written in response (e.g., Allen & Dana, 2004; Hibbard, 2003; Perry, 2003; Weiner, Spielberger, & Abeles, 2003). In this article, the responses of Rorschach advocates will be examined and evaluated. By doing so, we intend to bring readers up to date concerning recent developments in the Rorschach controversy. In addition, broader conceptual issues are addressed, some of which cut to the heart of clinical psychology. As Widiger (2001, p. 374) observed, “The dispute over the Rorschach may be a flashpoint of a wider and more fundamental dispute within the profession.” Psychologists hold disparate views on the importance of (a) informal clinical impressions and (b) findings from systematic clinical studies (Borkovec, 2004; Levant, 2004; Nathan, 2004; Peterson, 2004), and these differences may underlie differences in views on the CS. For example, in the *New York Times* (Goode, March 9, 2004, D1), the president-elect of the American Psychological Association (Ronald Levant) accused Rorschach critics and some other psychologists as “being overly devoted to the scientific method.” The dispute over the CS may not be resolved until psychologists come to terms with underlying issues such as the proper role of informal versus systematic observation in clinical psychology.

## 1. The controversy

The controversy surrounding the CS has elicited strong emotions. For example, three distinguished Rorschach proponents—including the current president of the Society for Personality Assessment and two past presidents of the American Psychological Association—have gone so far as to publicly compare recent criticisms of the test to the “burning of books” (Weiner, Spielberger, & Abeles, 2002, p. 11).

The same three distinguished Rorschach proponents also accused critics of seeking to impose “a death penalty for the teaching and use of methods that do not pass their muster” (Weiner et al., 2002, p. 11). This accusation not only portrays Rorschach critics as enemies of academic freedom, but it distorts what they have said. In reality, no one has made a recommendation to ban the teaching of the Rorschach, although we have made a distinction between teaching and training, and we have recommended that training in the Comprehensive System for the Rorschach be eliminated.

*Teaching* students about the Comprehensive System is much different from *training* them in its use. To oversimplify matters only slightly, teaching about the test encourages students to become scientific and appropriately critical thinkers, whereas training encourages them to join the ranks of technicians who all-too-frequently interpret Rorschach scores with a worrisome blend of certainty and credulity. (Wood, Nezworski, Lilienfeld, & Garb, 2003, p. 280)

Likewise, none of the authors of this article has suggested that a “death penalty” be imposed on the use of the Rorschach. To the contrary, in our book we explicitly endorsed the use of Rorschach scores that are reliable, well validated, and adequately normed (Wood, Nezworski, & Garb, 2003; Wood, Nezworski, Lilienfeld, et al., 2003). Perhaps Rorschach defenders are reacting to an article in which one of us (Garb, 1999), echoing Lee J. Cronbach (1955), called for a moratorium on many applications of the Rorschach, until further research demonstrates which scores are valid for which tasks. But a moratorium is not a “death penalty.” By definition, moratoriums are intended to be temporary.

The controversy surrounding the Rorschach raises important questions for the field of clinical psychology. What should the field do when questions are raised about the use of an assessment instrument? If evidence exists that the textbook use of a test is potentially harmful, should there be a moratorium on its use?

To help readers think about these questions, the Rorschach controversy will be described. We will review research on two topics: (a) adequacy of norms and (b) validity of scores. To provide necessary background, we will describe findings that have been reviewed previously. However, for the first time, we will respond to many of the criticisms made by Rorschach advocates.

### *1.1. Normative data*

Norms were a persistent problem for the Rorschach until the late 1970s. The leading Rorschach figure, Bruno Klopfer, was openly contemptuous of the need for norms (Klopfer & Kelley, 1946, p. 21) and refused to include them in his highly popular Rorschach system. The less popular systems of Samuel Beck and David Rapaport included norms, but their normative samples were small and unrepresentative (e.g., see critique by Hertz, 1959).

In contrast to these early systems, the CS has repeatedly been praised for its norms (e.g., see Anastasi, 1982). Since the late 1970s, Exner’s (1978, 1986, 1991, 1993, 2002b) books have provided dozens of densely printed pages that provide extensive normative statistics for nonpatient adults, children of various ages, and assorted patient groups.

Many psychologists were surprised, therefore, when in 2001 they learned of unexpected problems with the CS norms. In his books of the early 1990s, Exner (1991, 1993) had stated that the CS norms were based on a sample of 700 adults. However, in 2001 he published a clarification (Exner, 2001b, p. 172) revealing that these norms were actually based on the protocols of only 479 individuals, and that the scores for 221 of them had accidentally been counted twice, mistakenly swelling the total to an illusory 700.

To rectify these errors, Exner (2001b) created new CS norms by adding 121 more Rorschach protocols to the 479 already in his sample, bringing the total number to 600. However, these reconstituted norms also suffered from several shortcomings. First, the 600 protocols in the 2001 sample had been collected using convenience sampling strategies, not probability sampling strategies (Hunsley & Di Giulio, 2001). Second, the 600 protocols (including the 121 “new” protocols) had all been collected 15–25 years previously, during the 1970s and 1980s, and thus could be criticized for being out of date. Third, the 600 protocols had been scored in the 1980s with the CS scoring rules current at that time, but apparently were not re-scored with updated rules when the reconstituted norms were published in 2001 (Hibbard, 2003, p. 261; Meyer & Richardson, 2001). In other words, the 2001 norms for some important CS variables (e.g., ordinary and unusual form level) were based on old scoring rules that psychologists had stopped using more than a decade earlier.

Although Exner (2002a) announced that he is collecting data for a new set of norms, the reconstituted figures he published in 2001 still constitute the official CS norms (see Exner, 2002b). Rorschach proponents seem to disagree whether he should have re-scored the old protocols in his sample with updated rules before releasing these new norms. For example, Meyer & Archer (2001, p. 495) argued that “Whenever scoring modifications are introduced it is essential to rescore the reference protocols.” In contrast, while acknowledging that Exner did not rescore the protocols and that the norms for some CS variables were inaccurate to “some unknown extent,” Hibbard (2003, p. 261) contended that it would be an enormous task to rescore 600 protocols.

Aside from the issues just described, another important problem with the CS norms has come to light in recent years: Converging evidence from numerous laboratories indicates that the CS norms are in error and tend to make normal individuals appear psychologically disturbed. In a number of studies, investigators have administered the Rorschach to relatively normal groups of children or adults and then compared the results with the CS normative data. For many critical CS scores, the results for the children and adults have differed markedly from the CS norms (Hamel et al., 2000; Shaffer et al., 1999; Wood et al., 2001a, 2001b; but also see Meyer, 2001).

For example, in one study (Hamel et al., 2000) the Rorschach was administered to children recruited from a school who had no known history of mental health problems. They were healthier than average according to a well-validated measure (the Conners Parent Rating Scale-93; Conners, 1989). Yet when these children were compared with CS norms, the results wrongly indicated that “their distortion of reality and faulty reasoning approach psychosis” (Hamel et al., 2000, p. 291) and that many or most of the children were probably suffering from serious mood problems.

The research findings on the CS norms have been disputed by several leading Rorschach proponents. We will devote considerable space here to discussing their arguments and our own responses to them.

*Argument 1: The CS normative sample is “above average” in psychological functioning.* To explain why apparently normal adults and children in the community appear psychopathological when compared with the CS norms, several Rorschach proponents have argued that the members of the CS adult normative sample enjoyed superior psychological functioning and were healthier than the general US nonpatient population (Hibbard, 2003; Meyer, 2001; Meyer & Archer, 2001; Weiner et al., 2003). According to this argument, because there are some nonpatient individuals in the community with poor mental health, the average results for nonpatients in general should be expected to look “sick” compared with the CS norms.

