Incremental Validity of the Rorschach Prognostic Rating Scale Over the MMPI Ego Strength Scale and IQ

Gregory J. Meyer

Department of Psychology
University of Alaska Anchorage

A recent meta-analysis found that the Rorschach Prognostic Rating Scale (RPRS) had a strong ability to predict subsequent outcome ($r = .44$, $N = 783$; Meyer & Handler, 1997, this issue). However, that review did not directly address questions of incremental validity. This article focuses on the ability of the RPRS to predict outcome after taking into account other sources of data. Across studies that examined both the RPRS and the MMPI Ego Strength scale, the RPRS had a strong ability to predict outcome ($r = .40$, $N = 187$), whereas the MMPI scale did not ($r = .02$, $N = 280$). Nine studies examined the RPRS along with an intelligence test and allowed direct numerical estimates of incremental validity to be calculated. Across studies, the RPRS demonstrated strong incremental validity after controlling for intelligence (incremental $r = .36$, $N = 358$). It is clear that the Rorschach can make unique contributions to understanding clinically relevant processes in ways that self-reports or measured intelligence cannot. Contemporary Rorschach scales should continue to be evaluated for their distinctive and incremental contribution to clinical practice.

A recent meta-analysis (Meyer & Handler, 1997, this issue) examined the Rorschach Prognostic Rating Scale (RPRS; Klopfer, Kirtner, Wisham, & Baker, 1951), which is thought to be a measure of ego strength that reflects promise for treatment. Although not part of the Comprehensive System for Rorschach scoring (Exner, 1993), the RPRS is derived from scores that relate to movement, color, shading, form quality, and thought organization. The RPRS meta-analysis examined 20 statistics derived from a combined sample of 783 participants. With outcome criteria obtained an average of 1 year after initial testing, the uncorrected correlation between RPRS scores and outcome was found to be $r = .44$ (95% confidence interval = .39–.50). After artifact corrections were made to estimate the va-

lidity of the RPRS if all participants, all RPRS scores, and all outcome scores had been included in the final statistical analysis, *r* increased to .55.

Although these findings were substantial, the meta-analysis primarily focused on the univariate validity of the RPRS as a predictor of outcome. *Univariate validity* refers to the association between a single predictor and criterion. Recently, research attention has begun to focus on more complex questions concerning the incremental validity of Rorschach scores (see Archer & Krishnamurthy, 1997; Dawes, 1999; Hunsley & Bailey, 1999; Viglione, 1999). In its most basic definition, *incremental validity* refers to the capacity of one measure to improve prediction over one or more alternative measures (Sechrest, 1963; Wiggins, 1973/1988). For instance, if a newly revised scale of psychosis is able to improve the prediction of an appropriate criterion measure over the original psychosis scale, then the new scale demonstrates incremental validity over the original, and one can conclude that the new scale contributes meaningful information that could not have been obtained from the existing measure.

In a global fashion, research has indicated that the Rorschach is a valid instrument (Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999). What has yet to be fully clarified, however, is the extent to which the Rorschach provides valid and unique information that cannot be obtained from other sources. Two lines of reasoning suggest that the Rorschach should often provide incremental validity over other sources of test data. First, the available evidence indicates that the Rorschach produces univariate validity coefficients that are roughly equal to those obtained from self-report scales, although for some criteria the Rorschach appears to perform better, and for other criteria the reverse appears to be true (Atkinson, 1986; Bornstein, 1998, 1999; Hiller et al., 1999; Parker, Hanson, & Hunsley, 1988; for a discussion of Parker et al., 1988, see Garb, Florio, & Grove, 1998; Parker, Hunsley, & Hanson, 1999). Second, studies that have systematically explored the correspondence between Rorschach scores and self-reported characteristics have found little or no association between these sources of data under typical nomothetic analyses (e.g., Archer & Krishnamurthy, 1993; Meyer, 1997, 1999; Meyer, Riethmiller, Brooks, Benoit, & Handler, 2000). Given that both Rorschach scores and self-report scales are generally valid yet also generally uncorrelated with each other, information derived from the Rorschach should add to the information derived from a self-report instrument.

