Evidence to Correct Misperceptions About Rorschach Norms

Gregory J. Meyer

University of Alaska Anchorage

Abstract

Possible factors affecting normative shifts in Rorschach data are considered, including (a) genuine changes in mental health over time, (b) alterations in the type of target sample under consideration, (c) evolving scoring rules, and (d) variations in test administration skills or context. It is shown that Comprehensive System (CS) criteria for coding form quality have changed substantially over time. Building on the extensive research of others, it is also shown that CS data collected around the world from people tested outside of a clinical context look somewhat less healthy than Exner=s (1993) reference sample of socially/vocationally functioning nonpatients but somewhat more healthy than Exner=s reference sample of people starting outpatient psychotherapy. Furthermore, preliminary results from Exner=s new nonpatient sample recruited using the same procedures as before reveals scores that are generally quite similar to the existing reference values. The assertion that CS norms overpathologize people is not supported.

Keywords: Rorschach, Comprehensive System, Norms, Normative Changes, Test Scores, Types

Evidence to Correct Misperceptions About Rorschach Norms

Determining why normative values for a test may change is not easy. For instance, intelligence test scores have been changing at a dramatic rate, though why remains unclear (Flynn, 1999; Raven, 2000). Each time an IQ test is renormed, the same level of cognitive capacity is associated with a lower score. The ability that would have produced an IQ of 100 two generations ago would now only produce an IQ between 82 and 64, depending on the test. Thus, the average person from two generations ago would now look 1.2 to 2.4 standard deviations (SDs) more impaired (i.e., Cohen=s d of -1.2 to -2.4). Stated more dramatically, the day an IQ test is renormed, it begins Ato label thousands of children as mentally retarded who had escaped the label the previous day@ (Flynn, 1999, p. 12). End users of the test often are not aware of this shift because IQ is not reported in raw scores. Instead, each time a test is renormed, the average IQ value is recalibrated and, by definition, arbitrarily reset to equal 100.

Similarly, many users of the MMPI-2 or MMPI-A are not aware of how symptom reports changed between the initial and contemporary norms because raw scores are transformed so a T-score of 50 defines what is average, regardless of the actual symptoms reported. Nonetheless, people now endorse more symptoms. For instance, on Scale 4 (Psychopathic Deviate) and Scale 6 (Paranoia), adults now score about .50 to .75 of a SD higher, while adolescents now score about 1.0 to 1.5 SDs higher. These changes correspond to d values between .5 and 1.5, which are somewhat larger than the mean differences reported by Wood, Nezworski, Garb, and Lilienfeld (2001) for their selected subset of 14 Rorschach scores.

To understand possible normative shifts for the Rorschach Comprehensive System (CS),

at least four factors must be considered. First, it is possible that scores have changed because people have genuinely changed. Although Wood et al. (2001) did not address this issue, research suggests psychopathology has increased over time (e.g., Fombonne, 1994, 1998; Kelleher, McInerny, Gardner, Childs, & Wasserman, 2000; Swindle, Heller, Pescosolido, & Kikuzawa, 2000; Twenge, 2000). Accordingly, a valid measure of pathology should track those changes to show somewhat increased rates of mental health problems.

A second critical factor concerns the type of sample under consideration. Nonpatients defined by some positive evidence of health are not the same as nonpatients defined by no prior treatment history, and neither is equivalent to a nationally representative or census matched sample. According to the recent Surgeon General=s report, within a 12-month period about 28% of the U.S. population will have a diagnosable mental or addictive disorder, as determined by a limited subset of diagnoses (U.S. Department of Health, 1999). Lifetime rates of disorder are substantially higher. One of the studies informing the Surgeon General=s report (Kessler et al., 1994) found that 48% of the population has a lifetime history of at least one disorder (out of 14 examined). This study also found that about 28% of the population seeks treatment at some point in their life and about 8% seek treatment even though they do not have a diagnosable disorder.

