

Problems With the Norms of the Comprehensive System for the Rorschach: Methodological and Conceptual Considerations

James M. Wood, University of Texas at El Paso

M. Teresa Nezworski, University of Texas at Dallas

Howard N. Garb, Pittsburgh V. A. Health Care System and University of Pittsburgh

Scott O. Lilienfeld, Emory University

In "The Misperception of Psychopathology: Problems with the Norms of the Comprehensive System for the Rorschach," we reviewed results from 32 studies and concluded that the norms of the Comprehensive System (CS) are inaccurate and tend to make nonpatient Americans appear psychologically disordered. Hunsley and Di Giulio, Widiger, and Aronow all agree that our review uncovered serious problems with the CS norms. However, CS proponents Exner and Meyer disagree. In this reply we address criticisms and express the hope that CS proponents will eventually acknowledge the importance of the research findings and develop more accurate norms.

Key words: Rorschach, norms, projective techniques, psychometrics. [*Clin Psychol Sci Prac* 8:397-402, 2001]

In "The Misperception of Psychopathology: Problems with the Norms of the Comprehensive System for the Rorschach" (Wood, Nezworski, Garb, & Lilienfeld, this issue), we reviewed findings from 32 studies and concluded that the norms of the Comprehensive System (CS) for the Rorschach are seriously in error and tend to make nonpatient Americans appear psychologically disordered. In the present issue of *Clinical Psychology: Science and Practice*, several noted clinical psychologists provide stimulating comments on our review, both pro and con. We are grateful to the editor, David Barlow, for offering us the opportunity to respond.

Address correspondence to James M. Wood, Ph.D., Department of Psychology, University of Texas at El Paso, El Paso, TX 79968. Electronic mail may be sent to jawood@utep.edu.

RESPONSE TO HUNSLEY AND DI GIULIO

John Hunsley and Gina Di Giulio (this issue) agree that our analyses "provide compelling evidence that the CS adult norms are seriously flawed and are likely to result in frequent false-positive identifications of psychopathology." They conclude that "in terms of both size of the normative samples and quality of sampling used to obtain the norms, the quality of the CS norms are below that of the norms for the WAIS-III and MMPI-2." This conclusion is important for the field of psychological assessment because several researchers have wondered whether other commonly used tests could withstand the same level of scrutiny that has been directed at the Rorschach (Archer, 1999; Widiger, this issue).

To improve use of the Rorschach, Hunsley and Di Giulio propose that "the most prudent and responsible step would be to immediately begin the process of re-norming the CS." They do not recommend that psychologists use the numbers in our review as norms because these figures are based on convenience samples rather than on true probability samples and may not be adequately representative of the U.S. population. We agree. Nor do Hunsley and Di Giulio recommend that psychologists continue to use the current CS norms: "to continue to use norms that are highly likely to be inaccurate and overpathologizing . . . is ethically untenable." Again, we agree.

RESPONSE TO WIDIGER

Thomas Widiger is one of the foremost psychologists in the areas of psychopathology and clinical assessment. We are pleased that he is willing to share his thoughts regarding our article. If the controversy over the CS is to be resolved, a broad range of psychologists, including those not normally associated with projective techniques, need to evaluate the relevant evidence.

Widiger believes that we have provided "a compelling critique of the norms developed by Exner for the Rorschach Comprehensive System." He concludes that "There does appear to be a systematic error that results in clinicians and researchers identifying pathology that may not in fact be present."

However, Widiger believes that the Rorschach still retains some value. He argues that the CS has allowed the Rorschach to be examined scientifically:

Prior to the Comprehensive System, the interpretation and scoring of the Rorschach was so inconsistent across clinicians and researchers that it was difficult to subject to meaningful, replicable research. The Comprehensive System has undermined substantially the dismissal of negative findings on the basis of idiosyncratic or inadequate interpretation or scoring.

