Psychological Bulletin
2015, Vol. 141, No. 1, 236–249

0033-2909/15/$12.00 DOI: 10.1037/a0036005

COMMENT

A Second Look at the Validity of Widely Used Rorschach Indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013)

James M. Wood
University of Texas at El Paso

Howard N. Garb
Joint Base San Antonio, Lackland, Texas

M. Teresa Nezworski
William Beaumont Medical Center, Fort Bliss, El Paso, Texas

Scott O. Lilienfeld
Emory University

Misty C. Duke
University of Texas at El Paso

We comment on the meta-analysis by Mihura, Meyer, Dumitrascu, and Bombel (2013), which examined the validity of scores in Exner's Comprehensive System (CS) for the Rorschach. First, we agree there is compelling evidence that 4 categories of cognitive scores—the "Rorschach cognitive quartet"—are related to cognitive ability/impairment and thought disorder. We now feel comfortable endorsing the use of these scores in some applied and research settings. Second, we conducted new meta-analyses (k = 44) for the 4 noncognitive Rorschach scores with highest validity in the Mihura et al. findings. Unlike Mihura et al., we included unpublished dissertations (although we did not attempt to exhaustively unearth all unpublished studies), calculated correlations instead of semipartial correlations, and used the Rorschach International Norms for a larger proportion of comparisons. Our validity estimates for the *Suicide Constellation* and *Weighted Sum of Color* were similar to or even higher than those of Mihura et al., although we concluded that support for the *Suicide Constellation* is limited and that *Weighted Sum of Color* probably does not measure its intended target. Our validity estimates for Sum Shading and the *Anatomy and X-ray* score were much lower than those of Mihura et al. We conclude that their meta-analysis accurately reflects the published literature, but their exclusion of unpublished studies led to substantial overestimates of validity for some and perhaps many Rorschach scores. Therefore, the evidence is presently insufficient to justify using the CS to measure noncognitive characteristics such as emotionality, negative affect, and bodily preoccupations.

Keywords: Rorschach Inkblot Test, Comprehensive System, validity, meta-analysis

Supplemental materials: http://dx.doi.org/10.1037/a0036005.supp

The Rorschach Comprehensive System (CS; Exner, 1974, 2003) has been the object of scholarly debate for many years (e.g., Exner, 1996; Wood, Nezworski, & Stejskal, 1996). On the one side are authors who have contended that the CS Rorschach boasts a large number of well-validated and clinically useful indices (Exner & Erdberg, 2005; Viglione & Meyer, 2008). On the other side are authors who have maintained that the psychometric status of the CS is problematic and that only a few Rorschach scores are characterized by well-replicated empirical support (Wood, Nezworski, Lilienfeld, & Garb, 2003). To evaluate the validity of the CS, Mihura, Meyer, Dumitrascu, and Bombel (2013) have presented an ambitious and wide-ranging meta-analysis in *Psychological Bulletin.* By synthesizing validity coefficients for more than 50 CS Rorschach scores based on more than 200 published studies, they have taken an important step toward resolving the controversy that has surrounded the test.

The results of this meta-analysis provide support for both sides of the CS debate. On the one hand, Mihura et al. (2013) found "little to no validity support for over a third . . . of the targeted variables" (p. 574). Their meta-analysis revealed that several well-known Rorschach scores either performed poorly or lacked evi-

James M. Wood, Department of Psychology, University of Texas at El Paso; Howard N. Garb, Psychological Research Service, Joint Base San Antonio, Lackland, Texas; M. Teresa Nezworski, Department of Behavioral Health, William Beaumont Medical Center, Fort Bliss, El Paso, Texas; Scott O. Lilienfeld, Department of Psychology, Emory University; Misty C. Duke, Department of Psychology, University of Texas at El Paso.

The opinions expressed in this document are solely those of the authors and do not represent an endorsement by or the views of the United States Air Force, the Department of Defense, or the United States government. We wish to thank Lawrence D. Cohn of the University of Texas at El Paso for his guidance on aspects of the meta-analysis.

Correspondence concerning this article should be addressed to James M. Wood, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968. E-mail: jawood@utep.edu

dence of validity, including the Egocentricity Index (said to measure excessive self-focus), Space responses (said to measure oppositionality) and Adjusted D (said to measure stress and coping). On the other hand, Mihura et al. reported many positive Rorschach findings. Their meta-analysis identified numerous "strongly supported" Rorschach scores with validity coefficients above .30, particularly scores related to cognitive ability and cognitive impairment, leading Mihura et al. to recommend their use in clinical settings.

This Comment presents our reflections on the methodology and conclusions of the Mihura et al. (2013) meta-analysis. After briefly discussing Rorschach scores assumed to measure cognitive qualities, we discuss noncognitive Rorschach scores and report results from our reanalyses of the relevant research. Online supplemental materials show the calculations for the meta-analyses in our article.

Validity of Scores in the Rorschach Cognitive Quartet

Both Rorschach critics and Rorschach proponents have long agreed that some Rorschach scores are correlated with intelligence, cognitive impairment, and thought disorder (Dawes, 1994; Exner, 1974, 2003; Wood, Nezworski, & Garb, 2003). Many of these scores are listed in Table 1. As can be seen, they fall into four categories: (a) productivity, which reflects the number of distinct images that the patient has reported seeing in the inkblots; (b) complexity/synthesis, which reflects the patient's tendency to combine diverse blot elements into well-organized and integrated images; (c) form quality, which reflects the degree to which the images described by the patient show accurate correspondence or "good fit" to the shape of the blots; and (d) deviant verbalization, which indicates aberrant speech patterns or odd thoughts in the patient's responses. We refer to these categories and the scores associated with them as the Rorschach "Cognitive Quartet." Also included in the quartet are a few hybrids, most notably the *Perceptual-Thinking Index* (*PTI*), that combine two or more variables from the quartet into a single score.

On the basis of their meta-analytic results, Mihura et al. (2013, p. 570) concluded that validity was either "good" or "excellent" for all 13 of the cognitive scores in Table 1. As the authors noted (p. 575), the validity coefficients for these cognitive Rorschach scores were among the largest obtained (r = .35 to .50).

These positive findings for *Cognitive Quartet* scores are consistent with those of prior literature reviews (e.g., Wood, Lilienfeld, Garb, & Nezworski, 2000). We are in full agreement with Mihura et al. (2013) that these scores are correlated with various aspects of cognitive ability and cognitive impairment. Nevertheless, it should be borne in mind that Rorschach scores within the same category of the Cognitive Quartet tend to be substantially intercorrelated and largely redundant. For example, the intercorrelations of CS measures of complexity/synthesis such as *Synthesized Response* (*DQ+*), *Lambda*, the *Complexity Ratio* (*Blends/R*), and *Organizational Frequency* (*Zf*) range from approximately .40 to .70 (Wood, Krishnamurthy, & Archer, 2003, p. 264). Similarly, these scores predict essentially the same criteria, including IQ scores, dementia, and head injury, as can be verified by consulting the Appendix of the Mihura et al. meta-analysis. Just as the reliable variance in many personality measures can be accounted for largely by five underlying constructs known as the Big Five (McCrae & Costa, 1997), so the reliable variance in the best validated Rorschach scores can be accounted for largely by the Cognitive Quartet.

Noncognitive Rorschach Variables

We now turn to those Rorschach scores that are posited to measure noncognitive aspects of personality and psychopathology. Eight of these scores achieved substantial external validity ($r >$ .30) in the Mihura et al. (2013) meta-analysis. They are listed in Table 2, along with their interpretations and test–retest reliabilities (Sultan, Andronikof, Réveillère, & Lemmel, 2006; Sultan & Meyer, 2009). Seven of these scores were classified as "strongly supported" by Mihura et al. (p. 575). The eighth, *Weighted Sum of Color*, showed substantial validity (r = .38) but was classified as having only "good support" because it had been examined in relatively few studies.