Kessler et al.=s (1994) figures allow one to estimate the rate of lifetime disorder in different types of samples. The best evidence suggests that about 78% of the people seeking treatment would have a diagnosable disorder. About 48% of a nationally representative sample would have a lifetime disorder. In a nonpatient sample defined by the absence of a treatment history, about 39% would have a disorder. In nonpatients defined by no treatment history and positive evidence of health, such as functioning in the educational, vocational, and interpersonal

spheres (U.S. Department of Health, 1999), the rate of disorder is unknown, though it certainly would be lower than any of the prior categories because psychiatric disorders are associated with unemployment, withdrawal from social or community participation, and lower income, education, and work functioning (Kessler & Frank, 1997; Kessler et al., 1994; U.S. Department of Health, 1999). Thus, nonpatient samples that combine even a single positive criterion, such as gainful employment, with an absent treatment history will have the lowest rate of psychiatric disturbance.

Exner=s (1993) nonpatient reference sample consists of people with no history of mental health treatment and some positive evidence of healthy functioning. Specifically, 69% of the people were recruited through their job and an additional 25% were recruited through social or interest organizations (e.g., PTA, Audubon society). Thus, 94% of the sample demonstrated some positive evidence of functioning and health, in addition to having no prior treatment. Although Wood et al. (2001) dismissed the notion that Exner=s sample was healthier than average, the fact is those recruitment procedures should obtain healthier volunteers than (a) using the absence of a treatment history as the sole selection criterion or (b) obtaining a representative sample of the population that includes people regardless of their treatment history, psychiatric history, employment status, or social functioning.

Personality tests that use a representative sample of the population as the reference standard (e.g., MMPI-2, PAI), incorporate a substantial degree of pathology into their definition of what is normal or average. To the extent that MMPI-2 or PAI norms parallel a nationally representative sample, T-scores on the tests implicitly assume that it is Anormal@ for 48% of the population to have a lifetime history of psychiatric disorders. There is nothing wrong with using

this as a standard for health. However, it is quite different than the traditional CS standard.

Although Wood et al. (2001) referred to their 32 samples as Anonpatients,@ at least 5 explicitly included current or former psychiatric patients (Jansak, 1996; Jacques, 1990; Schiff, 1992; Van Horn, 1996; Waehler, 1991; see Waehler, 1995). Other samples included people expected to have psychiatric difficulties (e.g., Hallet, 1996, Wald, Archer, & Winstead, 1990), subjects known to be low functioning (Burns, 1993), poor people who participated because they needed the money (Jacques, 1990), or people tested under atypical circumstances (e.g., Goldfinger [1998] had subjects wear psychophysiological electrodes, did not let them touch the cards, and instructed them to remain as motionless as possible during the testing). Furthermore, the largest study (Meyer, 1989) accepted anyone who would volunteer to be tested for a full day, including people with active or prior psychiatric conditions and people who had used or were using psychiatric medications, outpatient therapy, or inpatient treatment. Eleven additional samples also did not employ any psychiatric screening criteria (Alexander, 1997; Calkins, 1980; Erstad, 1995 [2 samples]; Greenwald, 1990; Hayslip, McBride, Lowaman, & Aronson, 1992; Kranau, 1983; Perry & Kinder, 1992; Smith, Hillard, Walsh, Kubacki, & Morgan, 1991; Zacker, 1997; Zlotogorski, Hahnemann, & Wiggs, 1987). At least eight other samples used some screening criterion but nonetheless still would have included people with active or past disorders and/or treatment histories (Caine, Frueh, & Kinder, 1995; DeLucas, 1997; Frueh & Kinder, 1994; Hilsenroth, 1996; Kadle, 1989; Lipkin, 1988; Meisner, 1988; Netter & Viglione, 1994). Overall, at least 90% of Wood et al.=s samples fall into these categories. It should be obvious these samples target 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.