However, the conclusion that the Rorschach should add incremental validity to other sources of data is a logical deduction. The value of that conclusion ultimately depends on data documenting that it is true, not on the seeming accuracy of the logic. To further the evidence base concerning Rorschach incremental validity, I examined the ability of the RPRS to provide information about outcome beyond that which could be obtained from two other sources: (a) a self-report scale developed for this purpose and (b) measured intelligence.

THE RPRS AND THE MMPI EGO STRENGTH SCALE

Meyer and Handler (1997) conducted a secondary analysis of the Minnesota Multiphasic Personality Inventory (MMPI) Ego Strength (*Es*) scale (Barron, 1953). Like the RPRS, the *Es* scale was developed to predict response to psychotherapy. Unlike the RPRS, however, the *Es* scale is a self-report measure that does not require professional time for administration and scoring. This makes it a potentially less expensive and more attractive alternative to the RPRS. Meyer and Handler's meta-analysis of the *Es* across six studies revealed that the scale was unable to predict subsequent outcome (rs = .02 and .03 for uncorrected and corrected coefficients, respectively; $N = 280$).

Unfortunately, although these results suggest that the RPRS was a better predictor of outcome than the *Es* scale, Meyer and Handler (1997) did not conduct a head-to-head comparison. Instead, they examined an RPRS effect size that was calculated from 20 statistics but an *Es* effect size that was calculated from only 6 statistics.

samples, and both reported a lack of correlation between IQ and RPRS scores.

TABLE 1
Incremental Validity of the RPRS Over Measured Intelligence (IQ)

			r With Outcome			
Study	*IQ Measure*	*N*	*RPRS*	*IQ*	*RPRS Incremental r*	*Multiple R*
Bloom (1956)	Wechsler–Bellvue	25	.46	.40[a]	.37	.54
Filmer-Bennett (1955)[b]	"IQ"	22[c]	.36	.00	.36	.36

umn lists the number of participants in the analysis, the fourth column reports the RPRS–outcome correlation, and the fifth column indicates the IQ–outcome correlation. As can be seen in the table, five studies found univariate validity coefficients for the RPRS that were numerically larger than those for IQ scores (Bloom, 1956; Luborsky et al., 1979; Newmark et al., 1974; Newmark et al., 1973; Newmark, Konanc, Simpson, Boren, & Prillaman, 1979). Two other studies matched "improved" and "unimproved" patients on IQ prior to evaluating RPRS predictive validity (Filmer-Bennett, 1955; Novick, 1962). Matching groups on IQ forces the

good- and poor-outcome groups to be equated on intelligence, and it forces the IQ–outcome correlation to be zero (i.e., because outcome groups were equated on IQ, it is impossible for IQ to correlate with outcome). The final two studies found that IQ scores were more highly correlated with outcome than were RPRS scores in at least one of their analyses (Johnson, 1953; Mindess, 1957). Johnson's (1953) study contained some ambiguity, so I discuss it in more detail.

As indicated previously, Johnson (1953) presented two sets of RPRS raw scores for each participant in her study. One set of scores was obtained from Rorschach protocols administered at the start of treatment, and the other set was obtained from Rorschach protocols administered at the end of treatment. Johnson also presented raw scores for two IQ measures: the Terman–Merrill Binet Scale and Raven's Progressive Matrices. Johnson clearly stated that the Binet was administered at the start of treatment (p. 321; also see Johnson, 1952). In contrast, it appears that the Raven was administered at or near the end of treatment, although Johnson is never explicit about this point. Nonetheless, this conclusion seems warranted for two reasons. First, Johnson (1953) indicated that treatment lasted for an average of 15 weeks. Second, in a slightly expanded sample from the same setting (Johnson, 1952), she reported that on average there was an 11-week lag between the administration of the Binet and the administration of the Raven. Thus, it seems reasonable to conclude that the Raven was administered toward the end of treatment, likely at about the same time as the end-of-treatment Rorschach.

