



A Scientific Critique of Rorschach Research

Revisiting Exner's *Issues and Methods in Rorschach Research* (1995)

Jason M. Smith¹, Carl B. Gacono², Patrick Fontan³,
Enna E. Taylor⁴, Ted B. Cunliffe⁵, and Anne Andronikof⁶

¹FCC Hazelton, Bruceton Mills, WV, USA

²Private Practice, Asheville, NC, USA

³Circonscription Saint Denis 1, Paris, France

⁴Private Practice, San Francisco, CA, USA

⁵Private Practice, Miami, FL, USA

⁶Laboratoire IPSé, Université Paris Ouest, Nanterre, France

Abstract: Exner's (1995a) *Issues and Methods in Rorschach Research* provided a standard of care for conducting Rorschach research; however, the extent to which studies have followed these guidelines has not been examined. Similarly, meta-analytic approaches have been used to comment on the validity of Exner's Comprehensive System (CS) variables without an evaluation as to the extent that individual studies have conformed to the proposed methodological criteria (Exner, 1995a; Gacono, Loving, & Bodholdt, 2001). In this article, 210 studies cited in recent meta-analyses by Mihura, Meyer, Dumitrascu, and Bombel (2013) were examined. The studies were analyzed in terms of being research on the Rorschach versus research with the Rorschach and whether they met the threshold of validity/generalizability related to specific Rorschach criteria. Only 104 of the 210 (49.5%) studies were research on the Rorschach and none met all five Rorschach criteria assessed. Trends and the need for more stringent methods when conducting Rorschach research were presented.

Keywords: Rorschach, Exner Comprehensive System, psychological assessment

Exner's (1995a) *Issues and Methods in Rorschach Research* set a standard of care for conducting Rorschach research. At that time, he stated: "A huge number of published investigations ... are clearly marked by errors in design, implementation, and/or analysis" (Exner, 1995b, p. 3). Since Exner's (1995a) cautions, no one has examined the degree to which Rorschach research has conformed to the guidelines offered by the chapter authors. Meta-analytic findings, in particular, have been accepted at face value with little consideration for the degree to which individual studies fell within the parameters outlined by Exner (1995a) and others (e.g., Gacono, Loving, & Bodholdt, 2001).

The following is a brief overview of each chapter in *Issues and Methods in Rorschach Research*. In addition to highlighting issues to be considered when conducting Rorschach research, Exner (1995b) cautioned the researcher to be aware of the complexity involved with this type of research. He presented many essential ideas such as: (1) offering cautions about small sample size (at least 15 subjects are needed in an experimental group for every dependent variable to be included in the analysis); (2) presenting considerations related to personality/response style (Lambda & Introversive/Extratsensive); (3) suggesting how normative data can be misused (any comparisons with normative data would find meaningless significant results; “ingenuous conclusion that they have made a great discovery,” p. 17); (4) cautioning against overemphasizing or overgeneralizing results; and (5) not using the “shotgun study” (p. 14) where there is no a priori model specified before analyzing Rorschach data. This latter approach can increase the probability of finding significant results by chance (Type I error; also see Viglione, 1995; Weiner, 1995b). Further, and most importantly, Exner stressed that not everything “appearing in the literature was truth” (p. 4).

Dies (1995a, 1995b) discussed issues with sample size (a sample size of less than 20 could lead to deceptive results due to deviant subjects), missing/inappropriate control groups, and problems related to administration/scoring (also see Exner, Kinder, & Curtis, 1995; Gacono, Evans, & Viglione, 2008; Ritzler & Exner, 1995). Dies also opined that a theoretical model needs to proceed a study and it guides analyses (also Weiner, 1995b). Likewise, he identified certain Rorschach studies that had methodological bias, some of which were included in recent meta-analyses (e.g., Ball, Archer, Gordon, & French, 1991). Additionally, Dies (also see Acklin & McDowell, 1995; Mcguire, Kinder, Curtiss, & Viglione, 1995) found that researchers did not provide basic demographic data. This is necessary for interpreting findings, especially when a study is focused on validating the corollaries of a Rorschach variable, including sample parameters such as Lambda, Responses, mean, standard deviations, and frequencies for the variables studied (also see Gacono et al., 2001; Gacono & Gacono, 2008).

Viglione and Exner (1995) cautioned against inappropriate control groups (i.e., using an Exner normative sample as the comparison group; also see Shaffer, Erdberg, & Haroian, 1999 for other normative data) and stressed the importance of critically evaluating previous Rorschach literature. Failure to critique the literature can and has allowed studies with inaccurate and ambiguous results to infiltrate published studies creating deceptive impressions (also see Cunliffe et al., 2012). Proper statistical methods were also discussed, specifically the importance of accurately applying parametric versus nonparametric statistical procedures (many Rorschach variables form J-shaped curves that are not conducive to analysis with

parametric procedures; Viglione, 1995). Further, Viglione (1995) identified the variables appropriate for parametric and nonparametric statistics. Weiner (1995b) discussed the differences of using the Rorschach as a dependent and independent measure (doing research *with* the Rorschach or *on* the Rorschach), to provide adequate interrater reliabilities (80% agreement or better for Rorschach variables) and that poorly designed research will contain Type I and II error (stating a relationship exists when it does not or failing to identify relationships that exist). Ritzler and Exner (1995) discussed the limitations of clinical research and its chance for confounding variables (i.e., R, EB, and Lambda). Mcguire and colleagues (1995) stated that Rorschach variables may need to be categorized (i.e., dichotomous variables). Finally, Zillmer and Vuz (1995) provided information to perform factor analyses with Rorschach data.

Overall, many of the authors of *Issues and Methods in Rorschach Research* stressed the importance of not using a small sample size, providing key variables means/standard deviations (R, Lambda) as these can confound other variables, and the need to critically evaluate Rorschach research that may contain problems with methodology.

Gacono and colleagues (2001), Gacono and Gacono (2008), and Cunliffe et al. (2012) have added to the necessary parameters for conducting Rorschach research in order for a study to be generalizable and to allow reviewers to accurately interpret findings (e.g., reporting mean, standard deviation, and frequencies for IQ, Lambda, number of Responses). They offered five conceptual and four methodological criteria for evaluating the Rorschach/psychopathy research (Gacono et al., 2001, p. 32; also see Cunliffe et al., 2012; Gacono & Gacono, 2008; Gacono et al., 2008). Only Methodological Criteria 2-4 (see below) are provided as they apply to Rorschach research.

Methodological Issues in the Assessment of Rorschach/Psychopathy Findings (Gacono et al., 2001, p. 32)

2. Studies need to account for (control or delineate) the limitations imposed by factors such as gender, sexual deviance, concurrent Axis I psychosis, age, IQ, testing setting, and legal status. These factors can influence the production of certain Rorschach variables.
3. R (number of responses) must be considered. Increased R is found in certain sex offender groups, (Bridges, Wilson, & Gacono, 1998; Gacono, Meloy, & Bridges, 2000), whereas low R is typical among many criminal groups (Viglione, 1999). Thus, R can act as a moderator influencing the relationship

between Rorschach variables and criterion variables. Research should investigate this hypothesis by controlling for R and examining the relationship between Rorschach variables and criterion constructs at different levels of R (e.g., $R = 14-17$, etc.).

