JOURNAL OF PERSONALITY ASSESSMENT, *85*(2), 219–237

SOCIETY FOR PERSONALITY ASSESSMENT

The Status of the Rorschach in Clinical and Forensic Practice: An Official Statement by the Board of Trustees of the Society for Personality Assessment

This statement is intended for psychologists, other mental health professionals, educators, attorneys, judges, and administrators. Its purpose is to present a summary of the issues and evidence concerning the Rorschach. This statement affirms that the Rorschach possesses reliability and validity similar to that of other generally accepted personality assessment instruments, and its responsible use in personality assessment is appropriate and justified.

STATEMENT OF THE ISSUE

We are concerned that the Rorschach controversy of the past several years[1] has placed clinical and forensic psychologists in a conflicted position, where they have questioned whether they can continue to use the Rorschach in practice. Of even greater concern, some authors have called for a ban or moratorium on the use of the Rorschach and have recommended that psychology departments and organizations discontinue Rorschach training and practice.[2] As a positive development, the current controversy has led to an intense examination of the instrument, which has resulted in more systematic and well-designed research. Given the findings of psychometric adequacy and clinical utility that have emerged from these extensive investigations,[3] the Board of Trustees of the Society for Personality Assessment submits the following as our official statement on the status of the Rorschach in clinical and forensic practice. To support our position, we have assembled for the members of the Society of Personality Assessment and other interested psychologists and professionals the endnotes and tables in this statement covering the scientific status of the Rorschach.

HISTORY OF THE RECENT CONTROVERSY

The current controversy questioning the utility of the Rorschach extends back to 1995.[4] Since that time, it has been the topic of special sections in all three of the major journals devoted to the science and practice of psychological assessment.[5] Furthermore, multi-article sections have been published in several specialty journals[6] and a substantial number of stand-alone articles on the topic have attempted to address legitimate criticisms of the Rorschach,[7] while redressing those that are erroneous and misguided. In the process, the Rorschach has recently received a more intensive level of scrutiny than that given any other personality test of which we are aware.

SUMMARY OF SCIENTIFIC EVIDENCE

Ultimately, examination of the scientific evidence with this degree of rigor should allow an informed conclusion about the Rorschach's status in relation to other personality instruments and its appropriateness for clinical and forensic use. With the publication of the two installments of the special series in *Psychological Assessment,*[8] we think that such a conclusion becomes possible. Furthermore, an important empirical review served to place psychological assessment validity in context relative to other measures used throughout the health sciences.[9] That article presented the findings of over 125 meta-analysis and 800 multimethod assessment studies. The authors' most general conclusion was that psychological assessment instruments perform as effectively as measures in a variety of other health services areas, such as electrocardiograms, mammography, magnetic resonance imaging (MRI), dental radiographs, Papanicolaou (Pap) smears, positron emission tomography (PET) scans, and serum cholesterol level testing.[10] To illustrate, we have included Table 2 from this article (see the Appendix, pp. 224–231), which provides 144 validity coefficients for psychological and medical tests.

Another article,[11] written by authors with opposing views on the Rorschach, moved to a level of specificity that, we be-

Editor's Note: This is published as an official statement by the Board of Trustees for the Society for Personality Assessment. A copy may be obtained from the SPA Web page at www.personality.org.

lieve, allows a clear response to questions about the Rorschach's clinical and forensic utility within the overall context of psychological assessment instruments. We include Tables 2, 3, and 4 from this article (see the Appendix, pp. 232–234), which provides extensive meta-analytic data comparing the validity of the Rorschach to the validity of intelligence scales and the Minnesota Multiphasic Personality Inventory (MMPI) or its revision (MMPI–2).[12] Summarizing these findings, the authors' conclusion is explicit about the Rorschach's validity:

> there is no reason for the Rorschach to be singled out for particular criticism or specific praise. It produces reasonable validity, roughly on par with other commonly used tests.[13]

This article goes on to state that scientific validity is always conditional; that is, questions of validity for any test can only be addressed in the context of specific uses. As such, the Rorschach is like other tests for which research supports their general validity—all have purposes for which they are more or less valid.[14] It should be emphasized that this limitation presents an ongoing challenge for all psychological and medical assessment instruments, and a refined understanding of conditional validity is an important direction for ongoing research.