Our main focus will be on the four scores with the highest validity coefficients: the *Suicide Constellation*, *Weighted Sum of Color*, *Sum of Shading*, and the *Anatomy and X-ray* score. These variables present a special challenge to critics of the Rorschach, who have been skeptical of claims that the Rorschach is useful for measuring noncognitive aspects of personality and psychopathology (Wood, Nezworski, & Garb, 2003). We conducted new liter-

Table 1
Categories of the Rorschach Cognitive Quartet and the Variables Belonging to Them

Category	Variable name (acronym)	Interpretation
1. *Productivity*	*Number of Responses* (*R*)	Ability or tendency to respond with many ideas
2. *Complexity/Synthesis*	*Lambda* (*L*)	Avoidance vs. attentiveness to complexity, subtlety, or nuance
	Synthesized Response (*DQ+*)	Ability to synthesize concepts
	Organizational Frequency (*Zf*)	Ability to sustain cognitive effort
	Complexity Ratio (*Blends/R*)	Psychological complexity
	Human Movement (*M*)	Mental abilities including planning
	Experience Actual (*EA*)	Cognitive and emotional resources and imagination
3. *Form Quality*	*Conventional Form* (*X+%*)	Tendency to perceive the world as others do
	Distorted Form (*X-% and XA%*)	Distorted perceptions
	Appropriate Form (*WDA%*)	Reasonably appropriate perceptions
4. *Deviant Verbalizations*	*Critical Special Scores*	Thought disturbance
	Critical Special Scores, Severe	Thought disturbance, severe
Hybrid Score	*Perceptual-Thinking Index*	Disturbed thinking and distorted perceptions

Table 2
Noncognitive Rorschach Variables With Validity > .30 in the Mihura et al. (2013) Meta-Analysis

Variable name	Interpretation	Validity		Test–retest reliability
		External criterion	Introspective criterion	
Suicide Constellation (SCON)	Suicide risk	.41	.30	.47
Weighted Sum of Color (WSumC)	Emotions influence thoughts and experiences	.38	.30	.69
Sum of Shading	Distressing or irritating internal stimuli	.37	.09	.42
Anatomy and X-ray (An + Xy)	Preoccupations with body vulnerability or its functioning	.33	.01	.53
Inanimate Movement (m)	Mental distraction or irritation, often as a reaction to a moderate or severe stressor	.33	.14	.47
Form-Color Ratio (FC:CF + C)	Emotional impulsivity or reactivity	.32	.07	.52
Difference Score (D score)	Current level of coping abilities	.31	.11	.34
Cooperative Movement (COP)	Tendency to perceive positive interpersonal interactions	.31	−.04	.38

Note. Variable names, interpretations, and validity coefficients are from Mihura et al. (2013). Three-month test–retest reliability coefficients are from Sultan et al. (2006) and Sultan and Meyer (2009).

ature searches and meta-analyses for these four noncognitive Rorschach scores.

Procedures for New Meta-Analyses

Our procedures for conducting new meta-analyses were the same as those described by Mihura et al. (2013), with three exceptions. First, Mihura et al. included only published studies in their meta-analysis. In contrast, we included both published studies and relevant dissertations. Only dissertations that were available for download from the ProQuest Dissertations & Theses Database through our college libraries without paying a special fee were included. In general, this meant that all relevant dissertations from 1997 onward were included, plus a few earlier dissertations that could be downloaded without special payment. Our inclusion of dissertations allowed us to tap into a substantially larger body of research than Mihura et al. accessed. It also allowed us to evaluate the potential impact of publication bias on their findings.

A second difference between our procedures is that we used a different method for computing validity coefficients. Mihura et al. (2013, pp. 560–561) computed many of the validity coefficients in their meta-analysis by calculating semipartial correlations of Rorschach scores with relevant criteria while controlling for the number of responses (*R*) that participants gave to the blots. In contrast we computed validity coefficients by calculating zero-order correlations (Pearson's *r* or ϕ) between the Rorschach scores and the criteria. Correlation coefficients were reported, rather than semipartial correlations, because correlation coefficients directly reflect the relationship of Rorschach scores with relevant clinical phenomena. For example, a correlation coefficient straightforwardly addresses the clinical question: "How strongly are *Sum of Shading* responses related to negative emotions among patients?" In contrast, a semipartial correlation coefficient addresses a more abstract and hypothetical question of doubtful clinical value: "How strongly would *Sum of Shading* responses be related to negative emotions if all patients were to give the same number of Rorschach responses (which in fact they do not)?" Furthermore, we reported correlation coefficients because they have been used in past meta-analyses on Rorschach validity and in virtually all other meta-analyses in the field of clinical assessment. We wanted our findings to be in a form that would allow researchers and clinicians to easily make comparisons with the findings from earlier relevant meta-analyses.

The third important procedural difference between the meta-analysis of Mihura et al. (2013) and our new meta-analyses concerned the use of the CS International Norms (Meyer, Erdberg, & Shaffer, 2007). In most studies, Rorschach results for a targeted sample (usually a clinical group) were compared with results for a comparison sample. Mihura et al. (pp. 559–560) used the International Norms as a comparison sample under two circumstances: (a) if a study included only a target sample but no comparison sample, or (b) if a study used the CS norms as a comparison sample.

Mihura et al. (2013) chose not to use the CS norms (e.g., Exner, 1974, 2003) as a comparison sample because many Rorschach investigators no longer believe they are accurate. For example, based on a review of the relevant literature, Viglione and Meyer (2008) concluded that the CS norms for most Form Quality scores, color scores (including *Weighted Sum of Color*), and human representational variables are inconsistent with the findings of most researchers, and that therefore other norms should be used in their place (see also Hunsley, Lee, Wood, & Taylor, in press; Wood, Nezworski, Garb, & Lilienfeld, 2001).

Although we used the International Norms as a comparison sample under the same circumstances as Mihura et al. (2013), we also used them under two additional circumstances. First, we used the International Norms as a comparison whenever a study included any set of Rorschach norms other than the International Norms as a comparison sample. For example, Yamamoto et al. (2010) compared chronic pain patients with Japanese Rorschach norms developed by Takahashi, Takahashi, and Nishio (2007). When we conducted detailed analyses (available from the first author upon request), we discovered that the Takahashi et al. norms were seriously discrepant from both the International Norms and a Japanese sample used in creating the International Norms (Nakamura, Fuchigami, & Tsugawa, 2007). Second, we substituted the International Norms for a study's comparison group if that comparison group shared the very criterion quality that the Rorschach score was intended to measure. For example, a study by Exner (1986)

used a sample of patients with schizotypal personality disorder (PD) as a comparison group, and Mihura et al. included these patients as a comparison group in the meta-analysis of *Sum of Shading*, which is a Rorschach score that supposedly measures "distressing or irritating internal stimuli" (Mihura et al., 2013, p. 550). We did not use the schizotypal group as a comparison group when computing an effect size for this study because patients with schizotypal PD show substantial elevations on neuroticism (Saulsman & Page, 2004) and therefore share the "distressing or irritating internal stimuli" that are posited as characteristic of *Sum of Shading*. Instead we used the International Norms as a comparison group.

Results

Suicide Constellation

The top row of Table 3 shows the meta-analytic results for the *Suicide Constellation*. The first set of columns on the left side of the table displays the validity coefficient reported by Mihura et al. (2013) for this variable, along with the number of participants (*N*) and the number of studies (*k*) that were used to compute the coefficient (when discussing meta-analytic results, we use the term *validity* to refer to what Mihura et al. term *external criterion validity*).

The second set of columns shows our own calculations for the same studies with the same comparison groups that were included by Mihura et al. (2013). We calculated correlations, whereas Mihura et al. calculated semipartial correlations. For our calculations we used the same comparison groups that Mihura et al. did, even if we considered these groups to be problematic. The third set of columns in Table 3 shows our calculations when we used the same studies as Mihura et al. but with the International Norms as a comparison sample for some studies.

The fourth set of columns in the table shows our calculations for what we have called "new" studies. These were relevant studies, nearly all of them dissertations, that were not included in the Mihura et al. (2013) meta-analysis. The fifth set of columns presents our validity estimates based on all relevant published and unpublished studies.

As can be seen in Table 3, Mihura et al. (2013) included four studies that examined the validity of the *Suicide Constellation* as a measure of suicide risk. On the basis of these four studies, Mihura et al. estimated the validity of the *Suicide Constellation* to be .41, which was the highest coefficient reported in the meta-analysis for a noncognitive Rorschach score. When we reviewed this same sample of studies using the same comparison groups as Mihura et al., we arrived at a slightly higher estimate of .44.

We searched the literature for any additional relevant published or unpublished studies on the *Suicide Constellation* but, finding none, we concluded that the Mihura et al. (2013) meta-analysis included all relevant studies on this variable. When we examined the four studies included in the meta-analysis, however, we identified one that seemed substantially out of place. Unlike the other three studies, the investigation by Lundbäck et al. (2006) did not examine the relationship of the *Suicide Constellation* to attempted or completed suicides. Rather, it examined the relationship of *Suicide Constellation* scores to cerebrospinal fluid levels of 5-hydroxyindoleacetic acid (5-HIAA), a serotonin metabolite that may be a risk factor for suicide attempts (Åsberg & Forslund, 2000). There are many risk factors for suicide attempts and completions, including major depressive disorder, bipolar disorder, and substance abuse (Kirkcaldy, Richardson-Vejgaard, & Merrick, 2010), but the Mihura et al. meta-analysis excluded studies that examined the relation of these well-known risk factors to the *Suicide Constellation*. Similarly and for the same reasons, we believe that the meta-analysis should have excluded the Lundbäck et al. study. The central validity question concerning the *Suicide Constellation* is not whether it is related to risk factors such as depression, bipolar disorder, substance abuse, or 5-HIAA levels, but whether it is related to suicide attempts or completions.