4. Response style must be considered (Bannatyne, Gacono, & Greene, 1999). Variables and styles such as R, Lambda, Extratensive, and Introversive can impact the production of certain Rorschach variables (Exner, 1995b), contributing to seemingly discrepant findings among studies.

While the wording of Methodological Criteria 2, 3, and 4 is geared toward Rorschach studies, similar issues (with modifications) are essential when conducting research with other psychological assessment instruments such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), Personality Assessment Inventory (PAI; Morey, 1991), or Millon Clinical Multiaxial Inventory - IV (MCMI-IV; Millon, Grossman, & Millon, 2015). These criteria (2, 3, and 4) will be expanded because they are important caveats for Rorschach research.

Potential confounds within the parameters of the instrument used should be assessed. Rorschach responses and protocol constrictions can be significantly affected by mental illness, IQ, legal status, age, and testing environment (Methodological Criterion 2; Exner, 1974, 2003; Weiner, 1966). IQ and/or Educational level affect the production of certain research variables and must be accounted for by researchers (Exner, 2003; Gacono et al., 2001; Weiner, 1966). Meyer, Gromini, Viglione, Reese, and Mihura (2015) found years of adult education was correlated with different Rorschach variables related to complexity and cognitive synthesis. Therefore, a lack of consideration of these factors may result in methodological bias. Demographic information must be provided in order for reviewers to understand the meaning of any Rorschach data offered within the study.

As stated in Methodological Criterion 3, the number of Rorschach responses (R) relates to the stability and reliability of Rorschach variables (Viglione & Meyer, 2008; Weiner, 2003); protocols with less than 14 responses ($R < 14$) should be interpreted with caution as they typically do not have enough data for adequate retest reliability (Weiner, 2003). Although it is acknowledged that $R < 14$ protocols were considered acceptable if accompanied by a Lambda below 1.2 (Exner, 1986) and these protocols may have clinical significance (Gacono, 1997), the current standard per the Exner Comprehensive System (CS; Exner, 1993, 2003) requires 14 or more responses for interpretation. Further, the problem of R (Exner, 1992; Meyer, 1992; Wood, Nezworski, & Stejskal, 1996) may or may not impact Rorschach results; however, significant differences between experimental groups on R could impact research findings (Weiner, 1995a). One can state with greater

certainty that within a Rorschach protocol of over 30 responses having a T or Fr + rF equal to zero, the absence of these variables was most likely due to either the absence of the trait or the personality functioning of the subject. In a protocol with less than 14 responses, additional hypotheses need to be considered. Similarly, in a protocol with 12 responses and four reflections it is highly likely that the presence of reflections are stating something about the personality of the person (Gacono & Meloy, 1994).

Methodological Criterion 4 includes participants with $\Lambda > .99$ and/or $F\% = .50$ (Meyer, Viglione, & Exner, 2001). These “response style variables” (Gacono & Gacono, 2008 - Λ and R) are essential to interpreting Rorschach findings and are the two main validity criteria of a CS protocol (Exner, 2003). Further, if attempting to validate the Rorschach with meta-analyses, valid protocols (R and Λ reported; $R \geq 14$) are needed. Frequently seemingly discrepant findings can be explained by a predominance of high Λ protocols in their samples (Exner provided different normative samples for Introversive/Extravertive patients for $L > 0.99$ & $L < 1.00$). Further, Konishi (2003) found differences in many Rorschach variables between high and low Λ groups. While the “constriction” of the protocols is frequently interpretable related to the sample characteristics, comparing Rorschach variables between other studies (more normatively distributed Λ s) is impossible (Gacono & Gacono, 2008). It would be difficult to know if the presence or absence of a CS variable is due to a high Λ or the absence of the trait in a high Λ sample. For example, if a study of male offenders had an overall mean $L > .99$ (high Λ) and did not produce any reflections, it would be difficult to determine whether the lack of these variables was due to the protocol’s constriction or the absence of the trait represented by these variables. Certainly, one cannot compare the high Λ sample with a sample of male offenders with a mean $L = .75$ (avg. Λ range) where these variables were present and then conclude that the negative findings in the high sample negate the positive findings in the more normally distributed sample. Assuming equality in the samples where one is a high Λ sample is not justified. Generalizability is limited and only within-sample interpretive conclusions are justified. This reinforces the rationale for including Λ means, standard deviations, etc. for all samples included in Rorschach research so the reviewer can determine whether the lack of results was due to this potential confound, which tends to create either a Type I (stating a relationship exists when it does not) or Type II error (failing to identify a relationship that does in fact exist).

As noted by Gacono and Gacono (2008) when discussing the validity versus generalizability of findings in their forensic outpatient groups:

Atypical patterns of constriction are common and frequently represent accurate portrayals of the patient's psychology (referring to high Lambda samples) ... For example, the higher rate of positive SCZI and PTI and Level 2 Special Scores among Schizophrenic groups...compared to our outpatient forensic groups does not indicate a lack of psychotic process among these patients. It also does not, necessarily, suggest resistance to the testing process. Rather, this finding must be interpreted in light of the patient's chronicity and their cognitive and emotional impoverishment. Rorschach constriction or expansion must always be explained within the context of the entire assessment battery and the patient's psychosocial history. (p. 443)

Researchers have also indicated the importance of determining and reporting interrater reliability for all Rorschach variables studied (Meyer, 1999; Viglione & Meyer, 2008; Weiner, 2003; Wood, Nezworski, & Stejskal, 1996). Therefore, researchers should report interrater reliability ($\geq .60$ [kappa and/or ICC] and 80% agreement is characterized as good to excellent; Meyer, 1999; Weiner, 1995a, 2003). Failure to report good/excellent interrater reliability is problematic as without reliability one is not assured of finding valid studies (Borsboom, Mellenbergh, & van Heerden, 2004).

Further, though there is no standard that is considered an adequate sample size; it appears that less than 20 is not acceptable using effect size, power, and alpha tables for nonparametric tests (Cohen, 1992; Dies, 1995a; Exner, 1995b). In neuroscience research, having small samples negatively impacts reliability and has less power (Button et al., 2013), and these points are also relevant for Rorschach research. Therefore, when a study has low statistical power there is less chance of detecting a true effect and "low power also reduces the likelihood that a statistically significant result reflects a true effect" (p. 365; Type II error). Low-sample research can also be viewed as unethical, inefficient, and wasteful (Button et al., 2013).

In summary, many different criteria need to be addressed when completing Rorschach research based on previous research from prominent Rorschach experts. However, five were selected that were mentioned most by many researchers and that appear to be essential for the validity and generalizability of research findings when conducting Rorschach research. These five are: (1) IQ/Educational level; (2) Rorschach Responses; (3) Lambda/F%; (4) interrater reliability; and (5) sample size.

These criteria are paramount for Rorschach research as there has been an increase in using meta-analyses with the Rorschach. Meta-analyses may overcome some of these problems such as small sample size, low power, and small effect sizes, unless the individual studies used in the meta-analyses contain problems related to these five criteria.