Because it is more difficult to predict future functioning than it is to describe current functioning (Hulin, Henry, & Noon, 1990), it would be unfair to compare the predictive validity of pretreatment RPRS and Binet scores with the concurrent validity of posttreatment Raven scores. Predictive validity coefficients for RPRS and Binet scores obtained before the start of treatment should be lower than the concurrent validity coefficient for Raven scores obtained at the end of treatment. Unfortunately, because Johnson (1953) was not explicit about when the Raven was administered, disregarding the Raven scores could be viewed as a decision that was biased in favor of the RPRS. Consequently, to be conservative, my analysis proceeded as if the Raven had been administered at the start of treatment even though it probably was administered at the same time that outcome was determined.

Using Johnson's raw data, I calculated outcome correlations for both the Binet and the Raven scales. Table 1 indicates that the RPRS had a stronger correlation with outcome (.53) than the Binet did (.23) but a weaker correlation than the Raven did (.61).[4] Although the Binet did not have a statistically significant correlation

---

[4]Johnson's results are quite consistent with what would be expected if the Raven had been administered at the end of treatment. The two scales that were definitely administered at the start of treatment—the initial RPRS and the Binet—had correlations with outcomes of .5349 and .2253, respectively. The RPRS scale that was definitely administered at the end of treatment had a larger correlation with outcome ($r = .614$). The correlation between the Raven scores and outcome was of about the same magnitude ($r = .607$) as that observed with the posttreatment RPRS scores.

with outcome, when both Binet and Raven scores were forced into a regression equation, the combined IQ tests had a better ability to predict outcome ($R = .70$). However, in contrast to any reasonable theoretical expectations, the Binet had a negative weight in this equation and thus indicated that outcome was best when Raven IQ scores were relatively high but Binet IQ scores were relatively low. Despite this paradoxical result, the combined IQ predictor scores were retained for Johnson's (1953) study.

Returning to the overall analysis, from a univariate perspective and considering all three IQ test possibilities for Johnson (1953), 8 of 11 head-to-head comparisons found that the RPRS had a better ability to predict outcome than IQ. The remaining 3 comparisons found that IQ scores had higher validity coefficients than the RPRS, although both IQ and RPRS scores predicted outcome.

Incremental Validity of the RPRS Over IQ

The next question to address is whether RPRS scores make an incremental contribution to the prediction of outcome after considering IQ. This question can be addressed through multiple-regression equations. With multiple regression there are several ways to determine whether a target variable makes an incremental contribution to prediction. First, *incremental* could be defined in terms of a "better than" standard. For instance, if RPRS scores produce a higher correlation with outcome than IQ scores, or if RPRS scores are significantly correlated with outcome but IQ scores are not, RPRS scores would be better predictors than IQ, and they could be said to have incremental validity over IQ. Alternatively, this definition also could be met when both IQ and RPRS scores have significant correlations with outcome, but RPRS scores have the higher correlation and IQ scores no longer contribute to prediction once RPRS scores are entered into a regression equation. A second way to define *incremental* is according to a "contribute to" standard. For instance, if both IQ and RPRS scores have positive univariate correlations with outcome and if both independently contribute to a regression equation, both would be considered useful predictors that in combination produce results that are even more valid. A final way to define *incremental* is in terms of an "over and above" standard. According to this standard, IQ would be forced into a regression equation first, regardless of whether it had a significant association with outcome. Once this was done, RPRS scores would be evaluated for their degree of association with outcome after controlling for IQ. In this model, IQ scores are given preemptive importance. The predictive contribution of RPRS scores is computed only after IQ is allowed to account for as much of the outcome variance as it possibly can and regardless of whether IQ

For the RPRS, this last standard is demanding, because it gives the least amount of credit to the RPRS scale and the most amount of credit to IQ.[5]

To evaluate the RPRS against the demanding over and above standard, in which all the variance shared by IQ and RPRS scores is allocated to IQ and none is allocated to the RPRS, it is necessary to have, or be able to generate, this specific incremental validity information in each study. None of the studies listed in Table 1 presented the necessary results in the original article. Thus, for each study, this information had to be generated. Because Johnson (1953) presented actual IQ and RPRS scores for each patient in her sample, the necessary information could be readily calculated.