Relatedly, 16 of Wood et al.=s (2001) samples were college students (Alexander, 1997; Caine et al., 1995; Calkins, 1980; Frueh & Kinder, 1994; Greenwald, 1990; Hilsenroth, 1996; Meisner, 1988; Meyer, 1989; Perry & Kinder, 1992; Smith et al., 1991; Zlotogorski et al., 1987) or the elderly (Erstad, 1995; Hayslip et al., 1992; Kadle, 1989; Lipkin, 1988; Paul, 1987). Based on existing research, all the elderly samples were expected to produce atypical Rorschachs, much like the elderly show raw score decrements on other types of tasks (e.g., Baltes, 1997; Wechsler, 1997). Research also indicates college students have areas of symptomatology that exceed that seen in a representative sample or even a psychiatric population. The best evidence comes from Morey=s (1991) Personality Assessment Inventory (PAI). College students (N = 1,051) are more disturbed than a census matched sample (N = 1,000) and a representative clinical sample (N = 1,246) on the egocentric and stimulus seeking subscales of antisocial features, on mania and its subscales of grandiosity and activity level, and on verbal aggression. College students and the clinical sample are more disturbed than the census matched sample on psychotic experiences and manic irritability. Finally, college students fall between the clinical and census samples on antisocial behaviors and the borderline features of identity problems, self-harm, and affective instability. Given these data, asserting that college students or the elderly mimic Exner=s nonpatient sample would appear to serve no useful scientific goal.

An interesting feature of the PAI reference samples is the relative lack of difference between the clinical and census matched samples. About 35% of the PAI=s clinical sample were outpatients, about 25% inpatients, and the remainder came from alcohol abuse programs (13.6%), correctional facilities (10.2%), or other settings. While all 53 PAI scales show elevations in the clinical sample, the magnitude of difference between the clinical and census

sample is often relatively small. For instance, 34 of the scales (64%) have a Cohen=s d of .75 or less. No scales have a d > 1.11. Thus, a large clinical sample, 25% of whom were inpatients, produces PAI results that are often rather similar to the levels of symptomatology observed in a census matched sample. Similarly modest differences have been observed for a census matched MMPI-2 sample relative to an inpatient population (M |d| = .60 over 42 scales) and for the MMPI-A normative sample relative to a largely inpatient (81%) sample (M |d| = .48 over 35 scales; see Schinka, Elkins, & Archer, 1998; Schinka, LaLone, & Greene, 1998). These findings are not terribly surprising given that about half of a representative sample should have a history of at least one diagnosable disorder (Kessler et al., 1994). However, they also suggest CS scores should begin to look more like a patient sample as one moves from a target like Exner=s socially-vocationally functioning nonpatients to a target that is more representative of the full population.

A third important consideration is the prospect that subtle scoring changes may have emerged as the CS evolved. To the extent that scoring rules have been clarified over time or new standards implemented, more recent samples may appear different when that is not the case.

To address CS changes, one needs to examine the published scoring criteria in detail. One basic questions concerns the extent to which Exner=s form quality tables have changed. As part of a broader project, I have been tabulating this information. Although still unfinished, the initial results are illuminating. The 1974 and 1995 form quality tables for Cards I to VIII list about 1,750 identical percepts. Of the good form entries in 1974, only about half continue to be listed as good form (i.e., FQo) in 1995. The rest are now split between the two poor form categories (FQu . 45%; FQ- . 6%). Thus, there has been a substantial and apparently unnoticed shift in the

form quality tables over time. Because of this, relative to protocols that were collected and scored in the early years of the CS (and Exner=s nonpatients were collected prior to 1985), contemporary samples should have noticeably lower scores on X+% and F+%, noticeably higher scores on Xu% and somewhat higher scores on X-%, S-%, M-, and the SCZI.

With this in mind, I examined the early literature on form quality. I searched the Journal of Personality Assessment from 1974-1985 and found 48 samples from 14 articles that reported X+% scored by the CS. Sixty percent of the samples were expected to be psychologically impaired, while the remaining 40% were nonclinical or control samples (none of which were included in Wood et al.). Across all 48 samples, including outpatients, inpatients, and patients with schizophrenia, bipolar disorder, and borderline personality disorder, the weighted mean X+% was .70 (N = 1,517). Thus, the average X+% for clinical and control samples scored about 20 years ago is 1.26 SDs higher than the value of .51 that Shaffer, Erdberg, and Haroian (1999) found in a contemporary nonpatient sample. In fact, the clinical samples from 20 years ago had an X+% that was about .90 SDs higher than the Shaffer et al. nonpatient sample (i.e., .64 vs .51).