Overall, meta-analytic reviews and individual studies show the Rorschach possesses adequate psychometric properties. The research literature consistently demonstrates that the Rorschach can be scored reliably, has scores that measure important psychological functions, and has scores that provide unique information that cannot be obtained from other relevant instruments or clinical interviews. The extent to which a test provides unique information concerns incremental validity, which is an understudied topic in psychological and medical assessment in general.[15] However, Rorschach incremental validity has been documented in recent studies[16] and in a structured review of the literature.[17] A summary table from this structured review is provided in the Appendix (on p. 235).

INDEPENDENT BLUE-RIBBON PANEL EXAMINING MMPI–2 AND RORSCHACH VALIDITY

One challenge consumers face when evaluating evidence concerns the potential for researcher biases to influence the evidence that is considered or the manner in which that evidence is interpreted. The potential for such biases is a particular concern for traditional narrative literature reviews, and systematic meta-analytic summaries are preferred because they are less subject to these biases.[18] To obtain an impartial summary of the Rorschach validity evidence, a "blue-ribbon panel" led by Robert Rosenthal, a highly respected statistician, methodologist, and meta-analytic researcher, was commissioned to review and compare the validity of the two most commonly used clinical personality assessment measures, namely the Rorschach and the MMPI/MMPI–2. More important, Rosenthal had not previously conducted research on the Rorschach or MMPI/MMPI–2 and had no professional or personal investment in the outcome of the review.[19] We include Tables 4 and 9 from the review panel's initial article[20] on page 236, and Table 1 from their follow-up article[21] on page 237 of the Appendix. Both articles reached the same conclusions that the MMPI/MMPI–2 and Rorschach validity estimates were not reliably different from each other.[22] The panel also found that the magnitude of the Rorschach and MMPI/MMPI–2 validity was about as good as can be expected for personality tests.[23]

ETHICAL USE AND PROFESSIONAL PRACTICE

An important caveat to our statement regards the proper and appropriate use of the Rorschach for its intended purposes. Ethical and competent use of the Rorschach requires proper training, periodic evaluation and continuing education, and reliance on established and well-researched techniques for administration, coding, and interpretation. As with any test, those using the Rorschach are responsible for its proper application and interpretation. Several specific recommendations can be made that will enhance ethical and professional practice. First, as part of standard clinical care, Rorschach-based inferences, as with inferences from all psychological tests, should be integrated with information from other sources, such as clinical interview and collateral material. Second, clinicians should recognize factors specific to Rorschach testing that may affect or modify interpretation of its scores, such as how engaged a client was with the task.[24] Third, the importance of standardized administration and scoring cannot be overstressed. Atypical administration and scoring can lead to incorrect inferences and risk misinterpretation of Rorschach findings. Fourth, it is important to attend to the research literature to ensure Rorschach inferences are consistent with the evidence. For instance, data have consistently shown a common depression index (DEPI) does not identify interview-based diagnoses of major depression, though common psychosis indexes (SCZI, PTI, TDI) are associated with interview-based diagnoses of psychotic disorders.[25]

In addition, although members of the Board are not aware of psychologists who engage in this kind of practice, it has been asserted that some clinicians use Rorschach findings alone to draw a legal conclusion or determine if a historical event occurred, such as trauma or childhood sexual abuse. Such a practice is indefensible with the Rorschach, as it is with any other personality test.