We agree in part with Widiger. On the one hand, the CS has almost certainly brought greater uniformity to the administration and scoring of the Rorschach. On the other hand, the reliability of CS scoring has been substantially overstated by some CS proponents (see reviews by Lilienfeld, Wood, & Garb, 2000; Wood & Lilienfeld, 1999). For this reason, some scholars friendly to the Rorschach have recently proposed that procedures be established to evaluate the scoring competency of psychologists who use the CS.¹ Furthermore, Meyer (this issue) suggests that problems with CS administration may be sufficiently common to undermine the findings of some Rorschach studies. If Meyer is correct, the question arises whether use of the CS has in fact yielded a meaningful, replicable body of research, as Widiger believes. Perhaps it is still true that many negative research findings can be dismissed post hoc by Rorschach proponents, on the (usually untestable) grounds that administration of the test was poor.

Widiger is concerned about the use of the Rorschach in forensic settings: "The Rorschach is prone to excessive interpretation and should perhaps not have an appreciable or significant impact on the resolution of legal disputes." However, he believes that the Rorschach can sometimes be useful in psychotherapy. We consider Widiger's suggestion to be reasonable, although not without potential problems. We will comment on the use of the Rorschach as a therapeutic technique in our response to Aronow (this issue).

Widiger playfully suggests that our article might provide a substitute set of norms: "Wood et al. have perhaps made a major contribution to the Rorschach literature by so conscientiously accumulating data for norms that appear to be more valid than those originally developed by Exner." Of course, we do not endorse this tongue-in-cheek proposal. The numbers in our article are unsuitable for norms, for the reasons explained by Hunsley and Di Giulio.

Finally, we wish to clarify two misunderstandings. According to Widiger, "Another possible response is to

abandon the use of the Rorschach altogether. This is the outcome preferred by Wood et al." Widiger is mistaken that we recommended the abandonment of the Rorschach. It is true that one of us has called for a moratorium on the use of the Rorschach in clinical and forensic settings (Garb, 1999). However, this was not a call for abandonment; by definition, a moratorium is temporary. In this case, the moratorium was proposed to last until research can determine which CS scores are valid (and for which purposes) and which are not. The moratorium would not extend to research settings.²

Widiger also writes, "Wood et al. are indicating that the interests, perspective, and efforts of Exner and his colleagues are no longer wanted." Certainly we have never made a statement to this effect, nor have we intended such a conclusion. The efforts of Exner and his colleagues are welcome. However, we believe that many psychologists need to adopt a more critical attitude toward the CS than they have in the past.

RESPONSE TO ARONOW

Edward Aronow (this issue) considers the findings of our review to be credible:

The Wood et al. thesis as to the problematic nature of the CS norms provided by Exner actually seems quite reasonable, especially given the fact that the Wood article was stimulated by the publication of a carefully designed normative study carried out by Shaffer, Erdberg, and Haroian (1999). . . . In that this study was carried out by experienced researchers who can hardly be said to be opponents of the CS, Wood et al.'s general point on the CS norms appears to be justified.

Because Aronow is a champion of the Rorschach and considers it the "king of projectives," his acceptance of our general conclusions is particularly notable.

We appreciate Aronow's realistic approach to the Rorschach. The psychometric achievements of the CS have sometimes been seriously overstated (Wood & Lilienfeld, 1999). In contrast, Aronow freely concedes that the Rorschach has serious psychometric problems, but argues that it possesses utility as a "high band-width, low-fidelity instrument" that can be helpful in psychotherapy. In this respect his position is similar to Widiger's. Aronow views the Rorschach not as a highly valid assessment tool, but as a source of (possibly erroneous) hypotheses that can be fruitfully explored with clients in therapy. We consider

Aronow's position to be reasonable and consistent with important psychotherapeutic traditions in clinical psychology.³

RESPONSE TO EXNER

We turn now to the comment of John Exner, originator of the CS. Although appreciative of Exner's remarks, we are disappointed that he did not more directly address the practical implications of our findings. In their comments, Hunsley and Di Giulio, Widiger, and Aronow all agree that these findings are important and reflect serious problems with the CS norms. In contrast, Exner does not indicate that a reassessment of the norms is desirable, except possibly for $X+%$ and $X-%$.