After eliminating the article by Lundbäck et al. (2006), three studies on the *Suicide Constellation* remained with a combined validity coefficient of .44 ($N = 434$). The first, by Meyer (1993), found no correlation ($r = .00$, $N = 90$) between *Suicide Constel-*

Table 3
Meta-Analytic Summary of Effect Sizes of Validity Coefficients for Rorschach Comprehensive System Variables: Estimates From Mihura et al. (2013) Compared With Estimates From the Present Reanalyses

				Estimates from the present reanalyses											
				Same studies included by Mihura et al.											
	Estimates reported by Mihura et al.			Same contrast samples as in Mihura et al.[a]			Some contrast samples different than Mihura et al.[b]			"New" studies not included by Mihura et al.[c]			All studies[d]		
Variable	*N*	*k*	*r*	*N*	*k*	*r*	*N*	*k*	*r*	*N*	*k*	*r*	*N*	*k*	*r*
Suicide Constellation	411	4	.41	472	4	.44	434	3	.44	0	0		434	3	.44
Weighted Sum of Color	462	5	.38	462	5	.49	481	5	.47	88	2	−.01	569	7	.41
Sum of Shading	423	5	.37	370	4	.41	288	4	.22	706	10	.04	994	14	.09
Anatomy and X-ray	423	7	.33	423	7	.25	423	7	.25	2,245	16	.04	2,668	22	.07

Note. *N* = number of participants; *k* = number of studies; *r* = mean validity coefficient (partial correlation coefficients for the estimates reported by Mihura et al., 2013, but correlation coefficients for estimates from the present reanalyses).
[a] Our calculations for the same studies included by Mihura et al. (2013) with the same comparison groups that they used. [b] Our calculations when we used the same studies as Mihura et al. but with the International Norms as a comparison sample for some studies. [c] Our calculations for "new" studies, that is, for relevant studies that were not included by Mihura et al. [d] Our calculations for all relevant published and unpublished studies.

lation scores and patient suicide attempts. The second, by Exner and Wylie (1977), described the development of the *Suicide Constellation* and reported a sizable relationship with completed suicides ($r = .54$, $N = 240$). However, this was a derivation study that involved extensive "fishing": Numerous variables were considered for inclusion in the constellation, with some being retained and others discarded depending on whether they proved predictive in the study sample. The *Suicide Constellation* was derived and validated on the same sample, and for this reason its validity coefficient was almost certainly inflated by capitalization on chance. The third study, by Fowler, Piers, Hilsenroth, Holdwick, and Padawer (2001), was a methodologically sophisticated cross-validation study that found a very large relationship between *Suicide Constellation* scores and lethal suicide attempts ($r = .51$, $N = 104$).

What conclusions can be drawn from this body of research regarding the validity of the *Suicide Constellation*? Of the three relevant studies, the first found no evidence of validity (Meyer, 1993), the second was a derivation study whose validity coefficient was probably inflated (Exner & Wylie, 1977), and the third was a well-conducted cross-validation study that yielded clear-cut positive results (Fowler et al., 2001). These findings strike us as worthy of energetic follow-up, and we hope that numerous research teams, including suicide experts with no previous connection to the Rorschach, will conduct additional studies to examine the validity of the *Suicide Constellation*. At the same time, the research base is presently too thin and inconsistent to justify claims that this Rorschach score is strongly supported as a predictor of suicide attempts or completions. Until additional evidence of validity is available, we regard the use of the *Suicide Constellation* in clinical or forensic decision making as premature.

Weighted Sum of Color

As shown in Table 3, the Mihura et al. (2013) meta-analysis included 5 studies that examined the validity of *Weighted Sum of Color* as a measure of the extent to which "emotions influence thoughts and experiences" (p. 550). On the basis of these studies, Mihura et al. estimated the validity of *Weighted Sum of Color* as .38. In contrast, when we analyzed the same studies using the same comparison groups, we arrived at a substantially higher estimate of .49. When we analyzed these same studies using the International Norms for some comparisons, our results did not change much, and we arrived at an estimate of .47.

Mihura et al. (2013, p. 560) reported that their validity coefficient for *Weighted Sum of Color*, like ours, was based on correlations rather than semipartial correlations. It is puzzling, therefore that they arrived at an estimated validity of .38, whereas we arrived at an estimate of .49 for the same studies and samples. Because the Mihura et al. meta-analysis did not include a step-by-step paper trail, we have been unable to trace their calculations and determine exactly why their numbers differed from ours. Whatever the reason, our analyses suggest that Mihura et al. may have underestimated the validity of *Weighted Sum of Color* in their sample of studies.

As shown in Table 3 and Appendix A, we identified two unpublished studies (dissertations by Saraydarian, 1990, and De Vincent, 2009, with a combined N of 88) that reported additional validity findings for *Weighted Sum of Color*. These dissertations were not included in the Mihura et al. (2013) meta-analysis because it focused exclusively on published studies. A moderator analysis showed that the combined validity of *Weighted Sum of Color* for these two dissertations ($r = -.01$) was significantly lower ($\beta = .575$, $p = < .001$) than the validity coefficient for the five published studies included by Mihura et al. ($r = .47$), indicating publication bias. When we merged the results of these two "new" dissertations with the five "old" published studies that were included by Mihura et al., the combined validity coefficient for *Weighted Sum of Color* was $r = .41$ ($k = 7$; $N = 569$), which was still higher than the validity coefficient reported by Mihura et al.

Clearly the seven studies on *Weighted Sum of Color* have yielded some impressive validity coefficients. At the same time, a close examination of these studies leads to an unexpected conclusion: *Weighted Sum of Color* probably does *not* measure emotions influencing thoughts and experiences. In the following paragraphs, we summarize the most relevant studies on *Weighted Sum of Color* and discuss their implications.

Of the seven validity studies that have examined *Weighted Sum of Color*, five have yielded positive findings. The highest validity coefficient ($r = .64$) was reported by Mishra, Khalique, and Kumar (2010), who found that 150 Indian patients with bipolar disorder in the manic phase had substantially higher *Weighted Sum of Color* scores than a comparison group of 50 normal individuals. In addition, four studies all found higher *Weighted Sum of Color* scores for patients with borderline PD compared to other patient groups, although the difference was not always statistically significant (Exner, 1986; Saraydarian, 1990; Skinstad, Troland, & Mortensen, 1999; Zodan, Charnas, & Hilsenroth, 2009; see also Zodan, 2010).

The elevated scores of manic patients reported by Mishra et al. (2010) are consistent with the hypothesis that *Weighted Sum of Color* scores reflect the influence of emotions on thoughts. However, the scores of patients with borderline PD present a puzzle. The average *Weighted Sum of Color* score of patients with borderline PD was 3.52 ($N = 155$) in the four relevant studies (Exner, 1986; Saraydarian, 1990; Skinstad et al., 1999; Zodan et al., 2009). In comparison, the average *Weighted Sum of Color* score of nonpatients in the CS International Norms is 3.11, with a standard deviation of 2.17 (Meyer et al., 2007). Using these norms as a basis for comparison, the average score for the borderline group is very close to the average of nonpatients (z score = .19, T score = 52).

If *Weighted Sum of Color* is not elevated among patients with borderline PD compared with the CS International Norms, or is elevated only slightly, then why have studies repeatedly found that this score is related to borderline PD? The reason is that the *comparison patients* in these studies have generally scored substantially below normative expectations on *Weighted Sum of Color*. Thus, any attempt to understand this variable must explain why patients with borderline PD, a condition characterized by marked emotional dysregulation (Lieb, Zanarini, Schmahl, Linehan, & Bohus, 2004) tend to have relatively normal scores on *Weighted Sum of Color*. It is difficult to reconcile this finding with the hypothesis that *Weighted Sum of Color* is a measure of emotional influences on thoughts and experiences.