In the absence of raw data, the unique contribution made by RPRS scores can be calculated according to Equation 3.3.8 from Cohen and Cohen (1983). In the context of these analyses, this formula is

$$\text{RPRS incremental } r = \frac{\left(r_{\text{Outcome–RPRS}} - \left(r_{\text{Outcome–IQ}} \times r_{\text{IQ–RPRS}}\right)\right)}{\left(1 - r_{\text{IQ–RPRS}}^{2}\right)^{\frac{1}{2}}} \tag{1}$$

where the subscripts indicate the two variables being correlated. From the formula, it can be seen that three correlations are needed: (a) the RPRS–outcome correlation, (b) the IQ–outcome correlation, and (c) the IQ–RPRS correlation. Only Mindess's (1957) study presented all three of these correlations. The remaining studies pre-

---

[5]A brief example can clarify this standard. Say a researcher is interested in determining whether a new self-report depression scale is an improvement over the original and has a sample of 50 patients, all of whom have gold standard criterion ratings of depression. The researcher finds that the old and new scales correlate at .85 with each other. In terms of validity, the new scale is found to have a correlation of .40 with the criterion, whereas the original scale has a correlation of just .33. According to a "better than" standard, the new scale would have incremental validity over the old because the new scale has a stronger correlation with the criterion and because it would enter a regression equation first if allowed to do so. Once the new scale was in the regression equation, the old scale would not contribute any additional information (formulas provided in Cohen and Cohen, 1983, reveal that the incremental r for the old scale actually would be –.019 in this design). Thus, according to a better than standard, the new scale is a clear improvement over the old and should in fact replace the old scale. However, according to the "over and above" standard, this study would produce results indicating just the opposite, such that one would conclude that the new scale does not have incremental validity over the old scale. When the old scale is entered into a regression equation first, it is given all the credit for the substantial variance (i.e., $r = .85$) it shares with the new scale. The new scale is given no credit for this shared variance and is evaluated instead only on the basis of the unique contribution it makes to prediction. The incremental contribution from the new scale would be $r = .227$. Although this value is positive and in the expected direction, it would not be statistically significant in a sample of this size. Thus, one would have to conclude that the new scale is not an improvement over the old scale when using the more demanding over and above standard.

sented information on the correlation between IQ and RPRS scores with outcome but did not provide data on the IQ–RPRS correlation.

For the analysis to proceed, it was necessary to obtain an optimal estimate of the correlation between RPRS and IQ scores. I accomplished this by using all of the existing literature addressing the topic. Specifically, I used the following studies and correlations: Edinger and Weiss's (1974) process schizophrenic sample, $r =$ .679, $N = 15$; Edinger and Weiss's college sample, $r = -.006$, $N = 15$; Hathaway (1982), $r = .66$, $N = 52$; Johnson (1953), $r = .1699$, $N = 21$; Mindess (1957), $r =$ .097, $N = 68$; and Williams et al. (1967), $r = .00$, $N = 42$. Across the six samples and 213 patients included in this analysis, the average sample weighted correlation between IQ and RPRS scores was $r = .2562$. Given that Hathaway had deliberately maximized variance in her sample, the parameter estimate of .2562 is likely to be somewhat inflated. As such, using this estimate is likely to produce values that are smaller than appropriate for estimating the RPRS's incremental contribution to outcome prediction (this can be verified by consulting the formula given above). Nonetheless, to establish a conservative estimate of the incremental validity of the RPRS over IQ, the correlation of .2562 was used in subsequent calculations.