In our opinion, there are three reasons why this argument is unconvincing. First, it appears to be a post hoc attempt to explain away unwanted research findings. Specifically, although the norms were published in 1991, Rorschach proponents did not claim that the members of the normative sample were healthier than most nonpatients until 1999, when researchers began to uncover the problems with Exner’s numbers (e.g., Shaffer et al., 1999). If the normative sample was truly healthier than most nonpatients, one might reasonably expect Exner and other CS proponents to have unambiguously said so before 1999.

Second, there is no solid evidence to substantiate claims that the members of the CS normative sample were different from other American nonpatient groups in respect to psychological functioning. Specifically, according to Exner’s (1991, 1993) books, members of the normative sample were never administered diagnostic interviews or psychological tests (other than the Rorschach) to measure their level of psychological functioning, nor were they compared with other

nonpatient groups in systematic studies. Because such well-validated, non-controversial data are lacking, Rorschach proponents have only comparatively weak evidence to support their speculations regarding the supposedly superior psychological health of the normative sample. For example, Meyer (2001, p. 390) argued that because many members of the normative sample held jobs or belonged to clubs such as the Audubon society, they were likely to be above average in functioning compared with other groups of nonpatients.

Third, to support their argument that the CS normative sample is above average in functioning, proponents claimed that “Exner’s (1993) nonpatient reference sample consists of people with no history of mental health treatment” (Meyer, 2001, p. 390). Nevertheless, this claim, which has often been repeated by Rorschach proponents (e.g., Hibbard, 2003; Meyer & Archer, 2001; Weiner et al., 2003), is incorrect according to information we obtained from John Exner. Because descriptions of the CS normative samples have been vague, we wrote Exner a letter asking for clarification. He replied that, “In fact, to the best of my knowledge, between 80% and 85% of our sample have no psychiatric/psychological history whatsoever” (John Exner, personal communication, February 6, 2001). After being asked about the remaining 15% to 20%, he clarified that they had sought professional help for a range of problems including academic difficulties, occupational decisions, pastoral counseling, family difficulties, and grief counseling (John Exner, personal communication, March 29, 2001). In other words, the “nonpatients” in the CS normative sample had never been psychiatrically hospitalized, but a non-trivial proportion had sought professional help at some time in the past.

*Argument 2: So-Called “Non-Patient” Samples Included Psychiatric Patients.* Rorschach advocates have taken special aim at a literature review by the present authors (Wood et al., 2001b) in which we compiled findings from 32 Rorschach studies of nonpatient adults. We found marked discrepancies between results for the nonpatient adults and values listed for the CS norms, and we concluded that the norms are in error and can lead psychologists to overperceive psychopathology.

Several articles by Rorschach advocates (e.g., Meyer, 2001; Weiner et al., 2003) have claimed that our review (Wood et al., 2001b) was seriously flawed because it included psychiatric patients among its supposedly “non-patient” samples. For example, Weiner et al. (2003, p. 8) claimed that, “five of the Wood et al. samples included current or former psychiatric patients. . . .”

Of course, it would be an important matter if the “non-patient” samples in our review really did include psychiatric patients, as Meyer (2001) and Weiner et al. (2003) have claimed. Specifically, the inclusion of psychiatric patients would partially explain why these five samples appeared pathological when compared with the CS norms. Nevertheless, this allegation is in error. In the five studies that they cite there is no indication that any of the participants had had a psychiatric hospitalization or been on a psychiatric unit.

Why did Meyer (2001) and Weiner et al. (2003) claim that these five samples included psychiatric patients when in fact they did not? Apparently these Rorschach proponents considered any participants who had ever been in psychotherapy to be “psychiatric patients.” Four of the five studies named by the proponents included some individuals who were or had been in therapy, just as Exner’s normative sample did. For example, in one of the studies (Schiff, 1992/1993) participants were mental health professionals with MDs or PhDs who were in psychoanalytic psychotherapy as part of their training. We considered it entirely appropriate to include this group among the nonpatient samples in our review. Yet Meyer and Weiner et al. criticized us for including this sample on the grounds that it contained “current or former psychiatric patients.”

*Argument 3: Normative Discrepancies Are Due to Culture or Ethnicity.* Allen and Dana (2004) also criticized the Wood et al. (2001b) literature review. They argued that:

...the design of the Wood, Nezworski, Garb, et al. study does not allow the researchers to conclude that any observed differences in CS scores, when compared to the existing CS norms, are not instead due to cultural differences tapped by CS variables that underlie the ethnic group status of a number of the participants. (p. 192)

That is, several samples in the Wood et al. review were composed entirely of members of special cultural, occupational, or age groups, such as African Americans, Hispanic Americans, college students, or psychoanalytic trainees. According to Allen and Dana, cultural differences between these samples and the CS normative sample may have accounted for the discrepancies from the CS norms described in our review.

Allen and Dana (2004) are right to point out that cultural differences may exert a substantial impact on some CS scores (see Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Wood & Lilienfeld, 1999). However, such cultural influences cannot account for the discrepancies from the CS norms identified in our review (Wood et al., 2001b) and other recent studies, for the simple reason that these discrepancies have consistently been found in the large majority of relevant studies regardless of gender, culture, ethnicity, or nationality. In other words, these discrepancies extend generally across groups, and are not limited to particular cultural, demographic, or occupational groups.

This point can be illustrated by examining the numbers for Distorted Form ( $X\%$ ), an indicator of psychotic thinking. According to the norms published by Exner (1991, 1993, 2001b) in the 1990s and early 2000s, the average adult score for  $X\%$  was 0.07, with a standard deviation of 0.05. In our review (Wood et al., 2001b) we identified 15 samples of American nonpatients for which  $X\%$  had been reported. The  $X\%$  score was above 0.07 in every single one of these samples, with an overall mean of 0.19 (that is, about two and a half standard deviations above the Exner norms). The mean  $X\%$  score was 0.20 or higher in diverse samples, including normal unmarried women (Burns, 1993/1994), Navy and Army Air Force staff (DeLucas, 1997), undergraduates (Greenwald, 1990), middle-aged men who had never been married (Waehler, 1991), and police applicants (Zacker, 1997). In addition, our findings were virtually identical to those reported in samples of non-Americans. Specifically, Gregory Meyer (personal communication, February 8, 2001) informed us that for an international sample of 2125 nonclinical participants from nine different countries, mean  $X\%$  was 0.19, which was the exact figure we found for our American samples.

Thus, similar discrepancies from the CS norms have been reported again and again for  $X\%$  in both U.S. and international samples. The same is true for many other CS variables. When findings converge in this manner, it is virtually impossible to attribute them to the cultural composition of particular groups. Instead, the findings indicate that virtually all groups differ from the Exner norms and that there is a problem with the norms themselves.

*Argument 4: Psychopathology is becoming more prevalent.* Other Rorschach proponents have defended the CS norms by speculating that people in the community have become more pathological over time (Hibbard, 2003; Meyer, 2001). The 1993 and 2001 CS norms were based on protocols collected in the 1970s and 1980s. Thus, if psychopathology has been increasing over time, one would

expect recent community samples to appear more pathological on the Rorschach compared with these older CS norms. As Meyer (2001) conjectured:

It is possible that scores have changed because people have genuinely changed. Although Wood et al. (2001b) did not address the issue, research suggests psychopathology has increased over time. . . . Accordingly, a valid measure of pathology should track those changes to show somewhat increased rates of mental health problems. (p. 389)

In considering such arguments by Rorschach proponents, it is helpful to spell out their full implications. As already noted, striking discrepancies from the CS norms have been reported for adults and children, both within the US and internationally. Some of these discrepancies are very large (i.e., more than 2 standard deviations) and involve scores that ostensibly measure a broad spectrum of psychological problems, including psychosis, thought disorder, depression, anxiety, and narcissism. This suggests, for example, that on a measure of thought disorder the average person today would score in the upper four percent when compared with a group of individuals tested in the 1980s and 1990s. In their attempt to explain the problems with the CS norms, Rorschach proponents are in the position of claiming that there has been a huge upsurge in the prevalence of psychotic, mood, anxiety, and personality disorders.