Any attempt to understand *Weighted Sum of Color* must also come to terms with the findings of a dissertation by De Vincent (2009). Supervised by the well-known Rorschach researcher Robert McGrath, De Vincent collected the following data for 50 adults:

(a) *Weighted Sum of Color* scores on the Rorschach, (b) observer ratings from friends or family members, and (c) self-ratings. The observer and self-ratings were made for descriptor statements whose content reflected the standard interpretations of CS Rorschach scores (e.g., the standard interpretation of *Weighted Sum of Color*). The precise wording for these items came from the Rorschach Construct Scale (RCS), which is a rating instrument developed by Mihura, Meyer, and their colleagues for Rorschach validation research (Mihura et al., 2002; see also Mihura, Meyer, Bel-Bahar, & Gunderson, 2003). De Vincent found that the observer ratings and self-ratings for *Weighted Sum of Color* correlated $r = .48$, indicating substantial cross-method validity. In contrast, *Weighted Sum of Color* scores on the Rorschach itself were found to have near-zero correlations with the observer ratings ($r = -.08$) and self ratings ($r = -.02$).

The findings of De Vincent (2009) should not be interpreted as evidence that *Weighted Sum of Color* has no external correlates. To the contrary, as we have described, this variable has been found to be elevated in some patient conditions (e.g., mania). The reason for these effects, however, is unclear. The De Vincent study and others that we have already discussed indicate that *Weighted Sum of Color* may not measure what Mihura et al. (2013) assume it measures. Its psychological meaning is presently a mystery.

Sum of Shading

As shown in Table 3, Mihura et al. (2013) included five studies that examined the validity of *Sum of Shading* as a measure of "distressing or irritating internal stimuli" (p. 550). On the basis of these studies, Mihura et al. estimated the validity of *Sum of Shading* as .37. We attempted to analyze the same five studies but ran into a problem: Mihura et al. (p. 592) reported that they included a validity coefficient for *Sum of Shading* that was extracted from an article by García-Alba (2004) based on a comparison of "depression-related disorder versus nonclinical." Nevertheless, even after several close readings of that article we were unable to find such a comparison in a form that allowed the extraction of a validity coefficient. Therefore we set aside the article by García-Alba and based our estimates on the remaining four studies cited by Mihura et al. When we analyzed these studies using the same comparison groups as Mihura et al., we arrived at an estimated validity of .41 for *Sum of Shading* ($N = 370$, $k = 4$), which was very close to their estimate of .37. However, when the same studies were analyzed using the International Norms for some comparisons, the estimated validity of *Sum of Shading* dropped to .22 ($N = 288$, $k = 4$).

As can be seen, the decrease in estimated validity was substantial (nearly 50%) when the International Norms were used as a comparison group. This shrinkage occurred because several studies in the Mihura et al. (2013) meta-analysis used comparison groups that deviated sharply from the International Norms (e.g., the schizotypal comparison group of Exner, 1986; the Japanese norms of Yamamoto et al., 2010).

As shown in Table 3, we identified 10 additional studies (one article and nine dissertations) that reported validity findings for *Sum of Shading* but were not included in the Mihura et al. (2013) meta-analysis. A moderator analysis showed that the estimated combined validity of these 10 studies was .04 ($k = 10$; $N = 706$), which was significantly lower ($\beta = .451$, $p = .008$) than the validity coefficient of .22 for the four published studies analyzed by Mihura et al. In other words, there was clear evidence of publication bias. When we merged the results of the 10 "new" studies with the four "old" published studies, the combined validity coefficient for all 14 studies was $r = .09$ ($k = 14$; $N = 994$). This validity coefficient was quite low, but perhaps this disappointing finding could have been anticipated: As can be seen in Table 2, the test–retest reliability of *Sum of Shading* over periods of a few months is only $r = .42$, which is quite low when compared, for example, with that of most scales of the Minnesota Multiphasic Personality Inventory and the Wechsler Adult Intelligence Scale (Viglione & Meyer, 2008). Although in principle a scale with reliability of .42 can also have substantial validity, there are very few widely used clinical scales with reliability close to .40 that attain high validity. In summary, our analysis suggests that *Sum of Shading* is not a "strongly supported" Rorschach variable with "excellent" empirical support, as Mihura et al. (p. 570) concluded.

Anatomy and X-Ray

As shown in Table 3, Mihura et al. (2013) included seven studies that examined the validity of the *Anatomy and X-ray* score as a measure of "preoccupations with body vulnerability or its functioning" (p. 551). On the basis of these studies, Mihura et al. estimated the validity of the *Anatomy and X-ray* score as .33. When we analyzed these same studies using the same comparison groups but estimating correlation coefficients rather than semipartial correlations, we arrived at a somewhat lower figure of .25.

According to Meyer, Viglione, Mihura, Erard, and Erdberg (2011, p. 458), the *Anatomy and X-ray* score is "indicative of bodily concerns or preoccupations, as seen in its elevations in groups such as sexual offenders, people with a history of sexual abuse or trauma, and people with severe physical injuries." Viglione, Towns, and Lindshield (2012, p. 139) added that "anatomy content can be associated with physical disease." Accordingly, as shown in Table 3, we identified 16 additional studies that examined the relationship of the *Anatomy and X-ray* score to these criteria. If a study reported findings for the Anatomy score, but not the *Anatomy and X-ray* score, we used the Anatomy score as a substitute for the *Anatomy and X-ray* score, as Mihura et al. did (2013, p. 553, Table 1, footnote c).

These 16 "new" studies (including two articles and 14 dissertations) reported validity findings for the *Anatomy and X-ray* score but were not included in the Mihura et al. (2013) meta-analysis. Their estimated combined validity was .04 ($N = 2{,}245$; $k = 16$). A moderator analysis showed that this validity coefficient was significantly lower ($\beta = .501$, $p < .001$) than the validity coefficient of .25 for the seven published studies analyzed by Mihura et al. In other words, there is evidence of publication bias. When we merged the results of the 16 "new" studies with the seven "old" published studies, there were 22 studies in all rather than 23, because one study (Arenella, 2000; Arenella & Ornduff, 2000) contributed separate data to the new and old groups of studies. The combined validity coefficient for all 22 studies was .07 ($k = 22$; $N = 2{,}668$).

Post hoc explorations of these 22 studies lead to several tentative insights. Four of the studies, with a combined N of 1,321, examined the relationship of the *Anatomy and X-ray* score to

somatic concerns (Grosso, 1999; Hayman, 2000; Holaday & Blakeney, 1994; Moore, 2003). For example, in one of these studies (Grosso, 1999), 330 children on a psychiatric inpatient ward were rated by their parents or caretakers on the Somatic Concerns scale of the Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1977). The combined validity coefficient of the *Anatomy and X-ray* score for these four studies was –.01. These results raise questions regarding the assumption of Meyer et al. (2011) that the *Anatomy and X-ray* score is indicative of bodily concerns and preoccupations.

Three studies with a combined *N* of 237 examined whether the *Anatomy and X-ray* score is elevated among sexual offenders (Belcher, 1995; McCraw & Pegg-McNab, 1989; Smith, 2008). The combined validity coefficient for these studies was .09. These findings run counter to the claim of Meyer et al. (2011) that the *Anatomy and X-ray* score is elevated in this group of offenders.

Six studies with a combined *N* of 386 examined the relationship of the *Anatomy and X-ray* score to physical illness, including illness in a close relative (Auker-Keller, 1998; Bénony et al., 2002; Daley, 1992; Flahault & Sultan, 2010; Lottenberg Semer & Yazigi, 2009; Porcelli & Meyer, 2002). The combined validity coefficient for these studies was .22. These results offer support for the claim of Viglione et al. (2012) that the *Anatomy and X-ray* score is related to physical illness.

Finally, nine studies with a combined *N* of 554 examined the relationship of the *Anatomy and X-ray* score to a history of child sexual abuse victimization (CSA; Arenella & Ornduff, 2000; Bank, 2001; Black, 2002; Friedrich, Einbender, & McCarty, 1999; Perfect, 2004; Sakowicz, 2010; Scroppo, Drob, Weinberger, & Eagle, 1998; Shinaver, 1996; Zodan, 2010). The combined validity coefficient for these studies was .19. These findings support the claim of Meyer et al. (2011) that the *Anatomy and X-ray* score is elevated among victims of CSA, although the relationship is modest in magnitude.

As these post hoc analyses illustrate, it is currently difficult to identify the underlying construct that is measured by the *Anatomy and X-ray* score. Although research shows that elevations in this score are somewhat more common among individuals with a physical illness or a history of sexual abuse than among other individuals, the research suggests that they are not a sign of somatic concerns or bodily preoccupations. Furthermore, whatever quality is being measured by these responses must be very transitory. As shown in Table 2, the test–retest reliability of the *Anatomy and X-ray* score over periods of a few months is approximately $r = .53$, which is substantially lower than would be expected for a score that measures a stable characteristic of the person (Ayearst & Bagby, 2010; Hunsley & Mash, 2008).