Furthermore, Exner's comment neglects to mention that another problem with the CS norms has been identified. Exner (2001, p. 172; personal communication, March 23, 2001) has recently reported that 221 of the 700 protocols in the 1993 adult normative sample were duplicates. That is, the sample of 700 protocols was actually composed of 479 protocols, with 221 of these protocols counted a second time. The mistake was identified more than 2 years ago (Gregory Meyer, message to internet discussion group, Society for a Scientific Clinical Psychology at sscpnet@listserv.acns.nwu.edu, March 4, 2001). The question thus arises whether this error, or perhaps other errors not yet identified, may account for the discrepancies between the CS norms and the studies in our review. Unfortunately, the data needed to fully address this question are not available, although we strongly encourage further investigation of this issue.

Psychologists troubled by such news now confront important questions. In light of the accumulating evidence, can practitioners safely assume that the norms for CS variables are accurate? What are the ethical and legal implications for psychologists who dismiss the substantial scientific findings and continue to use the norms in clinical and forensic settings? Exner's silence on these topics provides little guidance to psychologists who use the CS in practice.

In the final part of his comment, Exner argues that we have misunderstood how the CS is used in actual practice. For instance, he criticizes our use of means and standard deviations to draw conclusions about overpathologizing: "It is, at best, naive to assume that interpreters of the Rorschach use a table of nonpatient descriptive statistics and systematically mark off seeming discrepancies between

means, modes, or standard deviations when reviewing a protocol." Furthermore, Exner suggests that our findings are irrelevant to the identification of psychopathology: "It is difficult to ascertain how the notion that differences among nonpatient sets, such as they have presented, leads to a misperception of psychopathology."

Finally, Exner insists that decisions about psychopathology are based not on individual Rorschach scores, but on a global assessment of all Rorschach findings: "The issue of whether or not someone is pathological is derived globally, in the context of all findings and, even then, is offered as a hypothesis to be evaluated against other information concerning the individual."

In response, we would like to make three points. First, we are surprised at Exner's suggestion that Rorschach interpreters do not "systematically" rely on the means, medians, and standard deviations in the CS normative tables when reviewing a protocol. If interpreters do not systematically note "discrepancies" between a client's scores and the CS norms, how can they accurately determine whether those scores are truly deviant?

Second, Exner's comments to the contrary, it is clear that inaccurate norms can lead to misperceptions of psychopathology. For example, if the norms for the MMPI-2 Schizophrenia scale were wrong by a full standard deviation (10 *T*-score points), then many normal patients would fall above the clinical cut-off of 65 and be considerably more likely to be perceived as seriously disturbed. The same reasoning applies to the Rorschach. In our review, using examples from Shaffer, Erdberg, and Haroian (1999), we showed how the CS norms can lead to substantial overpathologizing of nonpatients (e.g., classifying 16% of nonpatient adults as possibly schizophrenic and 29% as probably narcissistic). It is regrettable that Exner rejects this "notion" and dismisses the means and standard deviations in our review as irrelevant to the issue of overpathologizing.

Third, we strongly disagree with Exner's contention that a "global" approach to the CS eliminates the danger of overpathologizing. The problems of a global approach are illustrated in a dissertation by Mittman (1983). Mittman collected CS protocols from 6 nonpatient adults. The protocols were then rated by 90 alumni of Exner's Rorschach Workshops, who were asked to assign them to 12 diagnostic categories, including "normal." The results were disturbing: In 99 judgments, the alumni classified 12% of the nonpatient protocols as "major affective disor-

der,” 23% as “reactive depression,” and 43% as “personality disorder.” Only 21% of the nonpatient protocols were classified as “normal” or “malingering” (see summary by Exner, 1991, pp. 432–433).

The extremely high rate of false-positive diagnoses in Mittman’s (1983) study suggests that the global approach to CS interpretation overpathologizes approximately 75% of normal individuals, and that the false-positive diagnoses most likely to be assigned are depression, other types of mood disorder, and personality disorder. Psychologists who use the CS may justifiably wonder whether, like the psychologists in Mittman’s study, they too have been misidentifying psychopathology where none is present.

RESPONSE TO MEYER

Whereas Exner mainly disregards our findings, CS advocate Gregory Meyer challenges them at length and in meticulous detail. In the limited space available for a reply, we can address only the most serious of Meyer’s criticisms.