The sixth column of Table 1 indicates the extent to which RPRS scores correlate with outcome after first forcing IQ to account for as much outcome variance as possible. Estimated results are presented for Bloom (1956), Luborsky et al. (1979), and Newmark and colleagues (Newmark et al., 1974; Newmark et al. 1973; Newmark et al., 1979). Actual results are presented for the remaining studies (Filmer-Bennett, 1955; Johnson, 1953; Mindess, 1957; Novick, 1962). The results for several studies deserve comment. Filmer-Bennett (1955) and Novick (1962) equated their improved and unimproved groups on IQ. As described before, matching on IQ forces the IQ–outcome correlation to be zero. It also forces the RPRS–IQ correlation to be zero. By necessity, then, in these two studies the observed RPRS–outcome correlations (.36 and .42, respectively) also documented the incremental contribution of RPRS scores to the prediction of outcome.

For Johnson's (1953) study, three incremental validity figures are presented

When Equation 3.3.8 from Cohen and Cohen (1983) was applied to the data from Luborsky et al. (1979) and Newmark et al. (1973), IQ was found to have a suppressive effect on RPRS scores.[6] As such, the incremental contribution of RPRS scores was slightly larger than that suggested by the univariate RPRS–outcome correlations. Specifically, for Luborsky et al. the incremental r was calculated to be .1736 even though the univariate RPRS–outcome correlation was .155. For Newmark et al.'s (1973) study the relevant values were .5716 and .5525, respectively. To be conservative, the suppressive effects of IQ were ignored, and the RPRS incremental r was limited so it would not exceed the univariate correlation between RPRS scores and outcome.

Overall, as Table 1 indicates, for every analysis in which one could formally document or estimate the incremental validity of RPRS scores over IQ, the RPRS made an incremental contribution to outcome in the theoretically expected direction. Despite the small size of many studies, in 8 of 11 instances the incremental contribution was statistically significant. It is important to note that the RPRS made an incremental contribution to prediction even when the univariate IQ–outcome correlation was substantially larger than the RPRS–outcome correlation. For instance, in Johnson's (1953) study the Binet and Raven combined had a correlation of .70 with outcome, whereas the RPRS had a correlation of .53 with outcome. Despite the large magnitude of the IQ–outcome association, RPRS scores still added statistically significant incremental information to the prediction task. For all studies, the final column of Table 1 indicates the multiple R found when both RPRS and IQ scores were used to predict outcome.

Meta-Analytic Summary of the Incremental
Contribution of RPRS Scores Over IQ

As a final step, I conducted a meta-analysis on the incremental effect sizes reported in the sixth column of Table 1. To be conservative, I used the lowest RPRS incremental effect size for Johnson's (1953) study (i.e., .3286). In the nine studies contributing to this meta-analysis, 358 participants were evaluated for outcome a median of 9 months after baseline Rorschach testing. The average sample and quality weighted effect size was $r = .36$. After correcting for methodological artifacts, the average weighted correlation was .48. Sampling error explained all of the study-to-study variation observed in the incremental correlations, indicating that

---

[6]A *suppressor variable* is one that contributes undesirable variance to a target predictor variable and thus causes the univariate correlation between the target predictor and the criterion to be lower than it could or should be. When both the suppressor and target predictor variables are entered into a regression equation, the regression equation removes this undesirable variance and allows the target predictor to show a larger correlation with the criterion. See Cohen and Cohen (1983, pp. 94–96) for a more complete discussion.

the summary effect sizes of $r = .36$ and $r = .48$ are stable estimates of the underlying population parameters. Thus, the RPRS has substantial incremental validity for predicting outcome over and above the information that could be gleaned from IQ test scores.