Meta-Analysis

Meta-analytic procedures are frequently utilized in assessing the validity of psychological measures (Sánchez-Meca & Marín-Martínez, 2010) including the Rorschach (Mihura et al., 2013; Wood et al., 2010). However, meta-analyses are only as good as the individual studies that they include (Button et al., 2013; Cunliffe et al., 2012; Hunter & Schmidt, 2004). For example, recent meta-analyses by Wood et al. (2010) examining the Rorschach and psychopathy were found to contain methodological bias in the research studies suggesting that there were not enough appropriate studies to perform the meta-analyses (see Cunliffe et al., 2012). Out of the 22 studies utilized in the meta-analyses, only one study was valid for inclusion by Rorschach/psychopathy standards (Cunliffe et al., 2012; Gacono et al., 2001; Gacono et al., 2008).

Recent meta-analyses by Mihura et al. (2013) questioned the validity of several Rorschach Comprehensive System (RCS; Exner, 2003) variables (e.g., Egocentricity Index and Isolation Index). Mihura et al. found mean validity coefficients of $r = .27$ when using the Rorschach variables against externally assessed criteria (e.g., psychiatric diagnosis) and $r = .08$ for introspectively assessed criteria (e.g., self-report measures). Further, effect sizes were calculated for the different Rorschach individual variables and they used Hemphill (2003) criteria to interpret the effect sizes. In total, 13 variables had excellent support (e.g., An + Xy; Perceptual Thinking Index; $r \geq .33$), 17 had good support (e.g., Lambda, Affective Ratio; $r \geq .21$), 10 had modest support (e.g., Vista, PHR; $r = .15-.21$), 13 had little or no support (e.g., Pure Color, Food; $r < .15$), and 12 variables (e.g., Aspiration Ratio, Color Projection) did not have any validity studies.

Wood, Garb, Nezworski, Lilienfeld, and Duke (2015) critiqued the Mihura et al. (2013) meta-analyses and indicated that the research by Mihura et al. was biased, since they failed to include certain articles/unpublished dissertations. Mihura, Meyer, Bombel, and Dumitrascu (2015) responded to these criticisms and refuted the claims of publication bias. Tibon Czopp and Zeligman (2016) also critiqued the Mihura et al. (2013) meta-analyses. Tibon Czopp and Zeligman focused their critique on the 13 CS variables that Mihura et al. stated had little to no support (e.g., Pure Color, Food; $r < .15$). They argued that these variables should not be removed from the CS system. They suggested there were discrepancies with the way Mihura and colleagues defined the 13 CS variables in the meta-analyses, which was not comparable to the customary CS interpretation for these variables. They argued the 13 individual variables needed to be interpreted within the CS clusters (e.g., self-perception, affective, etc.) due to the Rorschach being a multidimensional method. Additionally, in order to validate the 13 CS variables, externally assessed criteria (i.e., observer ratings, diagnosis) are better than introspectively

assessed criteria (i.e., self-report measures). Tibon Czopp and Zeligman also postulated that there may have been studies on these 13 CS variables. These studies that were not included in the meta-analyses may have supported the CS variables. Therefore, findings that actually existed were not revealed in the Mihura et al. meta-analyses (Type II error). Mihura, Meyer, Bombel, and Dumitrascu (2016) responded, saying the statements used by Tibon Czopp and Zeligman were biased, their arguments would have lowered validity coefficients, and they needed to perform their own meta-analyses for these 13 variables.

There are different ways to critically evaluate studies within a meta-analysis. One way to evaluate validity (Borsboom et al., 2004), especially relevant to Rorschach studies, is to determine whether the articles used were studies *with the Rorschach* (application studies) or studies *on the Rorschach* (validation studies; Weiner, 1995b). Another way to conceptualize this would be using the Rorschach as the dependent variable (application) or the independent variable (validation). Typically, in application studies (studies with the Rorschach), Rorschach variables are assumed to be valid measures of a specific psychological construct and they can be replaced by another instrument measuring the same construct. A validation study would be attempting to determine the Rorschach variable meaning and it does not assume it is already valid. An example of an application study would be a researcher studying cognitive remediation effects on thought disorder in patients with schizophrenia. This researcher might choose to assess thought disorder with the Rorschach CS WSum6 variable. However, the WSum6 score could be replaced by the Thought and Language Disorder Scale (Kircher et al., 2014) without modification to the experimental design (thought disorder could be operationalized differently). In this example, the researcher assumes that the WSum6 score is already a valid measure of thought disorder. Although this example would seem to support concurrent validity of both psychological measures, the correlation may not be a true estimate of validity (Borsboom et al., 2004). Therefore, construct validity may be more important as without constructs and theory, psychology would not be classified as a science (Dies, 1995a). By nature, application studies (*with* the Rorschach) are not designed to validate specific Rorschach variables, rather they are used to study a clinical phenomenon. It would be problematic to use application studies (*with* the Rorschach) – which were not designed to *validate* Rorschach variables – in order to evaluate the *validity* of Rorschach scores. Ideally, meta-analyses on the validity of individual Rorschach variables should be based on validation studies (i.e., studies *on* the Rorschach; also see Borsboom et al., 2004).

Another method for assessing individual studies in a meta-analysis is to evaluate specific criteria relevant to the instrument or procedure studied (Cunliffe et al., 2012; Gacono et al., 2001). Hunter and Schmidt (2004) outlined the necessary

components of a meta-analytic study: author, date, sample size, standardized effect score, subject characteristics, diagnostic conditions (scope, duration, and severity), strength of study design, and individual study methodological concerns. Validity forms the basis of Hunter and Schmidt's (2004) comments concerning the importance of ensuring that reliable, valid, and methodologically sound studies are selected for inclusion in a meta-analysis. Further, although researchers appear to use mathematically sound methods (equations used as intended), the issue lies in the validity (i.e., meaning; Borsboom et al., 2004) of the application of these techniques to methodologically biased studies included in a meta-analysis. Therefore, it is important to analyze the Rorschach studies included in a specific meta-analysis and to evaluate them with specific criteria (e.g., sample size, interrater reliability).

Present Study

In this article, the studies from the Mihura et al. (2013) meta-analyses were analyzed examining the quality of the study in light of Exner (1995a) and others (Cunliffe et al., 2012; Gacono et al., 2001). The 210 studies (similar to Dies, 1995b) from the Mihura et al. (2013) meta-analyses were examined focusing on whether the articles were application or validation studies (studies *with* the Rorschach vs. studies *on* the Rorschach) as well as the five methodological Rorschach criteria related to the validity and generalizability discussed by many Rorschach researchers. The five criteria were: (1) IQ/Education level; (2) Responses; (3) Lambda/F%; (4) interrater reliability; and (5) sample size.

Method

All articles used in the Mihura et al. (2013) meta-analyses were obtained and reviewed. Then, all 210 articles were examined for the following five main areas:

1. IQ/Education level,
2. Responses,
3. Lambda/F%,
4. Interrater reliability, and
5. Sample size.

Within each five criteria, different questions were examined. The issues were tallied up and a sum total for all 210 articles was obtained when the answer to the questions (see below) was *no*. Therefore, the examinations were:

1. IQ/Education level

- a. Did the article have statistics related to IQ/Educational level (*M* and range)?
- b. If either mean or range was provided, did the article include both statistics?
- c. If IQ range was reported, did all participants have an IQ > 80?

2. Responses (R)

- a. Did the article have statistics related to R (*M* and range)?
- b. If mean was provided, did the article include range?
- c. If range was provided, did the article include mean?
- d. If range was reported, did all protocols have $R \geq 14$?