Conclusions Regarding Noncognitive Rorschach Scores

Our analyses of the four noncognitive Rorschach scores support several conclusions.

1. An impressive feature of the Mihura et al. (2013) meta-analysis is its thorough search of the published Rorschach literature. When we conducted our own exhaustive searches, we identified only a handful of relevant published studies that the authors of the meta-analysis overlooked.

2. Mihura et al. (2013) based many of their validity estimates on semipartial correlations. We wish that they had also reported correlations, for reasons we have already explained. Nevertheless, our reanalyses do not suggest that their computational procedures led to systematic over- or underestimation of Rorschach validity.

3. Mihura et al. (2013) included only published studies in their meta-analysis. Our analyses indicate that this decision led to two important negative consequences. First, a large number of relevant studies, primarily dissertations, were omitted from the meta-analysis: For two of the four Rorschach scores that we examined—*Sum of Shading* and *Anatomy and X-ray*—more than half of all relevant studies were omitted. Second, because unpublished studies were omitted, the meta-analysis sometimes substantially overestimated the validity of Rorschach scores in the complete scientific literature. For three of the four scores we examined, the validity of unpublished dissertations was significantly less than that of published studies. For two of these three scores, the estimated validity was substantially reduced when unpublished dissertations were taken into account: When Mihura et al. (2013) included only published studies on *Sum of Shading* and the *Anatomy and X-ray* score, the validity of both of these variables was estimated as above .30. However, when we reanalyzed the results from all studies, including unpublished dissertations, the validity of *Sum of Shading* and *Anatomy and X-ray* scores was found to be below .10. Hence, we found clear evidence of publication bias. Some might argue that Mihura et al. (2013) made a wise decision in including only published studies, because these studies are peer reviewed and thus may be of better methodological quality than unpublished dissertations. Such an argument would be based on conjecture rather than evidence from the Rorschach literature; it runs contrary to the conclusions of many experts in meta-analysis who recommend inclusion of both unpublished and published studies to detect possible publication bias or file drawer effects (Begg, 1994; Lipsey & Wilson, 2001). All or nearly all of the dissertations that we identified and included in our reanalyses were supervised by experienced researchers and approved by committees of academics with research training. Many of the dissertations were supervised or guided by leading Rorschach researchers (e.g., Philip Erdberg, Mark Hilsenroth, Robert McGrath, David Pogge, John Stokes, and David Viglione). Hence, there is no a priori reason to assume that these studies are of lower methodological quality than published studies. In fact, the only meta-analysis to systematically compare published Rorschach studies and Rorschach dissertations with respect to methodological quality found no difference (Wood et al., 2010).

4. Our reexamination of the literature also raised serious questions about the construct validity of three of the four

Rorschach scores that Mihura et al. (2013) reported as having the highest validity. Our meta-analyses indicate that the validity coefficient of *Sum of Shading* as a purported measure of "distressing or irritating internal stimuli" was .09. Similarly, *Anatomy and X-ray* responses correlated –.01 with somatic concerns, and the overall validity of this Rorschach variable as a purported measure of preoccupation with body vulnerability or bodily functioning was .07. These findings undermine claims regarding the construct validity of these variables. It is unclear whether *Sum of Shading* measures any important psychological quality. In contrast, however, *Anatomy and X-ray* scores have consistently shown modest elevations among individuals with a physical illness or history of child sexual abuse, suggesting that these scores are related to some unspecified quality of psychological interest, though apparently not to bodily preoccupation as posited by Mihura et al.

Similar difficulties surround the construct validity of *Weighted Sum of Color*. The relatively few studies to examine this variable have yielded an impressive overall validity coefficient of .41. However, these studies also show that patients with borderline PD do not receive elevated scores on *Weighted Sum of Color*, raising serious questions about whether this score measures emotional influences on thoughts and experiences.

Summary and Recommendations

We concur with Mihura et al. (2013) that many cognitive Rorschach scores are valid, and that their validity is usually higher than that found for noncognitive Rorschach scores. Our conclusions regarding noncognitive Rorschach scores are less encouraging. To evaluate the strengths and weaknesses of the Mihura et al. (2013) meta-analysis, we conducted new meta-analyses for the four noncognitive Rorschach scores with the highest reported validity coefficients. On the positive side, we found that the Mihura et al. meta-analysis included virtually all the relevant published studies on these variables, and that the estimated validity coefficients reported by the authors provided an unbiased and trustworthy summary of the published literature. However, the Mihura et al. (2013) meta-analysis included only published studies and therefore omitted dissertations with relevant results. Many of the dissertations reported relatively low validity coefficients, suggesting significant publication bias for *Weighted Sum of Color*, *Sum of Shading*, and the *Anatomy and X-ray* score. We also found evidence that raised serious questions about these scores' construct validity.

On the basis of the foregoing evidence, we offer the following conclusions and recommendations:

1. There is abundant scientific evidence to justify clinical use of CS scores based on *Form Quality* and *Deviant Verbalizations*. Nevertheless, the CS norms for several of these scores, especially Form Quality scores, have been criticized as inaccurate, with a tendency to overpathologize normal individuals (Hunsley et al., in press; Viglione & Meyer, 2008; Wood et al., 2001). We therefore recommend that the International Norms of the Rorschach (Meyer et al., 2007), though far from perfect, be used instead, because they are less likely to misidentify relatively normal patients as psychologically disordered.

2. There is substantial scientific evidence that several CS scores based on *Productivity* and *Complexity/Synthesis* differentiate cognitively impaired groups (e.g., patients with Alzheimer's dementia or head injury) from normal groups, and that these scores are correlated with intelligence tests (Mihura et al., 2013). However, there is little evidence that these scores are as valid as commonly used intelligence tests for identifying impaired groups or for measuring cognitive ability. At present, therefore, cognitive impairment and cognitive ability are more appropriately assessed by a well-validated intelligence test than by the Rorschach.

3. Nearly 15 years ago, one of the authors of this Comment published a recommendation that a moratorium be placed upon use of the Rorschach in clinical and forensic settings because of the test's weak scientific foundation (Garb, 1999). He and the other authors of this Comment agree that, in light of the compelling evidence laid out by Mihura et al. (2013), the time has come to withdraw this recommendation so far as it applies to the Cognitive Quartet of Rorschach scores. We are convinced that these scores provide valid information regarding cognitive ability and cognitive impairment that can be helpful in some applied and research settings, provided that interpretations are based on comparisons with the CS International Norms (Meyer et al., 2007) rather than with the regular CS norms (Exner, 2003).

4. We extended the Mihura et al. (2013) meta-analysis for the *Suicide Constellation*, *Weighted Sum of Color*, *Sum of Shading*, and *Anatomy and X-ray* scores. Our analyses revealed that none of these scores was truly "strongly supported." Our analyses indicated that significant publication bias was a problem for most of these scores, and we found cases in which Mihura et al. substantially overestimated the validity of Rorschach scores due to the exclusion of unpublished studies. We also found substantial evidence that most of these Rorschach scores probably do not measure what they are purported to measure.

5. In our view, the most important contribution of the Mihura et al. (2013) meta-analysis is as a guide to future research. The authors have identified many Rorschach variables with mixed but potentially promising support, including the *Suicide Constellation*. The Mihura et al. meta-analysis should be seen as a valuable impetus for further investigation, rather than the final word regarding the scientific status of clinical psychology's most contested measure.

References

References marked with an asterisk indicate studies included in the meta-analyses.

*Arenella, J. (2000). *Rorschach indicators of body image disruption and precocious maturity in sexually and physically abused girls* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9959123)

*Arenella, J., & Ornduff, S. R. (2000). Manifestations of bodily concern in sexually abused girls. *Bulletin of the Menninger Clinic, 64,* 530–542.

Åsberg, M., & Forslund, K. (2000). Neurobiological aspects of suicidal behavior. *International Review of Psychiatry, 12,* 62–74. doi:10.1080/09540260074148

*Auker-Keller, A. A. (1998). *Female breast cancer patients: The differences in psychological and family functioning based on severity of illness* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9934015)

Ayearst, L. E., & Bagby, R. M. (2010). Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 23–61). New York, NY: Guilford Press.

*Bank, C. A. (2001). *Predicting sexually abused victims using the Rorschach Inkblot Test* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3014406)

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York, NY: Russell Sage Foundation.