When the Data Are Reanalyzed, Our Main Findings Remain Unchanged

Meyer points out that our review overlooked a relevant study by Perry, Potterat, Auslander, Kaplan, and Jeste (1996), as well as some means and standard deviations in other articles. We appreciate Meyer’s feedback and have conducted reanalyses that incorporate these overlooked data. As it turns out, these reanalyses yield results that are virtually identical to those reported in Tables 2 and 3 of our review. We will be happy to provide corrected versions of these tables on request.

Meyer’s International Sample Replicates Our Own Findings

Meyer summarizes the results from 9 international CS studies that included 2,125 nonclinical participants. Averaging the results for 69 Rorschach variables, he reports that the international sample was about four-tenths of a standard deviation more impaired than Exner’s normative sample (mean $d = -.38$). When form-quality variables were set aside, the international sample was about three-tenths of a standard deviation more impaired than the normative sample (mean $d = -.31$). Meyer concludes, “it should be clear that the CS norms do not overpathologize.”

Of course, important information can be lost when many numbers are averaged, as Meyer has done. The question arises, “What would Meyer have found had he

looked at only the 14 variables in our review?” These 14 variables are central to CS interpretation. We included them in our review because (a) they exhibited substantial discrepancies from the CS norms in the study by Shaffer et al. (1999) and (b) multiple studies in our review had included them because of their theoretical and practical importance to Rorschach interpretation. To examine the performance of these same 14 variables in Meyer’s international sample, we requested his data.

When we reviewed Meyer’s numbers, we found that the findings for the international sample were remarkably similar to our own. We encourage interested readers to obtain the numbers from Meyer (which he offers to provide in his comment) and examine them directly. For example, averaging the results for the variables in our review,⁴ the international sample was about eight-tenths of a standard deviation more impaired than Exner’s normative sample (mean $d = -.77$), while the aggregated nonpatient sample in our review was about seven-tenths of a standard deviation more impaired than the normative sample (mean $d = -.71$). As can be seen, the international sample actually looked slightly more pathological than our own. When form-quality variables were set aside, the mean values for the international sample and our own nonpatients were identical to the second decimal place (mean $d = -.62$). These effect sizes were calculated in an extremely conservative manner and would have been substantially larger had we used the standard deviations of the CS norms.

The findings from Meyer’s international sample most certainly do not refute our conclusions. To the contrary, they replicate our findings and bolster our contention that the CS norms tend to make nonpatients look pathological.

How Healthy Was Exner’s Normative Sample?

Our review found that nonpatient American adults tend to appear disturbed when compared with the CS norms. To explain this finding, Meyer claims that the CS normative sample was in fact more psychologically healthy than the nonpatients in our review: “Exner’s (1993) nonpatient reference sample consists of people with no history of mental health treatment and some positive evidence of healthy functioning.”

However, Meyer overstates the psychological health of the CS normative sample. First, contrary to Meyer’s claim that the CS nonpatient sample had “no history of men-

tal health treatment,” Exner (personal communication, February 6, 2001) has informed us that between 80% and 85% of the CS normative sample had no psychiatric/psychological history whatsoever. The remainder had consulted a professional at some time in their lives (no more than eight visits) for academic difficulties, occupational decisions, pastoral counseling, family difficulties, grief counseling, or other problems (John Exner, personal communication, March 29, 2001). Second, the only positive evidence of healthy functioning mentioned by Meyer is that most members of the CS normative sample were recruited at their work or through organizations such as the PTA or bowling leagues (Exner, 1993, p. 258). However, the fact that an individual is employed or belongs to a bowling league does not constitute very compelling evidence of healthy functioning. Even seriously disturbed individuals may hold a job or belong to a club. Third, Meyer neglects to mention that 42 low-income individuals were deliberately recruited for the CS normative sample “through the assistance of social service agencies” (Exner, 1993, p. 258). As Exner (1991) has stressed, the members of the CS normative sample “are not necessarily normal; they are simply not patients. . . . ranging from introverted to extroverted, well controlled to poorly controlled, gregarious to isolated, strange to sturdy, and so on” (p. 461).

Were Our Nonpatient Samples More Pathological Than Exner's Normative Sample?

Meyer argues that the nonpatient samples in our review were not really comparable with the CS normative sample because they included “a broader range of the population than Exner's reference group and thus should obtain somewhat less healthy CS scores if the CS is a valid personality measure.” However, Meyer's attempt to portray our nonpatient samples as less healthy than the CS normative sample is unconvincing for several reasons.