Although Wood et al. (2001) attributed the large difference they observed in form quality scores to an Aerror@ in Exner=s nonpatient norms, this was an incorrect attribution. Exner=s norms are quite consistent with the early CS literature. Had Wood et al. expanded the scope of their review, they would have seen that contemporary nonpatient samples produce worse form quality scores than Exner=s nonpatients and the patient samples described in the early CS literature.

Clearly, the shift in the CS form quality tables has generated an unappreciated problem. It must be factored in when considering any contemporary data. Furthermore, in order to make

meaningful comparisons with the existing reference samples, the protocols in the reference samples would have to be rescored according to current form quality standards.

With respect to other scoring changes, examples in Exner=s 1974 text can be contrasted with the same examples in Exner=s 1993 text (e.g., Tables 12, 13, and 14 in 1993 are equivalent to Tables 7, 8, and 10 in 1974). While there are no changes for some determinants (Fr+rF, FD, M, FM, m), many color or shading examples now contain additional elaborations that help justify the assigned code. For chromatic color, achromatic color, texture, vista, and diffuse shading, about 15% to 50% of the examples have been modified at least slightly. Some of these changes relate to revised scoring criteria for C= and Y (for which reference protocols were rescored; compare Exner, 1974 vs 1986). However, the overall pattern suggests that many determinants used to be assigned more liberally. Today an examinee must convey more specific information before the codes can be assigned, which may have implications for scores on Lambda or PureF%. Other changes in the scoring examples include the more frequent assignment of secondary contents, idiographic contents that are now placed in classified content categories, decreased use of FQ+ relative to FQo, modified rules for scoring DQ, and a slight modification in assigned z scores. In addition, there are now clearer rules that make it less likely to observe white space z-scores for Cards III and X. These latter changes suggest that more contemporary samples may observe slightly different values for Zf, Zd, the Isolation Index, and the Intellectualization Index.

A fourth salient issue concerns the skill of the examiners who collect Rorschach records. Like any complex test, the Rorschach can only be administered properly after adequate training. Training is particularly critical at the inquiry stage because it is the examiner=s skill that clarifies

information needed for scoring. In turn, the inquiry has an impact on many variables, including W, Dd, color, diffuse shading, achromatic color, texture, and vista, as well as the scores derived from these variables, including Lambda or PureF%, EA, es, D score, Adjusted D score, and EB styles. To date, more attention has focused on the adequacy of CS scoring than on procedures to ensure adequate administration. However, if contemporary samples are collected by relatively inexperienced examiners, regardless of scoring accuracy, the protocols could appear deviant.

It is well known that engagement with the Rorschach can be indexed by the number of responses (R) and Lambda (Meyer, 1999), and that task engagement moderates the elevation observed on many other CS scores (e.g., Meyer, 1992, 1993). Although task engagement can be a function of the person being tested or the testing context, it can also be a function of the examiner. Thus, it is worthwhile to examine the extent to which R and Lambda correlate with the 14 variables presented in Wood et al.=s review. This is particularly important because Wood et al. used some studies with odd data on these scores. In one study the original authors deliberately deleted about half of each participant=s responses (Perry & Kinder, 1992), reducing the average R to just 15. In another study, 44% of the sample had prior Rorschach training and produced an average R of 39 (Schiff, 1992). Obviously, these are dramatically different types of samples.

For this analysis, I used Wood et al.=s samples (after correcting omissions; see below), and, to increase the number of observations, used subsamples within a study when possible. Despite the latter, the analyses were quite underpowered, with the number of observations ranging from just 6 to 25. Nonetheless, R was significantly correlated with FC, WSumC, SumY, SumT, Pure H, WSum6, and Afr in the range between .65 and .90. Lambda was significantly correlated with FC, WSumC, and Afr in the range between -.57 and -.66. These data underscore

the importance of considering the structural relations among Rorschach variables when interpreting Wood et al.=s findings. They also highlight the importance of adequate task engagement on the part of examinees and adequate administration skills on the part of examiners.