The greatest problem with this notion, of course, is its lack of firm empirical support. If a psychiatric pandemic has been underway for the past 20 or 30 years, one would expect to have heard a great deal about it from scientific journals, public health officials, and the media. For example, if the level of psychotic symptoms and thought disorder has truly increased more than two standard deviations over the past 25 years (as comparison with the CS norms seems to indicate), one would expect a flood of psychiatric patients to have poured into state hospitals and community mental health centers. Although numerous examples of abnormal behavior have been reported in the media during the past 25 years, there is little or no scientific evidence of a surge in psychotic disorders.

Rorschach advocates (e.g., Meyer, 2001, p. 389) have cited several epidemiological articles that suggest that depression and anxiety may have increased in certain European and American groups over the past 30–50 years (Fombonne, 1994, 1998; Kelleher et al., 2000; Swindle et al., 2000; Twenge, 2000; also see Diener & Seligman, 2004; Grof, 1997). Nevertheless, the widespread claim that depression and similar disorders are increasing in prevalence is scientifically controversial and has been contested on methodological grounds. In particular, the findings are susceptible to recall bias because older individuals may be less capable of recalling depressive episodes in their youth compared with younger individuals (see Guiffra & Risch, 1994; Parker, 1987).

This recall bias may produce an illusory increase in the prevalence of depression across the 20th century. Moreover, the results of several recent investigations offer little evidence for an increase in depression over time. In a large-scale epidemiological study that controlled for recall bias by inquiring about current depression, Murphy, Laird, Monson, Sobol, and Leighton (2000) found no evidence of an overall increased prevalence of depression from 1952 to 1992. Therefore, whether depression is increasing in prevalence is still the subject of study, and the research is more ambiguous than implied by Meyer (2001).

Finally, even if one were to accept the controversial supposition that an outbreak of psychopathology has recently swept over the Western world, one would still have reason to reject the idea that it has had

an impact on Rorschach scores. Exner (2002a) has recently begun collecting new normative data for the CS, and he has reported that preliminary results show minimal changes from his previous normative samples. One can hardly defend the CS norms by saying that psychopathology has increased over time, if the new normative data are nearly identical to normative data collected in the 1970s and 1980s.

*Argument 5: Discrepancies From the Norms Are Due to Improper Test Administration.* Proponents have also argued that the Rorschach was not administered properly in several studies that found discrepancies from the CS norms. For example, Meyer (reported by Mestel, 2003) speculated that the Rorschach was administered improperly in the Hamel et al. (2000) study, although he has never published findings to support this claim. Similarly, Ritzler et al. (2002a, p. 240) and Weiner et al. (2003, p. 8) criticized studies for using graduate students to administer the Rorschach. According to Weiner (2001a):

Rorschach practitioners generally regard good inquiry as an acquired skill that even experienced examiners continue to hone, and most Rorschach students find learning to inquire properly a particularly challenging task. (p. 123)

Such arguments are unconvincing for two reasons. First, quality of test administration was not reported in the studies criticized by the Rorschach proponents. In fact, quality of administration has virtually never been formally assessed in any Rorschach research, and there is no validated, systematic method of quality control for ascertaining whether a particular Rorschach administration is adequate. In the absence of solid data, the objection that test administration was inadequate in these studies is highly speculative and necessarily ad hoc. Because virtually all Rorschach studies are subject to the same objection, it is arbitrary for proponents to single out for criticism only those studies whose findings have indicated problems with the CS norms.

Second, it is puzzling that Rorschach proponents have disparaged studies that used graduate students to administer the test, because a large proportion of important Rorschach research, including studies favorably cited by the proponents themselves, have used administrators without PhDs. For example, when Exner collected his normative data, he used both graduate students and individuals with even less clinical experience to administer the Rorschach. Exner (1986, p. x) described his administrators as follows:

Some have been professional psychologists or psychology graduate students, but more than half come from more varied backgrounds, ranging from a professional musician and a retired tailor, to an extremely talented high school senior. Other examiners have included physicians, dentists, nurses, social workers, educators, homemakers, and a few very adept secretaries who discovered that administering the Rorschach can sometimes be as boring as typing letters.

In a related argument, some Rorschach proponents (Exner, 2001a; Weiner, 2001a) have noted that Lambda scores were high in several studies that reported discrepancies from the CS norms (e.g., Shaffer et al., 1999). According to Rorschach proponents, high scores on Lambda can sometimes indicate that the administrators did not conduct a full inquiry or that the individuals being tested were defensive. Therefore, the proponents have argued, the Rorschach may not have been properly administered in these studies, or the participants may have been unusually defensive.

There is a logical flaw in these arguments. Specifically, the proponents have never explained how an incomplete inquiry or defensiveness can cause a relatively normal individual to appear to have a thought disorder and other symptoms of serious psychopathology on the Rorschach. It seems more likely that incomplete inquiry or defensiveness would have the *opposite* effect, by making individuals appear *less* disturbed than they really are. Thus, when studies in the US and elsewhere repeatedly show that people in



the community have high Lambda scores compared with the CS norms, the most likely explanation is that the norms for Lambda are too low, not that the people are defensive or that the test was administered improperly.

### 1.2. Validity

We turn next to recent developments concerning the validity of Rorschach scores. Advocates and critics of the test agree that at least a small subset of Rorschach scores is valid when used for particular tasks. Specifically, even psychologists who are critical of the test generally agree that some scores from various Rorschach systems can be helpful for detecting thought disorder, diagnosing mental disorders characterized by thought disorder, measuring dependency, and predicting treatment outcome (Lilienfeld et al., 2000; Wood, Nezworski, & Garb, 2003).

There is also general agreement among proponents and critics that many CS scores have not been adequately studied. According to Meyer and Archer (2001):

... many variables given fairly substantial interpretive emphasis have received little or no attention.... These include the Coping Deficit Index, Obsessive Style Index, Hypervigilance Index, active-to-passive movement ratio, D-score, food content, anatomy and X-ray content, Intellectualization Index, and Isolation Index. (p. 496)

In addition, the Perceptual Thinking Index (Exner, 2001b) is a new CS measure that has not been well validated (Meyer & Archer, 2001, p. 496).

The strongest evidence supporting the Rorschach comes from global and focused meta-analyses. *Global meta-analyses* involve pooling the results for a range of CS and non-CS scores. An overall effect size estimate that describes the average level of validity for all of the scores is then calculated. *Specific meta-analyses* involve pooling results for a single score, and calculating an effect size estimate for that score. Some positive results have been obtained from both types of meta-analyses.

The results from global meta-analyses indicate that at least a few Rorschach scores are valid: otherwise overall effect size estimates would not be positive. However, they do not tell us which scores are valid for which tasks. As explained by Meyer and Archer (2001):

Global meta-analyses are inherently limited because they provide diffuse information. They do not cumulatively organize evidence for specific test scales and thus fail to provide fine-grained and clinically useful information about the value of a scale in relation to specific criteria. This is a genuine limitation of global meta-analyses, and it is impossible to circumvent this shortcoming. (p. 491)

In considering the results from global meta-analyses, it is important to note that results have almost always been based on a relatively small number of studies. Statements by Rorschach advocates have sometimes been misleading on this point. For example, in discussing the results from a global meta-analysis conducted by Parker, Hanson, and Hunsley (1988), Weiner and Kuehnle (1998, p. 440) wrote, "Parker et al. used the effect sizes reported in 411 studies to derive population estimates of convergent validity of 0.41 for the Rorschach and 0.46 for the MMPI." This statement is incorrect: the convergent validity coefficient of 0.41 for the Rorschach is based on the results from *only five studies* and the convergent validity coefficient of 0.46 for the MMPI is based on the results from 30 studies (Parker et

al., 1988, p. 370, Table 1). Similarly, [Weiner et al. \(2002\)](#) incorrectly reported the number of Rorschach protocols that were analyzed in a meta-analysis conducted by [Hiller et al. \(1999\)](#).<sup>1</sup>

For focused meta-analyses, positive results have been reported for the detection of thought disorder and the diagnosis of psychotic disorders ([Jorgensen et al., 2000](#)), the prediction of treatment outcome ([Meyer, 2000](#); [Meyer & Handler, 1997](#)), and the assessment of dependent behavior ([Bornstein, 1996](#)). Most psychologists who have been critical of the Rorschach acknowledge that the test can be used for these tasks. Nevertheless, it is striking that focused meta-analyses have been conducted for so few Rorschach scores.