3. Lambda/F%

- a. Did the article have statistics related to Lambda/F% (*M* and range)?
- b. If mean was provided, were the means for Lambda/F% < .99/.50%?
- d. If mean was provided, did the article report range?
- d. If the mean of Lambda/F% > .99/.50, did the article report IQ/Education Level? (L/F% > .99/.50 and IQ/Education level were examined in combination as this affects generalizability.)

4. Interrater reliability

- a. Did the article have interrater reliability statistics?
- b. If interrater reliability was reported, were there values for ICC/ $\kappa \geq .60$ or $\geq 80\%$ agreement?

5. Sample size

- a. Did the article have comparison groups with ≥ 20 participants?

An overall analysis of the 210 articles was conducted after the aforementioned analyses. This was to determine how many problems an article had with the five main criteria. For example, if an article did not report Lambda and interrater reliability statistics, it was calculated that the study had two methodological issues. Therefore, the five criteria were tallied up to determine how many problems an article had (i.e., the article contained zero, one, two, etc. problems).

After assessing issues in the 210 articles, the articles were then classified as either validation or application studies. See the previous section to understand how validation and application studies were operationalized. After identifying

the validation studies, and since they are preferred for meta-analyses, these articles were reviewed with the same five criteria as all 210 articles.

Due to the importance of validation studies in a meta-analysis, only these articles were examined to determine whether the methodological bias stated earlier had an impact on findings. The validation studies were reviewed to determine whether there were any counterintuitive findings. A counterintuitive study that we propose can be operationalized as a study that had findings that were inconsistent with theory or did not replicate previous findings (potential Type II error; where biased designed studies fail to identify relationships that in fact exist). Further, a study was classified as counterintuitive if there was mixed support for the Rorschach variables studied. For example, many research studies in the Mihura et al. meta-analyses examined multiple Rorschach variables (i.e., Hart, 1991). Hart did not find results consistent for the Egocentricity Index (EGOI) but support was found for the other variables examined; however, it was classified as counterintuitive owing to the inconsistent finding for EGOI. The importance of this distinction is, simply stated, if counterintuitive findings occur within a methodologically biased design there is no way to tell if the findings are true findings independent of confounds introduced by the methodology (Type II error; Weiner, 1995b). Specifically, these type of studies (including those with confounds that are not addressed) rarely identify relationships that do not exist (Type I error); rather, counterintuitive findings in Rorschach research often result in a failure to identify relationships that do in fact exist (Type II error). These counterintuitive studies within the validation studies were tallied up and an overall number was calculated.

Results

All 210 articles had some issues in the criteria assessed. When looking at the five criteria globally (IQ/Education level, R, Lambda/F%, interrater reliability, sample size), every article had methodological issues (see Table 1). Four articles only had one methodological issue (1.9%), 17 had two (8.1%), 85 had three (40.5%), 87 had four (41.4%), and 17 had five (8.1%). An example of an article with two problems was the one by Dao and Prevatt (2006). They provided statistics for R, provided an adequate interrater reliability, and had comparison groups greater than 20; however, there was no mention of IQ/Education level or Lambda/F%. An example of an article with five issues was that by Abraham, Mann, Lewis, Coontz, and Lehman (1990). They had three comparison groups of 15 participants, did not provide statistics for R, Lambda/F%, or interrater reliability, and did not provide a range for IQ. The specific questions for the five criteria were also calculated

Table 1. Methodological issues analysis ($N = 210$)

Methodological issues (n)	Articles (n)	Percentage
0	0	0%
1	4	1.9%
2	17	8.1%
3	85	40.5%
4	87	41.4%
5	17	8.1%

Table 2. Five methodological criteria analysis ($N = 210$)

Criterion	Articles (n)	Percentage
IQ/Education level		
No M and range	56	26.7%
Missing M or range	118	56.2%
IQ $M < 80$	20	9.5%
Responses (R)		
No M and range	108	51.4%
M reported but no range	65	31.0%
Range reported but no M	2	1.0%
Responses < 14	22	10.5%
Lambda/ $F\%$		
No M and range	141	67.1%
$M = L > .99/F\% > .50$	40	19.0%
No range for $L/F\%$	45	21.4%
$M = L > .99/F\% > .50$ & No IQ/Ed. level reported	28	13.3%
Interrater reliability		
None reported	72	34.2%
$< .60$ (κ/ICC) and/or 80% agreement	16	7.6%
Sample size		
< 20 participants	56	26.7%

Note. M = mean; L = Lambda; $F\%$ = percentage of F responses.

(see Table 2). The total calculations show if the answer was *no* to the question asked.

Two independent raters then used the aforementioned operationalizations about application and validation studies and analyzed each of the 210 studies. In total, 106 studies were classified as validation studies; however, there was a

Table 3. Methodological issues in validation studies only ($N = 104$)

Methodological issues (n)	Articles (n)	Percentage
0	0	0%
1	0	0%
2	12	11.5%
3	34	32.7%
4	50	48.1%
5	8	7.7%

disagreement on three articles. The interrater reliability was measured by kappa ($\kappa = .97$). A third rater resolved the disagreement, and 104 of the 210 studies (49.5%) were classified as validation studies.

Since validation studies are more appropriate for meta-analyses, the 104 studies were analyzed with the same criteria used for all 210 articles (see Table 3 and Table 4). After these analyses, the counterintuitive finding operationalization was used with the 104 validation studies. It was found that 45 studies (43.3%) had counterintuitive findings.

Discussion

Exner and colleagues (1995a; Cunliffe et al., 2012; Gacono et al., 2001; Gacono et al., 2008) provided pertinent caveats for conducting Rorschach research. At this time, there are no set rules for conducting Rorschach research and few have critiqued the points provided by the authors of *Issues and Methods in Rorschach Research* and Gacono et al. (2001). Other researchers have proposed important concepts to consider when conducting Rorschach research (i.e., not using a normative sample as a comparison group); however, the criteria used in this article appear to be the most pertinent to validity and generalizability. Those who research the Rorschach must be aware of these complexities and pitfalls (Exner, 1995a). Failure to heed these cautions produced biased Rorschach studies that led to inaccurate claims about the test (many resulting from Type II error). Most of the researchers did not provide descriptive data for Rorschach samples (Vigilione & Exner, 1995). For example, 51.4% did not provide any statistics for R and 67.1% did not provide these data for $\Lambda/F\%$. Further, of the 210 articles, 90% had three or more issues, which indicated these articles manifested gaps in their methodology.

A little more than half the articles included in these meta-analyses were application studies (*with* the Rorschach; 50.5%) and this is problematic when

Table 4. Five methodological criteria analysis for validation studies only ($N = 104$)

Criterion	Articles (n)	Percentage
IQ/Education level		
No M and range	27	26.0%
Missing M or range	61	58.7%
IQ $M < 80$	9	8.7%
Responses (R)		
No M and range	58	55.8%
M reported but no range	26	25.0%
Range reported but no M	2	1.9%
Responses < 14	9	8.7%
Lambda/ $F\%$		
No M and Range	80	76.9%
$M = L > .99/F\% > .50$	16	15.4%
No range for $L/F\%$	14	13.5%
$M = L > .99/F\% > .50$ & No IQ/Ed. level reported	12	11.5%
Interrater reliability		
None reported	33	31.7%
$< .60$ (κ/ICC) and/or 80% agreement	7	6.7%
Sample size		
< 20 participants	25	24.0%

Note. M = mean; L = Lambda; $F\%$ = percentage of F responses.

evaluating the validity of Rorschach variables since this type of study is not meant to validate Rorschach variables. Even more concerning is the fact that these two very different types of research (studies *with* the Rorschach vs. studies *on* the Rorschach) were combined indiscriminately in the Mihura et al. (2013) meta-analyses. Indeed, in application studies, Rorschach variables are typically dependent variables of the research design (e.g., the effects of cognitive remediation on thought disorder in patients with schizophrenia as assessed by the CS WSum6). Treating application studies (*with* the Rorschach) as validation studies (*on* the Rorschach) is the equivalent of confusing dependent and independent variables of research designs.