First, Meyer criticizes our samples because “at least 5 explicitly included current or former psychiatric patients.” But according to Exner (personal communication, February 6, 2001), the CS normative sample also included individuals who had received brief treatment, though not for severe or prolonged disorders.

Second, Meyer argues that “at least eight other samples used screening criteria but nonetheless still would have included people with active or past disorders and/or treatment histories.” But according to Exner (personal com-

munication, February 6, 2001), approximately 15–20% of the CS normative sample also included individuals with brief treatment histories.

Third, Meyer argues that one of our samples included “poor people who participated because they needed the money.” But according to Exner (1993, p. 258), 42 members of the CS normative sample “were recruited through the assistance of social service agencies.”

In closing, we wish to express our hope that Meyer, Exner, and other CS advocates will eventually acknowledge the importance of the research evidence: The norms for many important CS variables are clearly unrepresentative of nonpatient Americans and tend to make normal individuals appear pathological. Although CS proponents may now disregard (e.g., Exner) or attempt to refute (e.g., Meyer) this unwelcome finding, perhaps in the future they will change their minds and revise the CS norms, as Hunsley and Di Giulio suggest. In the meantime, psychologists should not use the present CS norms in clinical or forensic work. Psychologists who use these norms run the risk of attaching false and negative labels to clients and causing them harm.

NOTES

1. In a message on November 19, 2000 to the internet Rorschach Discussion List at rorschach@maelstrom.stjohns.edu, Steven D. Hickman, Psy.D., announced that he had organized a symposium for the 2001 Midwinter Meeting of the Society for Personality Assessment to discuss a “gold standard” for CS scoring and an examination to assure competency among Rorschach scorers.

2. The moratorium was not intended to extend to the use of the Rorschach as a therapeutic technique. Unfortunately, this was not made clear in Garb (1999).

3. Nevertheless, Aronow's potentially promising approach may not be entirely without its problems. For example, through a process of confirmatory bias or anchoring (Garb, 1998; Tversky & Kahneman, 1974), clinicians who draw erroneous inferences from a Rorschach in therapy may tend to become biased when they interpret other information concerning a client.

4. The variables included in the average were Reflections (percent of protocols with 1 or more reflections), *EB* (percent of ambivalent protocols), and means for Reflections, $X+%$, $X-%$, *Afr*, *FC*, *Populars*, *Sum Y*, *Sum T*, *WSumC*, *Morbid responses*, *WSum6*, *Lambda*, and *Pure H*. We included Reflections twice (the mean and percent of protocols) because Meyer had done so in his analyses of the international sample. This double inclusion of Reflections had little effect on the overall mean.

ACKNOWLEDGMENT

We thank John Exner and Gregory Meyer for kindly providing us with information and data in response to our inquiries. We also thank Kim McKinzey for pointing out the relevance of Mittman's (1983) findings to the issue of overpathologizing by psychologists who use the CS.

REFERENCES

- Archer, R. P. (1999). Introduction to a special section: Perspectives on the Rorschach. *Assessment*, 6, 307–311.
- Exner, J. E. (1991). *The Rorschach: A Comprehensive System. Volume 2: Interpretation* (2nd ed.). New York: Wiley.
- Exner, J. E. (1993). *The Rorschach: A Comprehensive System. Volume 1: Basic Foundations* (3rd ed.). New York: Wiley.
- Exner, J. E. (2001). *A Rorschach workbook for the Comprehensive System* (5th ed.). Asheville, NC: Rorschach Workshops.
- 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.
- 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.
- Mittman, B. L. (1983). Judges' ability to diagnose schizophrenia on the Rorschach: The effect of malingering (doctoral dissertation, Long Island University, The Brooklyn Center, 1983). *Dissertation Abstracts International*, 44, 1248B.
- Perry, W., Potterat, E., Auslander, L., Kaplan, E., & Jeste, D. (1996). A neuropsychological approach to the Rorschach in patients with dementia of the Alzheimer type. *Assessment*, 3, 351–363.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current non-patient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, 73, 305–316.
- Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and probability. *Science*, 185, 1124–1131.
- Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment*, 6, 341–351.

Received March 12, 2001; accepted March 20, 2001.