To evaluate the validity of the Rorschach, we ([Lilienfeld et al., 2000](#)) used the following criteria: (1) studies on a score should be methodologically sound, (2) significant results should be replicated by independent investigators, and (3) results should be consistent across studies. These criteria are more rigorous than the implicit criteria employed by many clinicians. Such rigor is necessary, in part, because attempts to replicate significant findings for the CS have frequently failed (e.g., see [Nezworski & Wood, 1995](#); [Wood, Lilienfeld, Garb, & Nezworski, 2000](#); [Wood, Nezworski, Lilienfeld et al., 2003](#), pp. 245–248, 251–252, 265–266). Independent replication is necessary because replications within only one laboratory may be due to systematic sources of error (e.g., undetected confounds or idiosyncrasies in administration or scoring).

Using these three criteria, we concluded that the following Rorschach scores have been validated by research:

- (a) Thought Disorder Index for the Rorschach in the assessment of thought disorder, (b) Rorschach Prognostic Rating Scale in the prediction of treatment outcome, (c) Rorschach Oral Dependency Scale in the assessment of objective behaviors related to dependency, and (d) deviant verbalizations and poor form quality (as well as the CS Schizophrenia Index and other indexes derived from these variables) in the assessment of schizophrenia (and perhaps schizotypal personality disorder and bipolar disorder) and borderline personality disorder. ([Lilienfeld et al., 2000](#), p. 54)

To this list, we have added the [Piotrowski \(1957\)](#) signs of organic brain damage and the [Elizur \(1949\)](#) Hostility and Anxiety scales ([Wood, Nezworski, & Garb, 2003](#); [Wood, Nezworski, Lilienfeld, et al., 2003](#)). Considering that the CS is composed of more than 150 scores and that an even larger number of non-CS scores also exist (most of the scores listed above are non-CS scores), it is evident that the vast majority of Rorschach scores that are widely used do not satisfy our three criteria for being well-validated.

Rorschach advocates have not responded by compiling a list of scores that meet our three criteria. For example, [Hibbard \(2003\)](#) wrote a full-length article criticizing [Lilienfeld et al. \(2000\)](#), but did not comment on the validity of individual Rorschach scores. [Weiner et al. \(2002\)](#) argued for the validity of the Rorschach, but did not use the aforementioned criteria when arguing for the validity of specific scores. For example, they argued that the Rorschach “does very well” when used to measure “dysphoric mood and negative cognitions,” but they did not cite a single study to support this claim ([Weiner et al., 2002](#), p. 10). Similarly, [Weiner et al. \(2002\)](#) cited only one publication to support their claim that the CS

---

<sup>1</sup> According to [Weiner et al. \(2002, p. 8\)](#), the Rorschach results reported by [Hiller et al. \(1999\)](#) were based on 2276 protocols. According to [Hiller et al. \(1999, p. 286; also see Table 8, p. 288\)](#), Rorschach results were based on 1713 protocols. The Rorschach protocols were collected in 30 studies, which is a relatively small number when one considers that hundreds, if not thousands, of studies have been conducted on the Rorschach.

Egocentricity Index is valid. In contrast, Nezworski and Wood (1995, p. 196) reviewed the results from 59 independent studies and concluded that “there is insufficient evidence to support the validity of the EGOI [Egocentricity Index], pairs, and reflections as measures of self-esteem, self-focus, narcissism, ego-functioning, or depression.”

Adopting a different tack, Perry (2003, p. 582) argued that our criteria are too rigorous: “. . .the methods used to evaluate the validity of the Rorschach have been extreme, using stringent demands that have rarely been imposed on other assessment instruments and techniques.” This sentiment may be one reason why Rorschach advocates have not responded by compiling a list of scores that meet these criteria. However, we have pointed out that the Washington University Sentence Completion Test (Loevinger, 1998), a rigorously constructed projective technique, meets the aforementioned criteria (Lilienfeld et al., 2000). We have also described five broad advantages of the Minnesota Multiphasic Personality Inventory-II (MMPI-2; Butcher et al., 1989) over the Rorschach Comprehensive System (Wood, Garb, Lilienfeld, & Nezworski, 2002). Many MMPI-2 scales meet our criteria for empirical support (Greene, 2000).

## 2. Roots of the controversy

Having reviewed the “official” controversy regarding the norms and validity of the CS, we now turn to a more general question: Are there deeper disagreements at stake in the Rorschach controversy that account for the passion that has often intruded into recent debates? By addressing this question, we hope to shed some light on the underlying beliefs that divide Rorschach proponents and critics. We begin by taking a historical approach: we will describe the earliest days of the Rorschach in this country.

### 2.1. Bruno Klopfer, informal validation, and intuitive information integration

From 1940 to 1980, no figure in the United States was more closely associated with the Rorschach than Bruno Klopfer. Besides establishing the most popular American Rorschach system and co-authoring the most widely used textbook on the test (Klopfer & Kelley, 1946), Klopfer founded the Rorschach Institute (which eventually became the Society for Personality Assessment) and the *Rorschach Research Exchange* (which eventually became the *Journal of Personality Assessment*).

Klopfer periodically expressed contempt for the American psychometric approach to testing (see Wood Nezworski, Lilienfeld et al., 2003). As already noted, he rejected any suggestion that the Rorschach should have norms. Likewise, when validity studies failed to confirm many of his ideas about the test, he brushed aside the research with the remark: “Perhaps it is not necessary to be concerned with validity in the usual sense; or perhaps a new technique of validation is necessary” (Klopfer & Davidson, 1962, p. 24).

During the 1930s and 1940s Klopfer promoted his own notions about psychological testing to compete with the psychometric principles that had been developed by American psychologists during the preceding decades (see Wood, Nezworski, Lilienfeld et al., 2003, pp. 76–83). Two of Klopfer’s ideas concern us here because, as we will show, they continue to be influential among Rorschach proponents.

First, Klopfer (1939, p. 47) held that informal observation by individual interpreters was sufficient to demonstrate the validity of the Rorschach (a notion we will refer to as “the principle of informal validation”). Specifically, he recommended that clinicians interpret a Rorschach without knowing

anything else about a patient, and then compare their findings with case notes or observations. If the Rorschach interpretation appeared consistent with these other sources of information, the test could be considered “validated.” Downplaying the need for rigorous scientific validation of the Rorschach, Klopfer and his followers (see [Krugman, 1949](#)) cited enthusiastic testimonials by clinicians who had informally “validated” the test for themselves.

Second, Klopfer ([Klopfer & Kelley, 1946](#), pp. 14–18) held that although individual Rorschach scores do not usually bear a straightforward relationship to personality characteristics, a skilled interpreter can intuitively integrate the scores into a complete picture of a client’s personality (a notion we will refer to as the “principle of intuitive information integration”). According to Klopfer, the intuition of certain elite interpreters was so highly developed that they could extract everything important about a client from the Rorschach alone, without considering information from interviews, biographical data, or other tests.