Since validation studies are more pertinent for a meta-analysis looking to validate Rorschach scores, many of the Rorschach validation studies included in Mihura et al. (2013) meta-analyses did not meet the methodological criteria assessed (90% had three or more problems with the criteria). Failure to meet the five criteria assessed limits the validity and generalizability of the study. If there are problems in a study it may not be advisable to use it in a meta-analysis. Finding a relationship despite poor methodology is more impressive and comes

with more weight than failing to find a relationship because of poor methodology. This leaves open the possibility that the relationship exists and can be found with better methods (overcoming Type II error) as poor methodological studies rarely yield Type I error (identifying relationships that do not exist). A more refined analysis of the validation studies was undertaken to determine whether the studies had counterintuitive findings (i.e., were inconsistent with theory and/or previous research). In all, 45 studies had counterintuitive findings (43.3%). Owing to the methodological issues it is unclear if these findings are due to Type II error, the findings are real, an artifact of the methodology, a quirk, or an atypical sample. Moreover, negative results could be attributable to Lambda or EB. That is, a variable could be valid for some purpose among $L < 1.00$ records; however, it could be invalid among $L > 0.99$ records, and similarly a variable could be valid for some purpose among introversive but not extratensive records (EB style).

For example, Simon (1989) did not find support of the Isolation Index (one of the variables Mihura et al. suggested had no support) compared with an MMPI scale. However, Simon did not report $L/F\%$ and $IQ/$ Education level and included protocols with $R < 14$ ($R = 10$). Without even considering the problems when using MMPI scales to validate Rorschach indices, the failure to take into account these criteria, it makes it impossible for the reader to determine whether the lack of correlation between the MMPI scale and the Rorschach index was nothing more than the result of an atypical sample. Hart (1991) looked at the Egocentricity Index (EGOI; another variable identified by Mihura et al. as having poor support) and the results did not support previous findings. However, Hart did not have IQ scores for all the participants, some had $IQ < 80$, and $L/F\%$ and interrater reliability were not reported. Another study examining the EGOI (Brems & Johnson, 1990) also did not find support for the index. However, the participants (inpatient psychiatric patients) had a high Lambda ($M = 1.20$) and $IQ/$ Education level and R statistics were not provided. Ball and colleagues (1991; a study found to contain bias [Dies, 1995a]) found inconsistent findings with the DEPI; however, percent agreement in terms of some Rorschach variables was 60%, much less than the acceptable 80%. Again, these issues from all these studies makes it difficult to generalize the findings.

Although the counterintuitive findings pertain to Type II error, there are problems with Rorschach studies that can lead to Type I error. This would include using a shotgun study (where many statistical analyses are performed), normative samples being used as a comparison group (i.e., Exner, 1995b; Shaffer et al., 1999), or inappropriate statistical comparisons being used (i.e., using parametric statistics like ANOVA rather than a chi-square analysis). Exner (1995b) cautioned that using a normative sample would result in some statistically significant findings by chance and should not be used. This would also be found when using too many

comparisons (shotgun studies). Viglione (1995) provided a table of which CS variables are appropriate for parametric analyses (i.e., *t* test) and which are appropriate for non-parametric (i.e., chi-square) analyses. For example, any content variable is more appropriate for chi-square analyses, while X-% is generally appropriate for parametric analyses. It is often more appropriate with variables such as texture to evaluate this in terms of $T = 0$, $T = 1$, and $T > 1$. This is the case with many Rorschach variables and conducting the correct analysis is essential when examining studies to be included within meta-analyses. Examining the 210 articles in the Mihura et al. meta-analyses, 17 (8.1%) studies met criteria for a shotgun study, 12 (5.7%) had compared their sample with an Exner normative sample, and 70 (33.3%) used inappropriate statistical comparisons (i.e., using parametric statistics like ANOVA rather than a chi-square analysis). Therefore, many studies included in the Mihura et al. (2013) meta-analyses may be problematic because of both Type I and Type II error.

In 1995, Exner stated: "A huge number of published investigations ... are clearly marked by errors in design, implementation, and/or analysis" (Exner, 1995b, p. 3). Further, Viglione and Exner (1995) stated: "A substantial proportion [of Rorschach research] have been marked by flaws" and "all literature cannot be afforded equal weight" (p. 55). Though there have been improvements, this continues to be the case as identified in the analysis. Many of the caveats provided by researchers have not been followed (Dies, 1995a, 1995b; Exner, 1995; Gacono et al., 2001; Ritzler & Exner, 1995; Viglione, 1995; Viglione & Exner, 1995; Weiner, 1995b).

Implications

These results can factor into future Rorschach meta-analytic research. Using more appropriate and methodologically sound studies, the effect sizes may be larger and more power can be given to the findings (overcoming Type I and II errors). Currently, two systems are being used to administer, score, and interpret the Rorschach. The Exner CS and the Rorschach Performance Assessment System (R-PAS; Meyer, Viglione, Mihura, Erard, & Erdberg, 2011). The CS and the R-PAS are two entirely different systems; however, both use the same research foundation. Specifically, "R-PAS variable selection draws heavily on the meta-analyses by Mihura et al. (2013)" (Erard, Meyer, & Viglione, 2014, p. 172) and the "Mihura et al. (2013) results form the foundation for the statements made throughout Chapter 15 [of the R-PAS manual] on Variable Selection and Validity" (p. 172). Therefore, the results would also apply to the R-PAS.

However, the results of this paper should not dampen the spirits of those who use the CS and/or R-PAS or inspire those who have criticized the Rorschach (i.e., Wood and colleagues). It is imperative that researchers who use the

Rorschach produce methodologically sound studies so that any subsequent meta-analytic studies will have more power. Mihura et al. (2013) in a single article took on the challenge to address the validity of each Exner variable. Although critics (Tibon Czopp & Zeligman, 2016; Wood et al., 2015) and to a certain extent this article have commented on it, the Mihura et al. meta-analyses began to address the validity of each Rorschach variable.

Limitations and Future Directions

The five criteria used in the main analyses were gathered from different published Rorschach research. However, there might be other criteria researchers may find important to include in this type of analysis. The five criteria included were expressed by multiple authors; therefore, there appears to be consensus about their importance related to generalizability and validity. Additionally, the concepts expressed in this article may be novel to the reader (*application vs. validation studies*; counterintuitive findings). Nevertheless, these concepts are important when examining individual studies included in a meta-analysis. Another limitation is that this analysis only focused on articles referenced in the Mihura et al. (2013) meta-analyses. Therefore, other Rorschach articles not cited may have contained less methodological bias. The Mihura et al. (2013) meta-analytic articles were used because they form the current Rorschach validity research base. Additionally, all articles cited in Mihura et al. were published before 2012. Therefore, current articles may be more methodologically sound. Future studies may examine articles from 2012 to the present with the same analyses (five criteria).