*Belcher, P. L. (1995). *A comparison of male adolescent sexual offenders using the Rorschach* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9536333)

*Bénony, H., Daloz, L., Bungener, C., Chahraoui, K., Frenay, C., & Auvin, J. (2002). Emotional factors and subjective quality of life in subjects with spinal cord injuries. *American Journal of Physical Medicine & Rehabilitation, 81,* 437–445. doi:10.1097/00002060-200206000-00008

*Black, E. M. (2002). *The use of the Rorschach Inkblot Technique in the detection and diagnosis of child sexual abuse* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3068669)

*Burns, J. C. (2003). *Use of the Rorschach to identify trauma in a sample of homeless and indigent women* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3084479)

*Connolly, E. L. (1997). *Chronic stress in mothers of severely disabled children: A Rorschach study* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9730214)

*Daley, B. S. (1992). *A descriptive study of the Rorschach and chronic pain* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9236716)

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York, NY: Free Press.

*De Vincent, T. (2009). *A validation of the controls cluster of the Rorschach Comprehensive System* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3353725)

Exner, J. E. (1974). *The Rorschach: A Comprehensive System* (Vol. *1*). New York, NY: Wiley.

*Exner, J. E. (1986). Some Rorschach data comparing schizophrenics with borderline and schizotypal personality disorders. *Journal of Personality Assessment, 50,* 455–471. doi:10.1207/s15327752jpa5003_14

Exner, J. E. (1996). A comment on "The Comprehensive System for the Rorschach: A critical examination." *Psychological Science, 7,* 11–13. http://dx.doi.org/10.1111/j.1467-9280.1996.tb00659.x

Exner, J. E. (2003). *The Rorschach: A Comprehensive System. Vol. 1: Basic foundations and principles of interpretation* (4th ed.). New York, NY: Wiley.

Exner, J. E., & Erdberg, P. (2005). *The Rorschach: A Comprehensive System. Vol. 2: Advanced interpretation* (4th ed.). New York, NY: Wiley.

*Exner, J. E., & Wylie, J. (1977). Some Rorschach data concerning suicide. *Journal of Personality Assessment, 41,* 339–348. doi:10.1207/s15327752jpa4104_1

*Flahault, C., & Sultan, S. (2010). On being a child of an ill parent: A Rorschach investigation of adaptation to parental cancer compared to other illnesses. *Rorschachiana, 31,* 43–69. doi:10.1027/1192-5604/a000004

*Fowler, J. C., Piers, C., Hilsenroth, M. J., Holdwick, D. J., & Padawer, J. R. (2001). The Rorschach Suicide Constellation: Assessing various degrees of lethality. *Journal of Personality Assessment, 76,* 333–351. doi:10.1207/S15327752JPA7602_13

*Friedrich, W. N., Einbender, A. J., & McCarty, P. (1999). Sexually abused girls and their Rorschach responses. *Psychological Reports, 85,* 355–362. doi:10.2466/PR0.85.6.355-362

Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot in clinical and forensic settings. *Assessment, 6,* 313–317. doi:10.1177/107319119900600402

*García-Alba, C. (2004). Anorexia and depression: Depressive comorbidity in anorexic adolescents. *The Spanish Journal of Psychology, 7,* 40–52. Retrieved from http://www.ucm.es/info/Psi/docs/journal/

*Gregg, P. A. (1998). *The effect of impression management on correlations between Rorschach and MMPI-2 variables* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9807469)

*Grosso, C. G. (1999). *An investigation of the construct and concurrent validity of variables from the Comprehensive System for the Rorschach in a child psychiatric inpatient population* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9940446)

*Hayman, J. A. (2000). *An actuarial and confirmatory investigation of the interpersonal and self-perception clusters of the Rorschach Comprehensive System* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9965254)

*Holaday, M., & Blakeney, P. (1994). A comparison of psychologic functioning in children and adolescents with severe burns on the Rorschach and the Child Behavior Checklist. *Journal of Burn Care & Rehabilitation, 15,* 412–415. doi:10.1097/00004630-199409000-00006

*Holaday, M., Moak, J., & Shipley, M. A. (2001). Rorschach protocols from children and adolescents with Asperger's disorder. *Journal of Personality Assessment, 76,* 482–495. doi:10.1207/S15327752JPA7603_09

Hunsley, J., Lee, C. M., Wood, J. M., & Taylor, W. (in press). Controversial and questionable assessment techniques. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in contemporary clinical psychology* (2nd ed.). New York, NY: Guilford Press.

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3–14). New York, NY: Oxford University.

Kirkcaldy, B. D., Richardson-Vejgaard, R., & Merrick, J. (2010). Suicide risk in adolescents and adults. In J. M. Brown & E. A. Campbell (Eds.), *Cambridge handbook of forensic psychology* (pp. 327–334). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511730290.041

*Leifer, M., Shapiro, J. P., Martone, M. W., & Kassem, L. (1991). Rorschach assessment of psychological functioning in sexually abused girls. *Journal of Personality Assessment, 56,* 14–28. doi:10.1207/s15327752jpa5601_2

Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *The Lancet, 364,* 453–461.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Long, J. S. (1995). *A comparative study between the Rorschach and MMPI-A in predicting symptoms and behaviors with inpatient adolescents* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9536085)

*Lottenberg Semer, N., & Yazigi, L. (2009). The Rorschach and the body: The study of self-esteem in enuretic children through the Rorschach method. *Rorschachiana, 30,* 3–25. doi:10.1027/1192-5604.30.1.3

*Lundbäck, E., Forslund, K., Rylander, G., Jokinen, J., Nordström, P., Nordström, A.-L., & Åsberg, M. (2006). CSF 5-HIAA and the Rorschach test in patients who have attempted suicide. *Archives of Suicide Research, 10,* 339–345. doi:10.1080/13811110600790942

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516. doi:10.1037/0003-066X.52.5.509

*McCraw, R. K., & Pegg-McNab, J. (1989). Rorschach comparisons of male juvenile sex offenders and nonsex offenders. *Journal of Personality Assessment, 53,* 546–553. doi:10.1207/s15327752jpa5303_11

*Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment, 60,* 153–180. doi:10.1207/s15327752jpa6001_13

Meyer, G. J., Erdberg, P., & Shaffer, T. W. (2007). Toward international normative reference data for the Comprehensive System. *Journal of Personality Assessment, 89,* S201–S216. doi:10.1080/00223890701629342

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual.* Toledo, OH: Rorschach Performance Assessment System.

Mihura, J. L., Meyer, G. J., Bel-Bahar, T., & Gunderson, J. (2003). Correspondence among observer ratings of Rorschach, Big Five Model, and *DSM–IV* personality disorder constructs. *Journal of Personality Assessment, 81,* 20–39. doi:10.1207/S15327752JPA8101_03

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin, 139,* 548–605. doi:10.1037/a0029406

Mihura, J. L., Meyer, G. J., Viglione, D. J., Ritzier, B., Kaser-Boyd, N., Adrian, C., . . . Bornstein, R. F. (2002). *The Rorschach Construct Scale: Observer-rating (mixed gender, female, & male formats) and self-rating forms* [Unpublished scale and tables]. University of Toledo, Toledo, OH.

*Mishra, D., Khalique, A., & Kumar, R. (2010). Rorschach profile of manic patients. *Journal of Projective Psychology & Mental Health, 17,* 158–164. Retrieved from http://www.somaticinkblots.com/SISJournal/tabid/61/Default.aspx

*Moore, T. L. (2003). *Sexually victimized and non-sexually victimized adolescent sexual offenders: One group or two?* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 8217265)

*Nakamura, N., Fuchigami, Y., & Tsugawa, R. (2007). Rorschach comprehensive system data for a sample of 240 adult nonpatients from Japan. *Journal of Personality Assessment, 89*(Suppl. 1), S97–S102. doi:10.1080/00223890701583291

*Owens, T. H. (1982). *Personality characteristics of female psychotherapy patients with a history of incest* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 8217265)

*Perfect, M. M. (2004). *Incremental validity of the Minnesota Multiphasic Personality Inventory (MMPI-A) and Rorschach Inkblot Test in predicting the number and severity of adolescents' maltreatment histories* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3143446)

*Porcelli, P., & Meyer, G. J. (2002). Construct validity of Rorschach variables for alexithymia. *Psychosomatics: Journal of Consultation and Liaison Psychiatry, 43,* 360–369. doi:10.1176/appi.psy.43.5.360

*Ritsher, J. E. B. (1997). *Depression assessment of Russian psychiatric patients: Validity of MMPI and Rorschach scales* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9816520)