Klopfer’s principles implied that systematic studies of Rorschach validity were an unnecessary formality and likely to be futile. According to the principle of informal validation, validity studies were unnecessary for the experienced Rorschach interpreter because they merely confirmed what the interpreter already knew through first-hand observation. Furthermore, according to the principle of intuitive information integration, validity studies of individual Rorschach scores were futile and even potentially misleading because they failed to capture the extraordinary richness of the test, which lay in the *weaving together* of multiple scores. As stated by [Krugman \(1949, p. 132\)](#), a Klopfer follower and first president of the Rorschach Institute:

Psychologists have come to recognize that a complex instrument like the Rorschach cannot yield a simple score that may be validated by the application of a Pearson or other correlation, but that the entire configuration must be compared with the clinical picture obtained with such procedures as the full case study or the psychiatric examination. In other words, validation must be by clinical rather than by mathematical processes.

## 2.2. *Modern CS proponents, informal validation, and intuitive information integration*

Several recent examples illustrate the continuing popularity of Klopfer’s principles of informal validation and intuitive information integration. For instance, [Weiner \(2001b\)](#) pointed to informal validation as evidence of the test’s value: “The Rorschach has survived and prospered because generations of psychologists in diverse cultures have found that it helps” (p. 4). Weiner later elaborated on his point of view:

How likely is it that so many Rorschach assessors have been using the instrument for so long, in so many places and contexts, solely on the basis of illusory correlation? If this seems unlikely, is it unreasonable to infer that there has been some utility in their work? ([Weiner et al., 2002](#), p.11)

Similarly, in the *American Psychologist*, [Silver \(2001\)](#) conceded that evidence of the CS validity is weak, but he nevertheless defended his continuing use of the Rorschach: “I am unwilling to discard this instrument. I think of the Rorschach as sampling a domain of behavior, which gives information that I have found to be useful” (p. 1009). As can be seen, both [Weiner et al. \(2002\)](#) and [Silver \(2001\)](#) affirmed informal validation as a justification for using the Rorschach.

As such quotes illustrate, informal validation is alive and well among Rorschach defenders. The notion of intuitive information integration is also thriving. Specifically, Gregory Meyer, now editor of

the *Journal of Personality Assessment*, and his colleagues (Meyer et al., 2001) argued in the *American Psychologist* that the validity of Rorschach and other test findings is likely to increase when clinicians integrate the results with other assessment information:

Because most research studies do not use the same type of data that clinicians do when performing an individualized assessment, the validity coefficients from testing research may underestimate the validity of test findings when they are integrated into a systematic and individualized psychological assessment. (p. 152)

As already noted, Rorschach promoters in the 1940s argued that the intuitive integration of Rorschach scores was such a complex process that it could not be evaluated by validity studies. Similarly, Meyer et al. contended that it is “virtually impossible” for research to evaluate the validity of “contextually embedded inferences” made by sophisticated psychologists:

More generally, to the extent that clinicians view all test data in a contextually differentiated fashion, the practical value of tests used in clinical assessment is likely greater than what is suggested by the research on their nomothetic associations. However, trying to document the validity of individualized, contextually embedded inferences is incredibly complex- and virtually impossible if one hopes to find a relatively large sample of people with the same pattern of test and extratest information (i.e., history, observed behavior, motivational context, etc.). Research cannot hope to approximate such an ideal. (Meyer et al., 2001, p. 153; see also Merlo & Barnett, 2001, for similar arguments).

Meyer et al. neglected to note that a respectable body of research has examined clinicians’ “individualized, contextually embedded inferences” using the Rorschach and other tests (Garb, 1998). As will be described in the next section, these studies have examined the validity of clinical judgments made by mental health professionals.

### 2.3. *Evaluation of informal validation and intuitive integration*

We briefly summarize the evidence bearing on the principles of informal validation and intuitive integration.

*Informal validation: does it work?* The principle of informal validation—that tests or treatments can be satisfactorily “validated” by practitioners’ informal observations rather than by systematic scientific evaluations—has been thoroughly discredited. When “informally” validated treatments and tests have been studied by the scientific method, they have often not been supported. Because a detailed discussion of the subject could fill a thick tome, we only briefly summarize four lines of evidence.

First, the history of astrology, palm reading, phrenology, and other pseudoscientific assessment techniques has demonstrated that they can be highly impressive to individuals who informally “validate” them (Dutton, 1988; Hyman, 1981; Paul, 2004). For instance, some of the greatest scientists of the early modern era (e.g., Copernicus, Kepler) cast horoscopes professionally and were impressed by the results, even though (as we now know) the basis of astrology has been refuted by research (Dean & Mather, 2000). Clearly, informal validation can yield highly misleading results, even when the “validators” are geniuses.

Second and similarly, the history of modern medicine provides numerous examples of treatments and assessment techniques that have become widely popular among intelligent physicians, even when they were later been found to be worthless or even dangerous (Haines, 2002; Lambert, 1978; McCoy, 2000;

Young, 1961, 1967). A tragic case is provided by the history of prefrontal lobotomy, a brain operation for the treatment of schizophrenia and other mental illnesses that became popular among neurosurgeons in the 1940s and 1950s on the basis of testimonials and informal case studies. As one physician who performed lobotomies stated, “I am a sensitive observer, and my conclusion is that a vast majority of my patients get better as opposed to worse after my treatment” (Dawes, 1994, p. 48). Systematic studies eventually showed lobotomies to be ineffective for treatment and seriously harmful in some cases (Valenstein, 1986).

Third, the history of clinical psychology provides similar examples of invalid tests and ineffective treatments that became highly popular among practitioners on the basis of testimonials and informal validation (Lilienfeld, Lynn, & Lohr, 2003). For instance, the single sign approach to projective drawings was vouched for by numerous psychologists in the 1950s and 1960s, even though research has shown overwhelmingly that this approach is invalid (Anastasi, 1982; Lilienfeld et al., 2000). Similarly, an extensive body of research on the phenomenon of illusory correlation indicates that psychologists are sometimes convinced that they have observed a relation between test scores and real-world client characteristics (e.g., personality traits) even when no such pattern exists (e.g., Chapman & Chapman, 1967, 1969; Garb, 1998).

Fourth and directly relevant to Klopfer’s proposal, Rorschach history provides numerous examples of the shortcomings of informal validation. For example, most Rorschachers in the 1950s sincerely but mistakenly believed in the validity of “color shock,” the Mother and Father cards, and other aspects of the test that have been discredited scientifically (Wood, Nezworski, Lilienfeld et al., 2003). Similarly, although thousands of psychologists in the 1990s used the CS Depression Index to screen for depression in millions of patients, the users neglected to note something that systematic studies eventually made clear: Scores on the Depression Index are largely or entirely unrelated to depression (Jorgensen et al., 2000; Wood et al., 2000).

*Intuitive information integration of test scores: does it work?* The principle of intuitive integration of test scores has fared little better than the principle of informal validation. As already noted, Rorschach promoters in the 1940s argued that the intuitive integration of Rorschach scores is such a complex process that it cannot be evaluated by validity studies. Similarly, Meyer et al. (2001) contended that it is “virtually impossible” for research to evaluate the validity of “contextually embedded inferences” made by sophisticated psychologists.

One thing that can be done to overcome the problem they describe is to study clinical judgment. Instead of calculating correlations between a test score and an external criterion, one can ask psychologists to make judgments based on Rorschach protocols alone or Rorschach protocols added to other sources of information. This method would allow psychologists to make judgments in the context of all of the Rorschach scores or all of the Rorschach scores plus other information that they usually have available in clinical practice.

The notion that some highly expert interpreters can extract extraordinary insights from the Rorschach was refuted by researchers as early as the 1950s and 1960s (see Wood, Nezworski, Lilienfeld et al., 2003). For example, in one famous study, Klopfer, Piotrowski and other leading Rorschach experts examined the test scores of cadets in a flight training school and attempted to identify those with “overt personality disturbances” (Holtzman & Sells, 1954). The accuracy of Klopfer and the other experts was no better than what could have been obtained by flipping a coin. Subsequent research has also refuted the notion that purported experts using the Rorschach are more accurate than other judges (e.g., graduate students learning to use the Rorschach; Garb, 1989; Whitehead, 1985).