With respect to Wood et al.=s (2001) analysis, several other problems can be noted. At times, they overlooked scores that were provided (e.g., using just 1 of 4 scores from Alexander, 1997) or used the wrong data (e.g., using (H) rather than H in Hallett, 1992). In other instances, they did not use the SDs that were given (e.g., Kranau, 1983) or that could be computed from the available information (e.g., Zlotogorski et al., 1987). In some instances Wood et al. combined different samples in the same source (e.g., Alexander, 1997; Burns, 1993; Kranau, 1983), while in other instances they did not (e.g., Erstad, 1995). For some studies they estimated scores from available data (e.g., WSumC in Goldfinger, 1998; Hallett, 1996; Van Horn, 1996; and Zlotogorski et al., 1987), while in other studies they did not (e.g., WSumC in Kranau, 1983). Importantly, Wood et al. did not estimate X+%, X-%, and Lambda from Ritzler and Nalesnik (1990), even though these subjects had scores that were almost identical to Exner=s nonpatients. Wood et al. also overlooked at least one relevant study that obtained form quality, Lambda, and Popular scores quite consistent with Exner=s (Perry, Potterat, Auslander, Kaplan, & Jeste, 1996). Further, Wood et al. used biased data from Zaker (1997) because the author only reported scores that deviated significantly from Exner=s. Scores that did not differ were never included in Wood et al.=s review. Wood et al. also aggregated information across studies for scores in which coding rules have been revised over time. For instance, WSum6 was revised in 1991 and Popular in 1986. Yet Wood et al. mixed data that had been collected before and after these revisions.

Finally, the scope of Wood et al.=s (2001) review should be considered. It is misleading to think it was based on 32 samples because many studies provided data for just one or two scores. Only two studies provided data for all 14 variables under consideration. Second, Wood et al. seem to lose sight of the fact that they deliberately selected a limited subset of variables. From these they then generalized to argue that the CS as a whole overpathologized patients.

To contend with many of the problems in Wood et al.=s (2001) review, it would be optimal to examine the full range of scores considered important to CS interpretation using relatively large samples that report all the relevant scores. Fortunately, Rorschach researchers from around the world have been collecting samples of people who were not tested in clinical settings. These samples provide a stringent test of the generalizability of the CS. They vary considerably in the procedures used to select participants, with some attempting to recruit nonpatients, some recruiting census matched samples, and some essentially recruiting any person who was willing to be tested. Furthermore, the samples were tested across very different cultures using a multitude of languages and examiners with many differences in skill and training.

The data considered here are the adult samples presented in Erdberg and Shaffer=s (1999) international symposium, including Shaffer et al.=s (1999) from the U.S. (N = 123), Noriko Nakamura=s from Japan (N = 240), Vera Campo=s from Spain (N = 520), Jan Ivanouw=s from Denmark (N = 72), Carl-Erik Mattlar=s from Finland (cf Mattlar et al., 1993; N = 343), Christian Mormont=s from Belgium (N = 300), Antonio Pires=s from Portugal (2000; N = 309), Matilda Raéz=s from Peru (N = 164), and Isidro Sanz=s from Argentina (N = 54). One other sample was excluded because it reported some erroneous results.

The weighted mean scores from these international samples were compared to two

reference samples: Exner=s (1993) 700 nonpatients and his 440 outpatients beginning therapy. The latter provides a very appropriate anchor for comparison. To the extent that the diverse samples from around the world are representative of a full population and full range of functioning, their scores should start to approximate Exner=s outpatient values more than his socially/vocationally functioning nonpatients. To test this postulate, I examined the 69 scores included in the lower portion of a CS structural summary, as these scores form the foundation for clinical interpretations. Cohen=s d was used to quantify deviations from Exner=s reference samples. The sign of each d was determined empirically by the nonpatient and outpatient means, with positive values indicating greater health and negative values indicating less health.

It should be recognized that the analyses considered here are much more comprehensive than those reported by Wood et al. (2001). Not only do the protocols come from many countries around the world, but the sample is about 3 times larger than Wood et al.=s average sample (2,125 vs 686.8), and this review examines about 4.5 times as many CS scores (69 vs 15).