*Sakowicz, K. (2010). *Rorschach correlates of childhood sexual abuse* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3437950)

*Saraydarian, L. (1990). *Diagnostic clarity of borderline personality disorder: A Rorschach study* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9027329)

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review, 23,* 1055–1085. doi:10.1016/j.cpr.2002.09.001

*Scroppo, J. C., Drob, S. L., Weinberger, J. L., & Eagle, P. (1998). Identifying dissociative identity disorder: A self-report and projective study. *Journal of Abnormal Psychology, 107,* 272–284. doi:10.1037/0021-843X.107.2.272

*Shinaver, C. S. (1996). *The impact of sexual trauma on the information processing of boys* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9706798)

*Skinstad, A. H., Troland, K., & Mortensen, J. K. (1999). Rorschach responses in borderline personality disorder with alcohol dependence. *European Journal of Psychological Assessment, 15,* 133–142. doi:10.1027//1015-5759.15.2.133

*Smith, A. L. (2008). *Verbal and social functioning among boys in the criminal justice system* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3320210)

Sultan, S., Andronikof, A., Réveillère, C., & Lemmel, G. (2006). A Rorschach stability study in a nonpatient adult sample. *Journal of Personality Assessment, 87,* 330–348. doi:10.1207/s15327752jpa8703_13

Sultan, S., & Meyer, G. J. (2009). Does productivity impact the stability of Rorschach scores? *Journal of Personality Assessment, 91,* 480–493. doi:10.1080/00223890903088693

Takahashi, M., Takahashi, Y., & Nishio, H. (2007). *Rorschach-test-kaishakuho* [Principles of interpretation of the Comprehensive Rorschach Test]. Tokyo, Japan: Kongo Shuppan.

*Verias, E. A. (2007). *Further investigation into the Rorschach and the utility of a modified DEPI in relation to adolescent depression* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3270246)

Viglione, D. J., & Meyer, G. J. (2008). An overview of Rorschach psychometrics for forensic practice. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L. A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (pp. 21–53). New York, NY: Routledge.

Viglione, D. J., Towns, B., & Lindshield, D. (2012). Understanding and using the Rorschach Inkblot Test to assess post-traumatic conditions. *Psychological Injury and Law, 5,* 135–144. doi:10.1007/s12207-012-9128-5

Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, R. D. (1977). *Multidimensional description of child personality: A manual for the Personality Inventory for Children.* Los Angeles, CA: Western Psychological Services.

Wood, J. M., Krishnamurthy, R., & Archer, R. P. (2003). Three factors of the Comprehensive System for the Rorschach and their relationship to Wechsler IQ scores in an adolescent sample. *Assessment, 10,* 259–265. doi:10.1177/1073191103255493

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. doi:10.1002/(SICI)1097-4679(200003)56:3<395::AID-JCLP15>3.0.CO;2-O

Wood, J. M., Lilienfeld, S. O., Nezworski, M. T., Garb, H. N., Wildermuth, J., & Allen, K. H. (2010). Validity of Rorschach Inkblot scores for discriminating psychopaths from nonpsychopaths in forensic populations: A meta-analysis. *Psychological Assessment, 22,* 336–349. doi:10.1037/a0018998

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. (2001). The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice, 8,* 350–373. doi:10.1093/clipsy.8.3.350

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.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10. doi:10.1111/j.1467-9280.1996.tb00658.x

*Yamamoto, K., Kanbara, K., Mutsuura, H., Ban, I., Mizuno, Y., Abe, T., . . . Fukunaga, M. (2010). Psychological characteristics of Japanese patients with chronic pain assessed by the Rorschach test. *BioPsychoSocial Medicine, 4,* Article 20. doi:10.1186/1751-0759-4-20

*Zodan, J. (2010). *Rorschach assessment of childhood sexual abuse and borderline pathology: A comparison of clinical samples* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3425222)

*Zodan, J., Charnas, J., & Hilsenroth, M. J. (2009). Rorschach assessment of reality testing, affect and object representation of borderline pathology: A comparison of clinical samples. *Bulletin of the Menninger Clinic, 73,* 121–142. doi:10.1521/bumc.2009.73.2.121

(*Appendix follows*)

Appendix

Meta-Analysis References

Table A1
Validity Criteria per Rorschach Variable in the Meta-Analyses

Rorschach variable & no. of studies	Validity criteria per Rorschach variable
	Suicide Constellation: Suicide risk
1	Suicide, effected vs. controls (patients with and without history of suicide attempts; nonclinical)[13]
1	Suicide attempters, near-lethal vs. parasuicidal and nonsuicidal inpatients[15]
1	Patients with serious suicide attempt vs. patients without serious attempt[28]
1	[–] CSF 5-HIAA (measure of serotonin turnover) in suicide-attempt patients[26, a]
	Weighted Sum of Color: Emotions influence thoughts and experiences
1	Bipolar, with mania and psychosis vs. nonclinical[29]
1	*Weighted Sum of Color* score on Rorschach Construct Scale, ratings by family or friends[11]
1	[–] Asperger's disorder vs. behavior- and emotional-problem controls[22]
3	Patients with borderline PD vs. patients with other diagnoses (schizotypal PD[12]; mixed PD[40]; schizotypal, narcissistic, or histrionic PD[30])
1	Patients with borderline PD vs. patients without borderline PD[43, 44, b]
	Sum of Shading: Distressing or irritating internal stimuli
1	Adult patients with history of incest victimization vs. CS International Norms[31]
1	Sexually abused children (genital contact) vs. medical patients (not in physical distress)[23]
3	Patients with depressive disorders or severe depression vs. non-depressed (CS International Norms[34]; non-depressed patients[41]; nonclinical[17, c])
1	Depressive symptoms, clinician ratings[24]
3	Patients with borderline PD vs. CS International Norms[12, 39, 43, 44, d]
2	Chronic pain patients vs. CS International Norms[10, 42, e]
1	Breast cancer patients vs. CS International Norms[3]
1	Mothers of disabled children vs. mothers of healthy children[9]
1	Homeless indigent women vs. CS International Norms[8]
1	Custody litigants vs. CS International Norms[18]
	Anatomy and X-ray: Preoccupations with body vulnerability or its functioning
4	Somatization or somatic concerns (caretaker ratings[19, 21, 30]; therapist ratings[20])
1	Physically abused children or adolescents vs. non-abused controls[1]
5	Sexually abused children or adolescents vs. non-abused controls[2, 4, 7, 16, 35]
1	Adult patients with history of sexual abuse (with or without borderline PD) vs. controls (no sexual abuse, no borderline PD)[43]
1	Physical or sexual abuse vs. neglect but no abuse[38]
1	Severity of sexual and physical abuse[32]
1	Dissociative identity disorder (high sexual or physical abuse rates) vs. psychiatric controls[37]
3	Sex offenders vs. controls (CS International Norms[5]; non-sexual offenders[27, 40])
1	Breast cancer patients vs. CS International Norms[3]
1	Spinal cord injury (paraplegic and tetraplegic) vs. nonclinical[6]
1	Chronic pain patients vs. CS International Norms[10]
1	Patients with irritable bowel syndrome vs. CS International Norms[33]
1	Enuretic children versus non-enuretic controls[25]
1	Psychological effects of parental cancer vs. more stable illnesses (diabetes or respiratory)[14]

Note. Superscripted numbers refer to the meta-analysis references listed after this table. Number of studies = total number of studies for correlation results or number of target samples for group contrasts. [–] = an inverse association was hypothesized. 5-HIAA = 5-hydroxyindoleacetic acid. PD = personality disorder.

[a] Lundbäck et al. (2006) was included by Mihura et al. (2013) but excluded from central analyses in present meta-analysis. [b] Samples of Zodan et al. (2009) and Zodan (2010) overlapped. Mihura et al. (2013) used Zodan et al. (2009) but central analyses of *Weighted Sum of Color* in present meta-analysis used Zodan (2010). [c] García-Alba (2004) was included by Mihura et al (2013) but was not included in present meta-analysis because no relevant effect size could be found. [d] Mihura et al. (2013) used patients with schizotypal PD as comparison for Exner (1986). Mihura et al. (2013) used Zodan et al. (2009) and compared borderline patients with non-borderline patients, but central analyses of *Sum of Shading* in present meta-analysis used Zodan (2010) and compared borderline patients with CS International Norms. [e] Mihura et al. (2013) used Takahashi et al. (2007) norms as comparison group for Yamamoto et al. (2010).

(*Appendix continues*)

References are listed by number to correspond to the superscripted criterion variables in the table above.