Although studies have discredited the myth that Rorschach “wizards” can perform miracles with the test, a modified version of the principle of intuitive information integration made its appearance in the 1950s (e.g., Little & Shneidman, 1959). According to this updated notion, although the Rorschach performed poorly when used alone, it could be expected to add valuable information when intuitively integrated with *other* information sources, such as tests or interviews. After all, this is how the Rorschach is ostensibly used in clinical practice. A psychologist uses the test to generate hypotheses, which can be confirmed or eliminated by examining non-Rorschach sources of information.

Despite the appeal of this version of the principle of intuitive information integration, numerous studies since the 1940s have shown that adding the Rorschach to other information sources does *not* typically improve the validity of clinical judgments (Garb, 1998; Wood, Nezworski, Lilienfeld et al., 2003). For tasks of making diagnoses and describing personality and psychopathology, the addition of Rorschach results to other information has *not* led to an increase in the validity of psychologists’ judgments, even when Rorschach protocols have been added to information as simple as demographic data (Garb, 1984, 1998, 2003). For example, in a study that used Exner’s CS (Whitehead, 1985), the task was to discriminate between (a) depressed vs. nondepressed back patients, (b) bipolar vs. schizophrenic psychiatric patients, and (c) back pain patients vs. psychiatric patients. Judges used the MMPI alone, the Rorschach alone, or the MMPI and Rorschach together. Hit rates increased from 58% to 74% when MMPI protocols were added to CS Rorschach protocols, but *decreased* (nonsignificantly) from 76% to 74% when the CS was added to the MMPI.

There is no evidence that results for a personality assessment instrument become more positive when the instrument is evaluated in the context of other assessment information. In a review of the relevant research, one of the authors of the present article (Garb, 2003) found that if a test score possesses little or no validity when used alone, validity will not be boosted by integrating the score with other sources of information. The addition of a test score to other assessment information led to an increase in validity only when the zero-order correlation for the test score by itself was significant.

#### 2.4. Implications for the current controversy over the CS

Recent debates between critics and proponents of the CS have often focused on psychometric issues. However, as the foregoing discussion suggests, other issues in the controversy have deep historic roots and involve more than the reliability, validity, and standardization of the test. By tracing recent statements of CS proponents back to the 1940s and the ideas of Bruno Klopfer, we can understand some of the undercurrents in the present debate. In this section we list four important contemporary questions regarding the Rorschach and demonstrate that they are answered in much different ways by critics who reject Klopfer’s principles and by proponents who accept them.

##### 2.4.1. Should a Rorschach score be scientifically validated before it is used clinically?

Critics of the CS believe that a Rorschach score should be used clinically only if it has been well validated in sound studies. In contrast, CS proponents who place credence in informal validation do not always consider scientific validation to be necessary if a score has already been personally “validated” (see the previous quote from Silver, 2001). Although most CS proponents endorse the general idea that research is important, they often stop short of endorsing the more specific principle that Rorschach scores should be used *only* for purposes for which they have been well validated in scientific studies.

#### 2.4.2. *What is to be concluded when research fails to confirm clinical lore?*

When research findings about the Rorschach repeatedly conflict with clinical lore, critics of the CS give substantially more credence to the research than to the lore. In contrast, most CS proponents who rely on informal validation may feel free to disregard such findings if they conflict with clinical impressions. For example, as already noted, leading CS advocates (e.g., Exner, 2000; Weiner, as quoted in Goode, 2001) continue to promote the Rorschach as a measure of depression and egocentricity despite overwhelmingly negative research findings (Jorgensen et al., 2000; Nezworski & Wood, 1995; Wood et al., 2000).

#### 2.4.3. *Who is qualified to have an opinion about the Rorschach?*

Because they regard systematic studies as the decisive source of information about the Rorschach, critics consider opinions about the test to be legitimate to the degree that they are well-reasoned and based on sound research. In contrast, some proponents who believe in informal validation consider opinions regarding the Rorschach to be illegitimate unless they are made by individuals who regularly use the test in clinical practice and have personally observed its usefulness. For example, Weiner et al. (2002, p. 11) indicated that criticisms of the CS cannot be “welcomed” unless they are made by psychologists “intimately familiar” with the test:

This issue concerns whether persons evaluating the scientific worth of a technique are intimately familiar with the nature of the technique and how it works. In what field of science are criticisms of procedures welcomed from persons who do not themselves use or study these procedures?

#### 2.4.4. *How Informative is the Rorschach?*

Critics of the CS believe that research-based validity coefficients provide the most realistic estimate of how informative the Rorschach is likely to be for a particular purpose. In contrast, CS proponents who subscribe to the notion of intuitive information integration assume that the validity of the Rorschach increases when it is combined with other information by skilled clinicians, and that validity coefficients therefore substantially underestimate the test’s true power. For example, as noted earlier, Meyer contended that “the practical value of tests used in clinical assessment is likely greater than what is suggested by the research on their nomothetic associations” (Meyer et al., 2001, p. 153).

#### 2.4.5. *Summary*

As can be seen, Klopfer’s principles of informal validation and intuitive information integration constitute a set of assumptions, one might even call it a philosophy, with broad implications. CS proponents who accept this philosophy give much different answers to the four preceding questions than do critics who reject it. Most important, to the extent that they accept Klopfer’s ideas, many CS proponents believe that the Rorschach can appropriately be used for clinical purposes for which it has never been scientifically validated, that its clinical value is likely to be substantially greater than formal studies indicate, and that criticism of the test is illegitimate except by psychologists who regularly use it.

Of course, the last notion—that the Rorschach can legitimately be criticized only by psychologists who regularly use it—raises interesting questions about the status of psychologists (such as the authors of the present article) who *formerly* used the Rorschach but stopped doing so,



largely because of the negative scientific evidence. Are criticisms by such “ex-users” legitimate? If not, then a Catch-22 situation seems to arise, in which only psychologists who are content with the Rorschach and use it regularly are allowed to criticize it! The notion that longtime users of the Rorschach possess an exclusive right to offer opinions about the test may also partially account for the heated tone with which proponents have sometimes referred to critics (e.g., [Weiner et al., 2002, 2003](#)).

### *2.5. A reflection of deeper divisions within clinical psychology?*

In closing, we note that the underlying issues in the Rorschach debate identified in this article are also central to other recent controversies within the field of clinical psychology. Specifically, disagreements regarding informal validation that first arose more than 50 years ago continue to play a major part in several current debates.

The central question regarding informal validation is this: Are informal clinical observations as dependable as systematic clinical studies for assessing the validity of tests and the efficacy of treatments? As indicated earlier in this article, the answer to this question is clearly “no.” Overwhelming evidence supports three insights that should be part of the training of every clinical psychologist: First, Bruno Klopfer notwithstanding, informal validation has been shown to be an exceedingly poor way to evaluate the effectiveness of assessment techniques and treatments. Second, systematic studies provide much more dependable evidence regarding test validity and treatment outcome than do personal testimonials, even when the testimonials are made by highly educated and sincere professionals. Third, psychologists who use assessment techniques and treatments that have been only informally validated increase the risk that they will misdiagnose or otherwise harm their clients.

In the 1940s, when Klopfer’s notion of informal validation was first promoted, several prominent psychologists recognized it as fallacious. One such figure was Donald Super of Columbia University, who made a statement that continues to be relevant today:

Unorganized experience, unanalyzed data, and tradition are often misleading. Validation should be accomplished by evidence gathered in controlled investigations and analyzed objectively, not by the opinions of authorities and impressions of observers. (excerpted in [Buros, 1949](#), p. 167)

More than 60 years after Super wrote these words, the dispute continues, as leading clinical psychologists still engage in high-profile debates regarding the merits and demerits of informal validation. For instance, in his recent successful run for president of the American Psychological Association, Ronald [Levant \(2004, p. 223\)](#) endorsed the principle of informal validation, arguing that research, clinical expertise, and patient values should be weighed equally in evaluating treatments:

A model, which values all three components equally, will better advance knowledge related to best treatment and provide better accountability.