Overall, across all 69 scores considered, the international sample was about 4 tenths of a SD more impaired than Exner=s nonpatients (i.e., M d = -.38) and about equal to Exner=s outpatients (M d = .03). However, from prior analyses, we know that form quality rules have changed over time. Indeed, the largest differences between the international sample and both of Exner=s reference samples were for the form quality variables. Setting these scores aside, the composite sample was about 3 tenths of a SD more Apathological@ than Exner=s nonpatients (i.e., M d = -.31) and about 1 tenth of a SD healthier than Exner=s outpatients (M d = .08).

Translated into more familiar assessment terms, if Exner=s nonpatient sample means were set at a T-score = 50, like on the MMPI-2, the pooled international sample would have a

mean T-score of 53. Alternatively, using an IQ metric, if Exner=s nonpatient means were set at 100, the international sample would have an average score of 95. This degree of increased Apathology@ is actually less than one would see if an intelligence test was renormed after 20 years (Flynn, 1999). Considering how the international sample also reflects a different type of normative target than Exner=s nonpatients, it should be clear that the CS norms do not overpathologize.

The central unresolved question is what a contemporary nonpatient sample would look like if it was collected using Exner=s initial procedures (i.e., trained examiners recruiting people with no treatment history though their jobs or social groups). This project has been underway for the past year. Although Exner will publish full results when data collection is complete, I compared the first 100 people in this new group to the initial nonpatients and outpatients. On average, the new sample is about 2 tenths of a SD less healthy than the existing nonpatients but about 3 tenths of a SD more healthy than the outpatients. Thus, the new sample would have an average T-score of 52 or an average IQ score of 97. These differences are trivial.

Overall, the data support four primary conclusions. First, there have been formerly unappreciated changes in CS form quality codes. It is now easier to obtain unusual and minus scores. Second, through the extensive effort of many individuals, CS reference data has been collected around the world, including the Northern and Southern hemispheres, Asia, Europe, and the Americas, using a diverse set of examiners with variable levels of training who administered the test in very different cultures and languages to people with a wide range of mental health characteristics. As would be expected, the international data appear to reflect the degree of pathology that would be fairly characteristic of a general population. Their scores look a bit

worse than Exner=s nonpatient reference sample, but also somewhat healthier than his outpatients starting therapy. Third, using the same recruitment procedures as before, preliminary results for Exner=s new nonpatient sample look quite similar to the existing reference values. Fourth, the data contradict Wood et al.=s (2001) conclusions.

Author Notes

Correspondence concerning this article can be sent to the author at the Department of Psychology, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508; e-mail: afgjm@uaa.alaska.edu.

References

Alexander, S. (1998). A comparison of black Americans from single- and two-parent families on the Rorschach (Doctoral dissertation, California School of Professional Psychology, Berkeley/Alameda, 1997). Dissertation Abstracts International, 58, 5184B.

Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. American Psychologist, 52, 366-380.

Burns, B., (1994). An object relations study: Relationship between the Rorschach Human Experience Variable and interpersonal relatedness among nonpatients (Doctoral dissertation, California School of Professional Psychology, San Diego, 1993). Dissertation Abstracts International, 54, 3847B.

Caine, S. L., Frueh, B. C., Kinder, B. N. (1995). Rorschach susceptibility to malingered depressive disorders in adult females. In C. D. Spielberger and J. N. Butcher (Eds.), Advances in Personality Assessment, Volume 10, (pp. 165-174). Hillsdale, NJ: Erlbaum.

Calkins, D. L. (1981). The Rorschach Egocentricity Index: A study of its relationship to self esteem (Doctoral dissertation, Brigham Young University, 1980; Dissertation Abstracts International, 42, 365B.

DeLucas, A. L. (1997). The quality of parental attachment and violent behaviors: A Rorschach study of convicted military offenders (Doctoral dissertation, California School of Professional Psychology, San Diego, 1997; Dissertation Abstracts International, 58, 2115B.

Erdberg and Shaffer (1999, July). International symposium on Rorschach nonpatient data: Findings from around the World I, II, III. Symposium presented at the XVIth Congress of the International Rorschach Society, Amsterdam, The Netherlands.

Erstad, D. (1996). An investigation of older adults= less frequent human movement and color responses on the Rorschach (Doctoral dissertation, Marquette University, 1995). Dissertation Abstracts International, 57, 4084B.