Conclusion

The following can be gleaned from this analysis:

1. Although it was published over 20 years ago, the Exner (1995a) edited book *Issues and Methods in Rorschach Research* is essential reading for any researcher using the Rorschach. The information provided should be followed, which will result in a methodologically sound study (also see Cunliffe et al., 2012; Gacono et al., 2001; Gacono & Gacono, 2008). Improved research designed studies will enhance the validity of the Rorschach.
2. The articles cited in the Mihura et al. (2013) meta-analyses suffered from methodological problems. Therefore, when referring to the meta-analytic

findings, some caution is advised. Further, Mihura et al. (2016) continued a trend to respond to criticism of the meta-analyses from a statistical perspective, rather than acknowledging and addressing issues raised by including poorly designed studies in the meta-analyses.

3. When researchers use the Rorschach in future studies, it is imperative that they understand what influences Rorschach research (i.e., IQ/Educational level), increase interrater reliability, report sample parameters (M , SD , and frequency for Lambda, R and any variables studied) and include appropriate sample sizes. Without this information the consumer cannot determine whether the statistical procedures in the studies were appropriate and whether the findings can be generalized beyond the sample studied (Gacono & Gacono, 2008; Gacono et al., 2008).
4. There is value in both application and validation studies; however, more validation studies are needed to perform appropriate meta-analyses.
5. It should be noted that some of these methodological issues are not unique to the Rorschach and they appear to be relevant in other psychological measure research (i.e., PAI, MMPI-2).
6. Only articles published before 2012 were reviewed owing to the focus on Mihura et al. (2013). Future research will investigate whether these issues are present in current research (2012 and beyond).

Acknowledgments

Thank you to Dr. Eric Lugo for the Spanish translation of the summary. Opinions expressed in this article are those of the author and do not necessarily represent the opinions of the Federal Bureau of Prisons or the Department of Justice.

References

- Abraham, P. P., Mann, T., Lewis, M. G., Coontz, B., & Lehman, M. (1990). The Cognitive Synthesis Test and the Rorschach indicators of internal and external reality in schizophrenic, borderline, and organic/psychotic adolescents. *British Journal of Projective Psychology*, 34, 34–44.
- Acklin, M. W., & McDowell, C. I. (1995). Statistical power in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 181–193). Hillsdale, NJ: Erlbaum.
- Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach Depression indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment*, 57, 465–476. https://doi.org/10.1207/s15327752jpa5703_6

- Bannatyne, L., Gacono, C., & Greene, R. (1999). Differential patterns of responding among three groups of chronic, psychotic and forensic outpatients. *Journal of Clinical Psychology, 55*(12), 1553–1565.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071.
- Brems, C., & Johnson, M. E. (1990). Further exploration of the Egocentricity Index in an inpatient psychiatric population. *Journal of Clinical Psychology, 46*, 675–679. [https://doi.org/10.1002/1097-4679\(199009\)46:5<675::AID-JCLP2270460521.3.O.CO;2-6](https://doi.org/10.1002/1097-4679(199009)46:5<675::AID-JCLP2270460521.3.O.CO;2-6)
- Bridges, M., Wilson, J., & Gacono, C. B. (1998). A Rorschach investigation of defensiveness, self-perception, interpersonal relations, and affective states in incarcerated pedophiles. *Journal of Personality Assessment, 70*, 365–385.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory–2 (MMPI–2): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Cunliffe, T. B., Gacono, C. B., Meloy, J. R., Smith, J. M., Taylor, E. E., & Landry, D. (2012). Psychopathy and the Rorschach: A response to Wood et al. (2010). *Archives of Assessment Psychology, 2*(1), 1–31.
- Dao, T. K., & Prevatt, F. (2006). A psychometric evaluation of the Rorschach Comprehensive System's Perceptual Thinking Index. *Journal of Personality Assessment, 86*, 180–189. https://doi.org/10.1207/s15327752jpa8602_07
- Dies, R. (1995a). Conceptual issues in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 25–51). Hillsdale, NJ: Erlbaum.
- Dies, R. (1995b). Subject variables in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 99–121). Hillsdale, NJ: Erlbaum.
- Erard, R. E., Meyer, G. J., & Viglione, D. J. (2014). Setting the record straight: Comment on Gurley, Piechowski, Sheehan, and Gray (2014) on the admissibility of the Rorschach Performance Assessment System (R-PAS) in court. *Psychological Injury and Law, 7*, 165–177. <https://doi.org/10.1007/s12207-014-9195-x>
- Exner, J. E. (1974). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (1st ed.). New York, NY: Wiley.
- Exner, J. E. (1986). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (2nd ed.). New York, NY: Wiley.
- Exner, J. E. (1992). R in Rorschach research: A ghost revisited. *Journal of Personality Assessment, 58*, 245–251.
- Exner, J. E. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic Foundations* (3rd ed.). New York, NY: Wiley.
- Exner, J. E. (Ed.). (1995a). *Issues and methods in Rorschach research*. Mahwah, NJ: Erlbaum.
- Exner, J. E. (1995b). Introduction. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 1–24). Mahwah, NJ: Erlbaum.
- Exner, J. E. (2003). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations and principles of interpretation* (4th ed.). Hoboken, NJ: Wiley.
- Exner, J. E., Kinder, B. N., & Curtis, G. (1995). Reviewing basic design features. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 145–158). Mahwah, NJ: Erlbaum.