In response, psychologists who supported the use of empirically supported treatments stressed the pre-eminent importance of research findings. For example, [Borkovec \(2004\)](#) argued that “. . . the single most important touchstone for everything . . . [we] do in therapy is contained in the following question: ‘What is the empirical evidence for what you just did with the client?’” (p. 212).

Similar disagreements over informal validation have also played an important part in other recent conflicts in clinical psychology, including controversies surrounding the use of anatomically detailed dolls in abuse assessment (Hunsley, Lee, & Wood, 2003; Koocher et al., 1995), clinical versus statistical prediction (Garb, 2000; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954), the status of empirically supported therapies (Chambless & Ollendick, 2001), and the efficacy of “critical incident” stress debriefing (McNally, Bryant, & Ehlers, 2003). In each of these cases, at least some proponents of scientifically questionable techniques have objected to the invocation of research to settle controversies on the grounds that clinical intuition should be permitted to override consistent negative findings (Lilienfeld et al., 2003).

It is troubling that clinical psychologists are still debating how to: (a) validate psychological tests and (b) evaluate the effectiveness of therapeutic interventions (e.g., Borkovec, 2004; Levant, 2004; Nathan, 2004; Peterson, 2004). The shortcomings of informal validation have been clearly recognized within the field of medicine and other disciplines for a very long time. It is our hope that the discipline of clinical psychology will soon discard informal validation and other discredited beliefs from half a century ago. As observed by Beutler (2004): “Contrary to Levant (2004), research, experience, and patient values are not equivalently valid. Scientific research is more likely to produce valid conclusions than sincere clinical opinion based on unsystematic experience. (p. 228).”

## References

- Allen, J., & Dana, R. H. (2004). Methodological issues in cross-cultural and multicultural Rorschach research. *Journal of Personality Assessment*, 82, 189–206.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Collier Macmillan.
- Beutler, L. E. (2004). The empirically supported treatments movement: A scientist-practitioner’s response. *Clinical Psychology: Science and Practice*, 11, 225–229.
- Borkovec, T. D. (2004). Research in training clinics and practice research networks: A route to the integration of science and practice. *Clinical Psychology: Science and Practice*, 11, 211–215.
- Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency Scale: 1967–1995. *Psychological Assessment*, 8, 200–205.
- Burns, B. (1994). An object relations study: Relationship between the Rorschach Human Experiences Variable and interpersonal relatedness among nonpatients (Doctoral dissertation, California School of Professional Psychology, San Diego, 1993). *Dissertation Abstracts International*, 54, 3847B.
- Buros, O. K. (1949). *The third mental measurements yearbook*. New Brunswick, NJ: Rutgers University Press.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193–204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271–280.
- Conners, K. (1989). *Manual for Conners’ rating scales*. North Tonawanda, NY: Multi Health Systems.
- Cronbach, L. J. (1955). Book review of developments in Rorschach technique: Volume I. Technique and theory. *Journal of Educational Psychology*, 46, 121–123.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Dean, G., & Mather, A. (2000, September/October). Sun sign columns. *Skeptical Inquirer*, 24, 36–40.

- DeLucas, A. L. (1997). The quality of paternal attachment and violent behaviors: A Rorschach study of convicted military male offenders (Doctoral dissertation, California School of Professional Psychology, San Diego, 1997). *Dissertation Abstracts International*, 58, 2115B.
- Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest*, 5, 1–31.
- Dutton, D. L. (1988). The cold reading technique. *Experientia*, 44(4), 326–332.
- Elizur, A. (1949). Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Research Exchange*, 13, 247–284.
- Exner, J. E. (1978). The Rorschach: A comprehensive system. *Current Research and Advanced Interpretation*, vol. 2. New York: Wiley.
- Exner, J. E. (1986). The Rorschach: A comprehensive system: Vol. 1. Basic foundations (2nd ed.). New York: Wiley.
- Exner, J. E. (1991). The Rorschach: A comprehensive system: Vol. 2. Interpretation (2nd ed.). New York: Wiley.
- Exner, J. E. (1993). The Rorschach: A comprehensive system: Vol. 1. Basic foundations (3rd ed.). New York: Wiley.
- Exner, J. E. (2000). A primer for Rorschach interpretation. Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2001a). A comment on “The misperception of psychopathology: Problems with the norms of the comprehensive system for the Rorschach”. *Clinical Psychology: Science and Practice*, 8, 386–388.
- Exner, J. E. (2001b). Rorschach workbook for the comprehensive system. (5th ed.) Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2002a). A new nonpatient sample for the Rorschach Comprehensive System: A progress report. *Journal of Personality Assessment*, 78, 391–404.
- Exner, J. E. (2002b). The Rorschach: Basic foundations and principles of interpretation (4th ed.). New York: Wiley.
- Fombonne, E. (1994). Increased rates of depression. Update of epidemiological findings and analytical problems. *Acta Psychiatrica*, 145–156.
- Fombonne, E. (1998). Increased rates of psychosocial disorders in youth. *European Archives of Psychiatry and Clinical Neuroscience*, 248, 14–21.
- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4, 641–655.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, 105, 387–396.
- Garb, H. N. (1998). Studying the clinician: Judgment research and psychological assessment. Washington, DC: American Psychological Association.
- Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot in clinical and forensic settings. *Assessment*, 6, 313–315.
- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there’s anything wrong with that! *Psychological Assessment*, 12, 31–39.
- Garb, H. N. (2003). Incremental validity and the assessment of psychopathology in adults. *Psychological Assessment*, 15, 508–520.
- Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Towards a resolution of the Rorschach controversy. *Psychological Assessment*, 13, 433–448.
- Goode, E. (2001, February 20). What’s in an inkblot? Some say, not much. *New York Times*, D1.
- Goode, E. (2004, March 9). Defying psychiatric wisdom, these skeptics say ‘Prove it.’ *The New York Times*, p. D1, D6
- Greene, R. L. (2000). The MMPI-2: An interpretive manual (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Greenwald, D. F. (1990). An external construct validity study of Rorschach personality variables. *Journal of Personality Assessment*, 55, 768–780.
- Grof, P. (1997). The age of misery: Affective disorders on the rise. *Canadian Journal of Psychiatry*, 42, 361–362.
- Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, and Law*, 5, 224–242.
- Grove, W. M., Barden, R. C., Garb, H. N., & Lilienfeld, S. O. (2002). Failure of Rorschach Comprehensive System based testimony to be admissible under the Daubert-Joiner-Kumho standard. *Psychology, Public Policy, and Law*, 8, 216–234.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.
- Guiffra, L. A., & Risch, N. (1994). Diminished recall and the cohort effect of major depression: A simulation study. *Psychological Medicine*, 24, 375–383.