Recall that an estimate of the "true" IQ–RPRS correlation was used to generate some of the incremental correlations for the preceding meta-analysis. When studies did not report the correlation between IQ and RPRS scores, this correlation was estimated to be $r = .2562$. In general, a lower IQ–RPRS correlation allows the RPRS to account for more incremental variance in outcome. As indicated above, the value of $r = .2562$ is likely to be an overestimate of the population IQ–RPRS correlation. Nonetheless, some researchers may speculate that my IQ–RPRS correlation was too small and that it led to an overestimate of RPRS incremental validity. To address this potential concern, I recomputed the meta-analysis twice. In the first instance, I assumed the IQ–RPRS correlation was really .40 rather than .2562. In the second instance, I assumed the correlation was really .70 rather than .2562. These assumptions lead to unrealistically conservative estimates of RPRS incremental validity. In fact, by assuming that IQ and RPRS scores had a correlation of .70, I gave IQ scores more credit than had ever been observed in any empirical investigation. Nonetheless, I inserted these IQ–RPRS correlations into Equation 3.3.8 from Cohen and Cohen (1983) and recomputed the results.[7] When the IQ–RPRS correlation was set at $r = .40$, the RPRS incremental contribution to predicting outcome was $r = .35$ for the uncorrected meta-analytic results and $r = .48$ for the artifact-corrected results. When the IQ–RPRS correlation was set at $r = .70$, the RPRS incremental contributions to outcome were $r = .36$ and $r = .48$ for the uncorrected and artifact-corrected meta-analytic results, respectively. Note that the summary effect sizes increased slightly in the last analysis. This occurred because using such a large estimate for the IQ–RPRS correlation produced suppression effects for IQ in Newmark et al.'s (1979) study. Ultimately, however, both of these revised meta-analyses revealed that the primary results are quite stable. Even when the RPRS is treated in a highly unrealistically and unfavorable manner, RPRS scores still make a large and important contribution to the prediction of outcome over and above IQ.

Although the foregoing discussion documents the incremental contribution of RPRS scores to outcome prediction over IQ, Mindess's (1957) study is atypical. Unlike all the other studies, which predicted the outcome of psychological treatment, Mindess predicted the outcome of training in a nursing program. Because the latter is not directly relevant to clinical assessment, I recomputed the analysis

---

[7]Because I decided to ignore IQ suppressor effects and limit the incremental contribution of RPRS scores to the univariate RPRS–outcome correlation observed in Luborsky et al. (1979) and Newmark et al. (1973), the modified correlations were used for only three studies: Bloom (1956), Newmark et al. (1974), and Newmark et al. (1979).

to focus solely on the prediction of treatment outcome. When the Mindess study was dropped from the meta-analysis, the eight remaining studies contained data from 290 patients who had been evaluated for outcome a median of 9 months after the baseline Rorschach testing. The average weighted RPRS incremental effect size was $r = .36$; it was $r = .49$ after correcting for methodological artifacts. Thus, excluding Mindess had no observable impact on the summary results. The RPRS has substantial incremental validity for predicting the outcome of psychological treatment after taking into account information that could be gleaned from intelligence tests.

Archer, R. P., & Krishnamurthy, R. (1997). MMPI–A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69,* 517–533.

Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology, 27,* 238–247.

Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology, 17,* 327–333.

*Bloom, B. L. (1956). Prognostic significance of the underproductive Rorschach. *Journal of Projective Techniques, 20,* 366–371.

Bornstein, R. F. (1998). Interpersonal dependency and physical illness: A meta-analytic review of retrospective and prospective studies. *Journal of Research in Personality, 32,* 480–497.

Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment, 11,* 48–57.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment, 11,* 297–302.

Edinger, J. D., & Weiss, W. U. (1974). The relation between the altitude quotient and adjustment potential. *Journal of Clinical Psychology, 30,* 510–513.

*Endicott, N. A., & Endicott, J. (1964). Prediction of improvement in treated and untreated patients using the Rorschach Prognostic Rating Scale. *Journal of Consulting Psychology, 26,* 342–348.

Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.

*Filmer-Bennett, G. (1955). The Rorschach as a means of predicting treatment outcome. *Journal of Consulting and Clinical Psychology, 19,* 331–334.

*Fiske, D. W., Cartwright, D. S., & Kirtner, W. L. (1964). Are psychotherapeutic changes predictable? *Journal of Abnormal and Social Psychology, 69,* 413–426.

Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and Minnesota