- Gacono, C. B. (1997). Is the Rorschach Aggressive Movement Response enough? *British Journal of Projective Psychology*, 42(2), 5–11.
- Gacono, C. B., Evans, F. B., & Viglione, D. J. (2008). Essential issues in the forensic use of the Rorschach. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L. A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (pp. 3–21). New York, NY: Routledge.
- Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Erlbaum.
- Gacono, C. B., Meloy, J. R., & Bridges, M. R. (2000). A Rorschach comparison of psychopaths, sexual homicide perpetrators, and nonviolent pedophiles: Where angels fear to tread. *Journal of Clinical Psychology*, 56, 757–777. [https://doi.org/10.1002/\(SICI\)1097-4679\(200006\)56:6<757::AID-JCLP6>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-4679(200006)56:6<757::AID-JCLP6>3.0.CO;2-I)
- Gacono, L. A., & Gacono, C. B. (2008). Some considerations for the Rorschach assessment of forensic psychiatric outpatients. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L. A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (pp. 421–444). New York, NY: Routledge/Taylor & Francis Group.
- Gacono, C. B., Loving, J. L., & Bodholdt, R. H. (2001). The Rorschach and psychopathy: Toward a more accurate understanding of the research findings. *Journal of Personality Assessment*, 77, 16–38.
- Hart, L. R. (1991). The Egocentricity Index as a measure of self-esteem and egocentric personality style for inpatient adolescents. *Perceptual and Motor Skills*, 73, 907–914. <https://doi.org/10.2466/PMS.73.7.907-914>
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58, 78–79. <https://doi.org/10.1037/0003-066X.58.1.78>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Kircher, T., Krug, A., Stratmann, M., Ghazi, S., Schales, C., Frauenheim, M., & Grosvald, M. (2014). A rating scale for the assessment of objective and subjective formal Thought and Language Disorder (TALD). *Schizophrenia research*, 160(1), 216–221. <https://doi.org/10.1016/j.schres.2014.10.024>
- Konishi, H. (2003). The lambda in the Rorschach comprehensive system [Original in Japanese]. *Shinrigaku Kenkyu*, 73(6), 502–505.
- Mcguire, H., Kinder, B. N., Curtiss, G., & Viglione, D. J. (1995). Some special issues in data analysis. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 227–250). Hillsdale, NJ: Erlbaum.
- Meyer, G. J. (1992). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment*, 58, 231–244. https://doi.org/10.1207/s15327752jpa5802_2
- Meyer, G. J. (1999). Simple procedures to estimate chance agreement and kappa for the interrater reliability of response segments using the Rorschach Comprehensive System. *Journal of Personality Assessment*, 72, 230–255.
- Meyer, G. J., Giromini, L., Viglione, D. J., Reese, J. B., & Mihura, J. L. (2015). The association of gender, ethnicity, age, and education with Rorschach scores. *Assessment*, 22, 46–64. <https://doi.org/10.1177/1073191114544358>
- Meyer, G. J., Viglione, D. J., & Exner, J. E. (2001). Superiority of Form% over Lambda for research on the Rorschach Comprehensive System. *Journal of Personality Assessment*, 76, 68–75. https://doi.org/10.1207/S15327752JPA7601_4
- 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., 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. <https://doi.org/10.1037/a0029406>
- Mihura, J. L., Meyer, G. J., Bombel, G., & Dumitrascu, N. (2015). Standards, accuracy, and questions of bias in Rorschach meta-analyses: Reply to Wood, Garb, Nezworski, Lilienfeld, and Duke (2015). *Psychological Bulletin*, *141*, 250–260. <https://doi.org/10.1037/a0038445>
- Mihura, J. L., Meyer, G. J., Bombel, G., & Dumitrascu, N. (2016). On conducting construct validity meta-analyses for the Rorschach: A reply to Tibon Czopp and Zeligman (2016). *Journal of Personality Assessment*, *98*(4), 343–350. <https://doi.org/10.1080/00223891.2016.1158182>
- Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV manual*. Minneapolis, MN: Pearson.
- Morey, L. C. (1991). *Personality Assessment Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Ritzler, B. A., & Exner, J. E. (1995). Special issues in subject selection and design. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 123–143). Hillsdale, NJ: Erlbaum.
- Sánchez-Meca, J., & Marín-Martínez, F. (2010). Meta-analysis in psychological research. *International Journal of Psychological Research*, *3*(1), 151–163.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, *73*(2), 305–316.
- Simon, M. J. (1989). Comparison of the Rorschach Comprehensive System's Isolation Index and MMPI Social Introversion score. *Psychological Reports*, *65*, 499–502. <https://doi.org/10.2466/pr0.1989.65.2.499>
- Tibon Czopp, S., & Zeligman, R. (2016). The Rorschach Comprehensive System (CS) psychometric validity of individual variables. *Journal of Personality Assessment*, *98*(4), 335–342. <https://doi.org/10.1080/00223891.2015.1131162>
- Viglione, D. J. (1995). Basic considerations regarding data analysis. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 195–226). Hillsdale, NJ: Erlbaum.
- Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment*, *11*(3), 251–265.
- Viglione, D. J., & Exner, J. E. (1995). Formulating issues in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 53–71). Hillsdale, NJ: Erlbaum.
- Viglione, D. J., & Meyer, G. J. (2008). An overview of 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. 21–54). New York, NY: Routledge.
- Weiner, I. B. (1966). *Psychodiagnosics in schizophrenia*. New York, NY: Wiley.
- Weiner, I. B. (1995a). Methodological considerations in Rorschach research. *Psychological Assessment*, *7*(3), 330–337.
- Weiner, I. B. (1995b). Variable selection in Rorschach research. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 73–97). Hillsdale, NJ: Erlbaum.
- Weiner, I. B. (2003). *Principles of Rorschach interpretation* (2nd ed.). Mahwah, NJ: Erlbaum.
- Wood, J. M., Garb, H. N., Nezworski, M. T., Lilienfeld, S. O., & Duke, M. C. (2015). A second look at the validity of widely used Rorschach indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013). *Psychological Bulletin*, *141*, 236–249. <https://doi.org/10.1037/a0036005>

- Wood, J. M., Lilienfeld, S. O., Nezworski, M. T., Garb, H. N., Holloway Allen, K., & Wildermuth, J. L. (2010). Validity of Rorschach inkblot scores for discriminating psychopaths from nonpsychopaths in forensic populations: A meta-analysis. *Psychological Assessment, 22*(2), 336–349. <https://doi.org/10.1037/a0018998>
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*, 3–10. <https://doi.org/10.1111/j.1467-9280.1996.tb00658.x>
- Zillmer, E. A., & Vuz, J. K. (1995). Factor analysis with Rorschach data. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 251–306). Hillsdale, NJ: Erlbaum.

Received December 3, 2016

Revision received October 22, 2017

Accepted February 5, 2018

Published online November 9, 2018

Jason M. Smith

FCC Hazelton

Bruceston Mills, WV

USA

jmsmithpsyd@gmail.com

Summary

Exner's *Issues and Methods in Rorschach Research* (1995a) provided a standard of care for conducting Rorschach research; however, the extent to which studies have followed these guidelines has not been examined. Similarly, meta-analytic approaches have been used to comment on the validity of Comprehensive System (CS) variables without an evaluation as to the extent the individual studies have conformed to proposed methodological criteria (Exner, 1995a; Gacono, Loving, & Bodholdt, 2001). In this article, the 210 studies cited in recent Mihura, Meyer, Dumitrascu, and Bombel (2013) meta-analyses were examined. Individual studies were analyzed for: research *on the Rorschach* versus research *with the Rorschach* and whether they met the threshold of validity/generalizability related to specific Rorschach criteria. These criteria were: (1) IQ/Education Level; (2) Responses; (3) Lambda; (4) interrater reliability; and (5) sample size. Out of 210, 104 (49.5%) studies focused on research *on the Rorschach* and none met all five Rorschach criteria assessed. Further, 90% of the studies examined had three or more issues related to the above criteria. Therefore, the Exner (1995a) edited book *Issues and Methods in Rorschach Research* is essential reading for any researcher using the Rorschach. When researchers use the Rorschach in future studies, it is imperative that they understand what influences Rorschach research (i.e., IQ/Education level), increase interrater reliability, use appropriate statistics, report sample parameters (*M*, *SD*, and frequency for Lambda/R and any variables studied), and include appropriate sample sizes. Without this information the consumer cannot determine whether the statistical procedures in the studies were appropriate and whether the findings can be generalized beyond the sample studied. Further, there is value in both application and validation studies; however, more validation studies are needed to perform appropriate meta-analyses. The results of this paper should not dampen the spirits of those who use the CS and/or R-PAS or inspire those who have criticized the Rorschach (i.e., Wood and colleagues). It is imperative that researchers who use the Rorschach produce methodologically sound studies so that any subsequent meta-analytic studies will have more power.