[1]Arenella, J. (2000). *Rorschach indicators of body image disruption and precocious maturity in sexually and physically abused girls* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9959123)

[2]Arenella, J., & Ornduff, S. R. (2000). Manifestations of bodily concern in sexually abused girls. *Bulletin of the Menninger Clinic, 64,* 530–542.

[3]Auker-Keller, A. A.(1998). *Female breast cancer patients: The differences in psychological and family functioning based on severity of illness* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9934015)

[4]Bank, C. A. (2001). *Predicting sexually abused victims using the Rorschach Inkblot Test* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3014406).

[5]Belcher, P. L.(1995). *A comparison of male adolescent sexual offenders using the Rorschach* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9536333)

[6]Bénony, H., Daloz, L., Bungener, C., Chahraoui, K., Frenay, C., & Auvin, J. (2002). Emotional factors and subjective quality of life in subjects with spinal cord injuries. *American Journal of Physical Medicine & Rehabilitation, 81,* 437–445. doi:10.1097/00002060-200206000-00008

[7]Black, E. M.(2002). *The use of the Rorschach Inkblot Technique in the detection and diagnosis of child sexual abuse* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3068669)

[8]Burns, J. C.(2003). *Use of the Rorschach to identify trauma in a sample of homeless and indigent women* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3084479)

[9]Connolly, E. L.(1997). *Chronic stress in mothers of severely disabled children: A Rorschach study* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9730214)

[10]Daley, B. S.(1992). *A descriptive study of the Rorschach and chronic pain* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9236716)

[11]De Vincent, T.(2009). *A validation of the controls cluster of the Rorschach Comprehensive System* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3353725)

[12]Exner, J. E. (1986). Some Rorschach data comparing schizophrenics with borderline and schizotypal personality disorders. *Journal of Personality Assessment, 50,* 455–471. doi:10.1207/s15327752jpa5003_14

[13]Exner, J. E., & Wylie, J. (1977). Some Rorschach data concerning suicide. *Journal of Personality Assessment, 41,* 339–348. doi:10.1207/s15327752jpa4104_1

[14]Flahault, C., & Sultan, S. (2010). On being a child of an ill parent: A Rorschach investigation of adaptation to parental cancer compared to other illnesses. *Rorschachiana, 31,* 43–69. doi:10.1027/1192-5604/a000004

[15]Fowler, J. C., Piers, C., Hilsenroth, M. J., Holdwick, D. J., & Padawer, J. R. (2001). The Rorschach Suicide Constellation: Assessing various degrees of lethality. *Journal of Personality Assessment, 76,* 333–351. doi:10.1207/S15327752JPA7602_13

[16]Friedrich, W. N., Einbender, A. J., & McCarty, P. (1999). Sexually abused girls and their Rorschach responses. *Psychological Reports, 85,* 355–362. doi:10.2466/PR0.85.6.355-362

[17]García-Alba, C. (2004). Anorexia and depression: Depressive comorbidity in anorexic adolescents. *The Spanish Journal of Psychology, 7,* 40–52. Retrieved from http://www.ucm.es/info/Psi/docs/journal/

[18]Gregg, P. A.(1998). *The effect of impression management on correlations between Rorschach and MMPI-2 variables* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9807469)

[19]Grosso, C. G.(1999). *An investigation of the construct and concurrent validity of variables from the Comprehensive System for the Rorschach in a child psychiatric inpatient population* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9940446)

[20]Hayman, J. A.(2000). *An actuarial and confirmatory investigation of the interpersonal and self-perception clusters of the Rorschach Comprehensive System* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9965254)

[21]Holaday, M., & Blakeney, P. (1994). A comparison of psychologic functioning in children and adolescents with severe burns on the Rorschach and the Child Behavior Checklist. *Journal of Burn Care & Rehabilitation, 15,* 412–415. doi:10.1097/00004630-199409000-00006

[22]Holaday, M., Moak, J., & Shipley, M. A. (2001). Rorschach protocols from children and adolescents with Asperger's disorder. *Journal of Personality Assessment, 76,* 482–495. doi:10.1207/S15327752JPA7603_09

[23]Leifer, M., Shapiro, J. P., Martone, M. W., & Kassem, L. (1991). Rorschach assessment of psychological functioning in sexually abused girls. *Journal of Personality Assessment, 56,* 14–28. doi:10.1207/s15327752jpa5601_2

[24]Long, J. S.(1995). *A comparative study between the Rorschach and MMPI-A in predicting symptoms and behaviors with inpatient adolescents* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9536085)

[25]Lottenberg Semer, N., & Yazigi, L. (2009). The Rorschach and the body: The study of self-esteem in enuretic children through the Rorschach method. *Rorschachiana, 30,* 3–25. doi:10.1027/1192-5604.30.1.3

[26]Lundbäck, E., Forslund, K., Rylander, G., Jokinen, J., Nordström, P., Nordström, A.-L., & Åsberg, M. (2006). CSF 5-HIAA and the Rorschach test in patients who have attempted suicide. *Archives of Suicide Research, 10,* 339–345. doi:10.1080/13811110600790942

[27]McCraw, R. K., & Pegg-McNab, J. (1989). Rorschach comparisons of male juvenile sex offenders and nonsex offenders. *Journal of Personality Assessment, 53,* 546–553. doi:10.1207/s15327752jpa5303_11

[28]Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment, 60,* 153–180. doi:10.1207/s15327752jpa6001_13

[29]Mishra, D., Khalique, A., & Kumar, R. (2010). Rorschach profile of manic patients. *Journal of Projective Psychology & Mental Health, 17,* 158–164. Retrieved from http://www.somaticinkblots.com/SISJournal/tabid/61/Default.aspx

[30]Moore, T. L.(2003). *Sexually victimized and non-sexually victimized adolescent sexual offenders: One group or two?* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 8217265)

[31]Owens, T. H.(1982). *Personality characteristics of female psychotherapy patients with a history of incest* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 8217265)

[32]Perfect, M. M.(2004). *Incremental validity of the Minnesota Multiphasic Personality Inventory (MMPI-A) and Rorschach Inkblot Test in predicting the number and severity of adolescents' maltreatment histories* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3143446)

[33]Porcelli, P., & Meyer, G. J. (2002). Construct validity of Rorschach variables for alexithymia. *Psychosomatics: Journal of Consultation and Liaison Psychiatry, 43,* 360–369. doi:10.1176/appi.psy.43.5.360

[34]Ritsher, J. E. B.(1997). *Depression assessment of Russian psychiatric patients: Validity of MMPI and Rorschach scales* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9816520)

[35]Sakowicz, K.(2010). *Rorschach correlates of childhood sexual abuse* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3437950)

[36]Saraydarian, L.(1990). *Diagnostic clarity of borderline personality disorder: A Rorschach study* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9027329)

[37]Scroppo, J. C., Drob, S. L., Weinberger, J. L., & Eagle, P. (1998). Identifying dissociative identity disorder: A self-report and projective study. *Journal of Abnormal Psychology, 107,* 272–284. doi:10.1037/0021–843X.107.2.272

[38]Shinaver, C. S.(1996). *The impact of sexual trauma on the information processing of boys* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 9706798)

[39]Skinstad, A. H., Troland, K., & Mortensen, J. K. (1999). Rorschach responses in borderline personality disorder with alcohol dependence. *European Journal of Psychological Assessment, 15,* 133–142. doi: 10.1027//1015–5759.15.2.133

[40]Smith, A. L.(2008). *Verbal and social functioning among boys in the criminal justice system* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3320210)

[41]Verias, E. A.(2007). *Further investigation into the Rorschach and the utility of a modified DEPI in relation to adolescent depression* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3270246)

[42]Yamamoto, K., Kanbara, K., Mutsuura, H., Ban, I., Mizuno, Y., Abe, T., . . . Fukunaga, M. (2010). Psychological characteristics of Japanese patients with chronic pain assessed by the Rorschach test. *BioPsychoSocial Medicine, 4,* Article 20. doi:10.1186/1751-0759-4-20

[43]Zodan, J.(2010). *Rorschach assessment of childhood sexual abuse and borderline pathology: A comparison of clinical samples* (Doctoral dissertation). Available from ProQuest dissertations and Theses database. (UMI No. 3425222)

[44]Zodan, J., Charnas, J., & Hilsenroth, M. J. (2009). Rorschach assessment of reality testing, affect and object representation of borderline pathology: A comparison of clinical samples. *Bulletin of the Menninger Clinic, 73,* 121–142. doi:10.1521/bumc.2009.73.2.121

Received September 3, 2013
Revision received November 16, 2013
Accepted December 2, 2013 ■