Exner, J. E., Jr. (1974). The Rorschach: A Comprehensive System. Volume 1. New York: Wiley.

Exner, J. E., Jr. (1986). The Rorschach: A Comprehensive System. Volume 1: Basic Foundations (2nd Ed.). New York: Wiley.

Exner, J. E., Jr. (1993). The Rorschach: A Comprehensive System. Volume 1: Basic Foundations (3rd Ed.). New York: Wiley.

Exner, J. E., Jr. (1995). A Rorschach workbook for the Comprehensive System. Asheville, NC: Rorschach Workshops.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. American Psychologist, 54, 5-20.

Fombonne, E. (1994). Increased rates of depression: Update of epidemiological findings and analytical problems. Acta Psychiatrica Scandinavica, 90, 145-156.

Fombonne, E. (1998). Increased rates of psychosocial disorders in youth. European Achieves of Psychiatry and Clinical Neuroscience, 248, 14-21.

Frueh, B. C., & Kinder, B. N. (1994). The susceptibility of the Rorschach Inkblot Test to malingering of combat-related PTSD. Journal of Personality Assessment, 62, 280-298.

Goldfinger, D. (1999). Rorschach patterns in Vietnam veterans with post-traumatic stress disorder: A study of cognition, psychophysiology, and psychological defense (Doctoral dissertation, University of Michigan, 1998). Dissertation Abstracts International, 59, 3691B.

Greenwald, D. F. (1990). An external construct validity study of the Rorschach personality variables. Journal of Personality Assessment, 55, 768-780.

Hallet, S. J. (1996). Trauma and coping in homicide and child sexual abuse detectives (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). Dissertation Abstracts International, 57, 2152B.

Hayslip, B., McBride, P. A., Lowman, R. L., & Aronson, H. J. (1992). Color vision deficits and Rorschach performance in aged persons. International Journal of Aging and Human Development, 34, 165-173.

Hilsenroth, M. J. (1997). A multidimensional assessment of narcissistic personality disorder (Doctoral dissertation, University of Tennessee, Knoxville, 1996). Dissertation Abstracts International, 57, 4709B.

Jacques, M. F. (1991). The Rorschach Egocentricity Index: A validation study (Doctoral dissertation, California School of Professional Psychology, San Diego, 1990), Dissertation

Abstracts International, 52, 1064B.

Jansak, D. M. (1997). The Rorschach Comprehensive System Depression Index, depression heterogeneity, and the role of self-schema (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). Dissertation Abstracts International, 57, 6576B.

Kadle, J. C. (1989). The suitability of standard Rorschach scoring categories for nonpatient and clinically depressed elderly (Doctoral dissertation, Case Western Reserve University, 1989). Dissertation Abstracts International, 50, 2154B.

Kelleher, K. J., McInerny, T. K., Gardner, W. P., Childs, G. E., & Wasserman, R. C. (2000). Increasing identification of psychosocial problems: 1979-1996. Pediatrics, 105, 1313-1321.

Kessler, R. C., & Frank, R. G. (1997). The impact of psychiatric disorders on work loss days. Psychological Medicine, 27, 861-873.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Archives of General Psychiatry, 51, 8-19.

Kranau, E. J. (1984). Level of acculturation/biculturation and Rorschach protocols of Hispanic Americans and Anglo Americans (Doctoral dissertation, Oklahoma State University, 1983). Dissertation Abstracts International, 44, 3200B.

Lipkin, L. R. (1989). Rorschach performance as a test of a theory of late-onset depression in women (Doctoral dissertation, Temple University, 1988). Dissertation Abstracts International, 49, 2864B.

Mattlar, C. E., Forsander, C., Norrlund, L., Öist, A.-S, Mäki, J., & Alanen, E. (1993).

Rasgos del contenido y del resumen estructural Rorschach de los Finlandeses: Una comparacion transcultural. Revista Sociedad Espanola Rorschach Met Y Proy, 21-27.

Meisner, S. (1988). Susceptibility of Rorschach distress correlates to malingering. Journal of Personality Assessment, 52, 564-571.