- Haines, J. D. (2002). The king of quacks: Albert Abrams, MD. *Skeptical Inquirer*, 26(3), 45–48.
- Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment*, 75, 280–294.
- Hertz, M. R. (1959). The use and misuse of the Rorschach method: I. Variations in the Rorschach procedure. *Journal of Projective Techniques*, 23, 33–48.
- Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) "The Scientific Status of Projective Techniques". *Journal of Personality Assessment*, 80, 260–271.
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11, 278–296.
- Holtzman, W. H., & Sells, S. B. (1954). Prediction of flying success by clinical analysis of test protocols. *Journal of Abnormal and Social Psychology*, 49, 485–490.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11, 266–277.
- Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice*, 8, 378–382.
- Hunsley, J., Lee, C. M., & Wood, J. M. (2003). Controversial and questionable assessment techniques. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 39–76). New York: Guilford Press.
- Hyman, R. (1981). Cold reading: How to convince strangers that you know all about them. In K. Frazier (Ed.), *Paranormal borderlands of science* (pp. 79–96). Buffalo, NY: Prometheus.
- Jorgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach depression index and the schizophrenia index: A review. *Assessment*, 7, 259–280.
- Kelleher, K. J., McInerney, T. K., Gardner, W. P., Childs, G. E., & Wasserman, R. C. (2000). Increasing identification of psychosocial problems: 1979–1996. *Pediatrics*, 105, 1313–1321.
- Klopfer, B. (1939). Shall the Rorschach method be standardized? *Rorschach Research Exchange*, 3, 45–54.
- Klopfer, B., & Davidson, H. H. (1962). *The Rorschach technique: An introductory manual*. New York: Harcourt Brace Jovanovich.
- Klopfer, B., & Kelley, D. M. (1946). *The Rorschach technique* (2nd ed.). Yonkers-on-Hudson, NY: World Book.
- Koocher, G. P., Goodman, G. S., White, C. S., Friedrich, W. N., Sivan, A. B., & Reynolds, C. R. (1995). Psychological science and the use of anatomically detailed dolls in child sexual abuse assessments. *Psychological Bulletin*, 118, 119–222.
- Krugman, M. (1949). Review of the Rorschach Inkblot Test. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 132–133). New Brunswick, NJ: Rutgers University Press.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology, Research and Practice*, 34, 491–498.
- Lambert, E. C. (1978). *Modern medical mistakes: Major errors of treatment in the twentieth century*. Bloomington, IN: Indiana University Press.
- Levant, R. F. (2004). The empirically validated treatments movement: A practitioner/educator perspective. *Clinical Psychology: Science and Practice*, 11, 219–224.
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27–66.
- Little, K. B., & Shneidman, E. S. (1959). Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs*, 73, 476.
- Loevinger, J. (1998). *Technical foundations for measuring ego development: The Washington University Sentence Completion Test*. Mahwah, NJ: Erlbaum.
- McCoy, B. (2000). *Quack! Tales of medical fraud from the Museum of Questionable Medical Devices*. Santa Monica, CA: Santa Monica Press.
- McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest*, 4, 45–79.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

- Merlo, L., & Barnett, D. (2001, September). All about inkblots. *Scientific American*, 285, 13.
- Mestel, R. (2003, May 19). Rorschach tested. Blot out the famous method? *Los Angeles Times*, F-1.
- Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment*, 75, 46–81.
- Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science and Practice*, 8, 389–396.
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment*, 13, 486–502.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69, 1–38.
- Meyer, G. J., & Richardson, C. (2001). *An examination of changes in Form Quality codes in the Rorschach Comprehensive System from 1974 to 1995*. Presented at the Midwinter Meeting of the Society for Personality Assessment, Philadelphia.
- Murphy, J. M., Laird, N. M., Monson, R. R., Sobol, A. M., & Leighton, A. H. (2000). A 40-year perspective on the prevalence of depression: The Stirling County Study. *Archives of General Psychiatry*, 57, 209–215.
- Nathan, P. E. (2004). When science takes us only so far. *Clinical Psychology: Science and Practice*, 11, 216–218.
- Nezworski, M. T., & Wood, J. M. (1995). Narcissism in the comprehensive system for the Rorschach. *Clinical Psychology: Science and Practice*, 2, 179–199.
- Parker, G. (1987). Are the lifetime prevalence estimates in the ECA study accurate? *Psychological Medicine*, 24, 375–383.
- Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367–373.
- Paul, A. M. (2004). *The cult of personality: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York: Free Press.
- Perry, W. (2003). Let's call the whole thing off: A response to Dawes (2001). *Psychological Assessment*, 15, 582–585.
- Peterson, D. R. (2004). Science, scientism, and professional responsibility. *Clinical Psychology: Science and Practice*, 11, 196–210.
- Piotrowski, Z. A. (1957). *Perceptoanalysis: A fundamentally reworked, expanded, and systematized Rorschach method*. New York: Macmillan.
- Ritzler, B., Erard, R., & Pettigrew, G. (2002a). A final reply to Grove and Barden: The relevance of the Rorschach Comprehensive System for expert testimony. *Psychology, Public Policy, and Law*, 8, 235–246.
- Ritzler, B., Erard, R., & Pettigrew, G. (2002b). Protecting the integrity of Rorschach expert witnesses: A reply to Grove and Barden (1999) re: The admissibility of testimony under Daubert/Kumho analyses. *Psychology, Public Policy, and Law*, 8, 201–215.
- Schiff, L. M. (1993). *A descriptive Rorschach study of psychoanalytic psychoanalysts (Doctoral dissertation, California School of Professional Psychology, Berkeley/Alameda, 1992)*. *Dissertation Abstracts International*, 53, 4968B.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, 73, 305–316.
- Silver, R. J. (2001). Practicing professional psychology. *American Psychologist*, 56, 1008–1014.
- Swindle Jr., R., Heller, K., Pescosolido, B., & Kikuzawa, S. (2000). Responses to nervous breakdowns in America over a 40-year period. *American Psychologist*, 55, 740–749.
- Tolman, A. O., & Mullendore, K. B. (2003). Risk evaluations for the courts: Is service quality a function of specialization? *Professional Psychology, Research and Practice*, 34, 225–232.
- Twenge, J. M. (2000). The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952–1993. *Journal of Personality and Social Psychology*, 79, 1007–1021.
- Valenstein, E. S. (1986). *Great and desperate cures: The rise and decline of psychosurgery and other radical treatments for mental illness*. New York: Basic Books.
- Wachler, C. A. (1991). Selected Rorschach variables of never-married men. *Journal of Clinical Psychology*, 47, 123–132.
- Weiner, I. B. (2001a). Considerations in collecting Rorschach reference data. *Journal of Personality Assessment*, 77, 122–127.
- Weiner, I. B. (2001b, December). The value of Rorschach assessment. *Harvard Mental Health Letter*, 4–5.
- Weiner, I. B., & Kuehnle, K. (1998). Projective assessment of children and adolescents. In A.S. Bellack and M. Hersen (Series Eds.) and C.R. Reynolds (Vol. Ed.), *Comprehensive clinical psychology: Vol. 4. Assessment* (pp. 431–458). Oxford, UK: Elsevier.

- Weiner, I. B., Spielberger, C. D., & Abeles, N. (2002). Scientific psychology and the Rorschach Inkblot Method. *The Clinical Psychologist*, 55, 7–12.
- Weiner, I. B., Spielberger, C. D., & Abeles, N. (2003). Once more around the park: Correcting misinformation about Rorschach assessment. *The Clinical Psychologist*, 56, 8–9.
- Whitehead, W. C. (1985). *Clinical decision making on the basis of Rorschach, MMPI, and automated MMPI report data*. Unpublished doctoral dissertation, University of Texas Health Science Center at Dallas.
- Widiger, T. A. (2001). The best and the worst of us? *Clinical Psychology: Science and Practice*, 8, 374–377.
- Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology*, 53, 519–543.
- Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach inkblot test: A case of overstatement? *Assessment*, 6, 341–349.
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach Test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology*, 56, 395–430.
- Wood, J. M., Nezworski, M. T., & Garb, H. N. (2003). What's right with the Rorschach? *Scientific Review of Mental Health Practice*, 2, 142–146.
- 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, 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, 397–402.
- Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). What's wrong with the Rorschach? Science confronts the controversial inkblot test. San Francisco, CA: Jossey-Bass.
- Young, J. H. (1961). *The toadstool millionaires: A social history of patent medicines in America before federal regulation*. Princeton, NJ: Princeton University Press.
- Young, J. H. (1967). *The medical messiahs: A social history of health quackery in twentieth-century America*. Princeton, NJ: Princeton University Press (This book can be read without charge on the World Wide Web at <http://www.quackwatch.com/>).
- Zacker, J. (1997). Rorschach responses of police applicants. *Psychological Reports*, 80, 523–528.