Résumé

L'ouvrage *Issues and Methods in Rorschach Research* (Exner, 1995a) présente des recommandations concernant les standards de la recherche Rorschach; cependant, la mesure dans laquelle les recherches publiées ont suivi ces lignes directrices n'a pas été examinée. De même, des approches méta-analytiques ont permis d'évaluer la validité des variables du Système Intégré, mais sans prise en compte de la qualité méthodologique des études examinées au regard des standards proposés par Exner, 1995a; puis Gacono, Loving & Bodholdt (2001). Dans cet article, les 210 études citées dans les dernières méta-analyses de Mihura, Meyer, Dumitrascu et Bombel (2013) ont été examinées selon les critères suivants : recherche de validation (sur le Rorschach) vs recherche d'application (avec Rorschach) d'une part et d'autre part les critères suivants : (1) IQ / niveau d'éducation; (2) Nombre de Réponses; (3) Lambda; (4) fiabilité inter-évaluateur; et (5) Taille de l'échantillon. Sur 210, 104 (49,5%) des études étaient des études de validation à proprement parler et aucune ne remplissait l'ensemble des critères méthodologiques évalués. En outre, 90% des études examinées présentaient au moins trois problèmes méthodologiques. Par conséquent, le livre *Issues and Methods in Rorschach Research* (Exner, 1995a) est une lecture essentielle pour tout chercheur utilisant le Rorschach. À l'avenir, il est impératif que les chercheurs comprennent les facteurs susceptibles d'influencer la recherche sur le Rorschach : QI, fiabilité inter-évaluateur, utilisation des statistiques appropriées, présentation détaillée des paramètres de l'échantillon (M , SD et fréquence pour Lambda / R et toutes les variables étudiées) et tailles d'échantillon appropriées. Sans ces informations, le consommateur ne peut pas déterminer si les procédures statistiques employées dans une étude étaient appropriées et si les résultats peuvent être généralisés au-delà de l'échantillon étudié. En outre, si les études d'application et de validation présentent chacune leur intérêt; il semble nécessaire d'entreprendre de nouvelles études de validation pour effectuer des méta-analyses appropriées. Les résultats de cet article ne devraient pas décourager ceux qui utilisent le CS et / ou le R-PAS ou encourager ceux qui ont critiqué le Rorschach (c'est-à-dire Wood et ses collègues). Il est impératif que les chercheurs qui utilisent le Rorschach produisent des études solides sur le plan méthodologique ce qui permettra d'augmenter la puissance de méta-analyses ultérieures.

Resumen

Exner (1995a) *Issues and Methods in Rorschach Research* proporcionó un estándar de precaución para la realización de la investigación de Rorschach. Sin embargo, no se ha examinado hasta qué punto los estudios han seguido estas pautas. Del mismo modo, los enfoques meta-analíticos se han utilizado para comentar la validez de las variables de CS sin una evaluación que incluya los estudios individuales ajustado a los criterios metodológicos propuestos (Exner, 1995a; Gacono, Loving, & Bodholdt, 2001). En este artículo se examinaron los 210 estudios citados en los recientes meta-análisis de Mihura, Meyer, Dumitrascu y Bombel (2013). Se analizaron los estudios individuales para: investigación *sobre el Rorschach* versus investigación *con el Rorschach* y si cumplían el umbral de validez / generalización relacionado con los criterios específicos de Rorschach. Estos criterios fueron 1) CI / Nivel de educación; 2) Respuestas; 3) Lambda; 4) Confiabilidad entre evaluadores; y 5) Tamaño de muestra. De 210, 104 (49.5%) estudios se enfocaron en la investigación del Rorschach y ninguno cumplió con los cinco criterios de Rorschach evaluados. Además, el 90% de los estudios examinados tenían tres o más problemas relacionados con los criterios mencionados anteriores. Por lo tanto, el libro editado de Exner (1995a) *Issues and Methods in Rorschach Research* es una lectura esencial para cualquier investigador que use el Rorschach. Cuando los investigadores utilicen el Rorschach en estudios futuros, es imperativo que entiendan qué

influencia tiene la investigación de Rorschach (e.g. CI), aumento de la confiabilidad entre evaluadores, uso de estadísticas apropiadas, información de los parámetros de la muestra (M, DE, y frecuencia para Lambda / R y cualquier variable estudiada) e incluye tamaños de muestra apropiados. Sin esta información, el consumidor no puede determinar si los procedimientos estadísticos de los estudios fueron apropiados y si los resultados pueden generalizarse más allá de la muestra estudiada. Además, hay validez en los estudios de aplicación y validación; sin embargo, se necesitan más estudios de validez para realizar los meta-análisis apropiados. Los resultados de este documento no deben desalentar a los espíritus de quienes usan el CS y / o R-PAS o inspirar a aquellos que han criticado el Rorschach (e.g. Wood y sus colegas). Es imperativo que los investigadores que usan el Rorschach produzcan estudios metodológicamente sólidos para que cualquier estudio meta-analítico posterior tenga más poder.

要約

Exner (1995a) のロールシャッハ研究の問題と方法の再検討

Exner (1995a) の“ロールシャッハ研究の問題と方法”はロールシャッハ研究をおこなう際に配慮すべき基準を提供したが、どの程度の研究がこのガイドラインに従って実行されたのかについては検討されてきてはいない。同様に、ロールシャッハ包括システムの変数の妥当性についてコメントするのにメタ分析アプローチが用いられてきているが、個々の研究が提唱された方法論的な基準 (Exner, 1995; Gacono, Bodholdt, & Loving, 2001) に対してどの程度適合しているかについての評価はされていない。本論文では近年、Mihura, Meyer, Dumitrascu, and Bombel (2013)に引用された210の研究のメタ分析をおこなった。個々の研究は次のように分析された: ロールシャッハの研究 対 ロールシャッハを用いての研究、これらの研究は特定のロールシャッハの基準に関係する妥当性 / 一般性に合致しているかどうか。これらの基準は次のようなものである。(1) IQ / 教育レベル、(2) 反応数、(3) ラムダ、(4) 評定者間の信頼性、(5) サンプルサイズ。210のうち、104 (49.5%) はロールシャッハの研究に焦点を当てたものであり、これらの研究にロールシャッハの5つの基準をすべて満たしたものはなかった。さらに90%の研究がこれらの基準に関連する3つかそれ以上の問題を有していることが明らかになった。ゆえに、Exner (1995)が編集したロールシャッハ研究の問題と方法はロールシャッハを用いる研究者にとっての必須の書籍であると言える。研究者が将来の研究でロールシャッハをもちようとする場合、何がロールシャッハの研究に影響を与え(換言すればIQ)、評定者間の信頼性を高めるかを理解し、適切な統計手法を使い、標本のパラメータ (M、SD、ラムダ / Rの頻度や、他の分析されている変数) を示し、適切なサンプルサイズを含むことが必然的になる。これらの情報なしでは消費者は、この研究の統計的手続きが適切かどうか、見出されたことは研究のサンプルを超えて一般化できるかどうかを決めることはできない。さらに、応用研究や妥当性の研究は有用であるが、さらに妥当性の研究は適切なメタ分析を実行する必要がある。本研究の結果は包括システム / R-PASを使用するもの熱意を低下させ、ロールシャッハを批判するもの (Woodやその同僚) を活気づけるものではない。ロールシャッハを用いる研究者は方法論的に健全な研究を生み出すことが必然であり、必然的にメタ分析研究は有力になるであろう。