Meyer, G. J. (1991). An empirical search for fundamental personality and mood dimensions within the Rorschach test (Doctoral dissertation, Loyola University of Chicago, 1989). Dissertation Abstracts International, 52, 1071B-1072B.

Meyer, G. J. (1992). The Rorschach's factor structure: A contemporary investigation and historical review. Journal of Personality Assessment, 59, 117-136.

Meyer, G. J. (1993). The impact of response frequency on Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. Journal of Personality Assessment, 60, 153-180.

Meyer, G. J. (1999). The convergent validity of MMPI and Rorschach scales: An extension using profile scores to define response/character styles on both methods and a re-examination of simple Rorschach response frequency. Journal of Personality Assessment, 72, 1-35.

Morey, L. C. (1991). The Personality Assessment Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources.

Netter, B. E. C., & Viglione, D. J. (1994). An empirical study of malingering schizophrenia on the Rorschach. Journal of Personality Assessment, 62, 45-57.

Paul, M. E. (1989). Geriatric normative data for the Rorschach for a community-living sample (Doctoral dissertation, Ball State University, 1987). Dissertation Abstracts International,

49, 3453B.

Perry, G. G., & Kinder, B. N. (1992). Susceptibility of the Rorschach to malingering: A schizophrenia analogue. In C. D. Spielberger and J. N. Butcher (Eds.), Advances in Personality Assessment, Volume 9 (pp. 127-140). Hillsdale, NJ: Erlbaum.

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.

Pires, A. A. (2000). National norms for the Rorschach normative study in Portugal. In Dana, R. H. (Ed.), Handbook of cross-cultural and multicultural personality assessment (pp. 367-392)

Raven, J. (2000). The Raven=s Progressive Matrices: Change and stability over culture and time. Cognitive Psychology, 41, 1-48.

Ritzler, B., & Nalesnik, D. (1990). The effect of inquiry on the Exner Comprehensive System. Journal of Personality Assessment, 55, 647-656.

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.

Schinka, J. A., Elkins, D. E., & Archer, R. P. (1998). Effects of psychopathology and demographic characteristics on MMPI-A scale scores. Journal of Personality Assessment, 71, 295-305.

Schinka, J. A., LaLone, L., & Greene, R. L. (1998). Effects of psychopathology and demographic characteristics on MMPI-2 scale scores. Journal of Personality Assessment, 70,

197-211.

Shaffer, T. W., Erdberg, P., and Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. Journal of Personality Assessment, 73, 305-316.

Smith, J. E., Hillard, M. C., Walsh, R. A., Kubacki, S. R., & Morgan, C. D. (1991). Rorschach assessment of purging and nonpurging bulimics. Journal of Personality Assessment, 56, 277-288.

Swindle, R., Jr., Heller, K., Pescosolido, B., & Kikuzawa, S. (2000). Responses to nervous breakdowns in America over a 40-year period. American Psychologist, 55, 740-749.

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.

U.S. Department of Health and Human Services. (1999) Mental health: A report of the Surgeon General. Rockville, MD: Author.

Van Horn, P. J. (1996). A Rorschach study of posttraumatic stress disorder in Vietnam theater veterans (Doctoral dissertation, Pacific Graduate School of Psychology, 1996). Dissertation Abstracts International, 57, 2890B.

Waehler, C. A. (1991). Selected Rorschach variables of never-married men. Journal of Clinical Psychology, 47, 123-132.

Waehler, C. A. (1995). Relationship patterns of never-married men and their implications for psychotherapy. Psychotherapy, 32, 248-257.

Wald, B. K., Archer, R. P., & Winstead, B. A. (1990). Rorschach characteristics of mothers of incest victims. Journal of Personality Assessment, 55, 417-425.

Wechsler, D. (1997). WAIS-III administration and scoring manual. San Antonio, TX: The

Psychological Corporation.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001). The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. Clinical Psychology: Science and Practice.

Zaker, J. (1997). Rorschach responses of police applicants. Psychological Reports, 80, 523-528.

Zlotogorski, Z., Hahnemann, L. E., & Wiggs, E. A. (1987). Personality characteristics of hypnotizability. American Journal of Clinical Hypnosis, 30, 51-56.

Footnotes