In conclusion, the Board encourages assessment professionals to serve their clients by avoiding undisciplined practice, as such behavior risks harming patients and other clients, discrediting tests, and discrediting the profession more generally. We encourage psychologists who are aware

of practitioners using the Rorschach or other assessment instruments in an unethical manner to confront those practitioners and if necessary to take further action.[26]

RORSCHACH AND LEGAL SETTINGS

We wish to address as well challenges to the use of the Rorschach in court.[27] Although court and legal settings require a higher level of expertise in the use of the Rorschach for expert testimony, articles summarizing the utility of the Rorschach as an instrument indicate that the Rorschach meets the variety of legal tests for admissibility, including validity, publication in peer reviewed journals, and acceptance within the relevant professional community.[28]

CONCLUSIONS

We recognize that differences of opinion are crucial to the scientific enterprise and we welcome rigorous investigations of specific claims for the validity of specific Rorschach indexes, as we do with all personality assessment techniques. We also recognize that the use of particular instruments in practice is, in part, a matter of personal preference. However, we disagree with the wholesale rejection or discounting of any particular technique where the scientific data do not warrant it. Therefore, it is the position of the Board of Trustees of the Society for Personality Assessment that the Rorschach possesses documented reliability and validity similar to other generally accepted test instruments used in the assessment of personality and psychopathology and that its responsible use in personality assessment is appropriate and justified.

ENDNOTES

[1]For example, for criticisms of the Rorschach, see Garb, Wood, Nezworski, Grove, and Stejskal (2001), Lilienfeld, Wood, and Garb (2000), Wood, Lilienfeld, Garb, and Nezworski (2000b), Wood, Nezworski, Garb, and Lilienfeld (2001a), Wood, Nezworski, Lilienfeld, and Garb (2003), and Wood, Nezworski, and Stejskal (1996a, 1996b); for reviews of evidence supporting reliability and validity, see Meyer (2004), Meyer and Archer (2001), Meyer et al. (2002), Viglione (1999), and Viglione and Hilsenroth (2001). For a broader perspective, see Bornstein and Masling (2005) and Exner (2003) who provide historical reviews of the various controversies that have arisen about the Rorschach since its original publication in 1921 (Rorschach, 1921).

[2]For a review of these arguments, see Garb (1999), Grove and Barden (1999), Grove, Barden, Garb, and Lilienfeld (2002), Lilienfeld et al. (2000), and Wood et al. (2003); although for rejoinders, see Hibbard (2003) and Ritzler, Erard, and Pettigrew (2002a, 2002b).

[3]For recent meta-analytic reviews of Rorschach validity or incremental validity, see Bornstein (1999), Grønnerød (2004), Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neulieb (1999), Jørgensen, Andersen, and Dam (2000, 2001), Meyer (2000), Meyer and Archer (2001), Meyer and Handler (1997, 2000), Rosenthal, Hiller, Bornstein, Berry, and Brunell-Neulieb (2001); for meta-analytic reviews of interrater reliability, see Meyer (2004) and Meyer et al. (2002); for meta-analytic reviews of test–retest reliability or the stability of scores over time, see Grønnerød (2003) and Roberts and DelVecchio (2001). For a review of research documenting incremental validity, see Viglione and Hilsenroth (2001); and for a contemporary study examining the reliability of clinicians interpreting the Rorschach, see Meyer, Mihura, and Smith (2005).

[4]See Exner (1995, 1996), Nezworski and Wood (1995), and Wood et al. (1996a, 1996b).

[5]For example, *Psychological Assessment* (Meyer, 1999, 2001); *Assessment* (Archer, 1999; Wood, Nezworski, Stejskal, Garven, & West, 1999); *Journal of Personality Assessment* (Kinder, 2001).

[6]See *Clinical Psychology: Science and Practice* (Aronow, 2001; Exner, 2001; Hunsley & DiGuilio, 2001; Meyer, 2001; Widiger, 2001; Wood, Nezworski, Garb, et al., 2001a, 2001b); *Journal of Clinical Psychology* (Garfield, 2000a, 2000b; Lerner, 2000; Weiner, 2000; Wood et al., 2000a, 2000b); *Journal of Forensic Psychology Practice* (Gacono, 2002; Hamel, Gallager, & Soares, 2001; Wood, Nezworski, Stejskal, & McKinzey, 2001), and *Psychology, Public Policy, and Law* (Grove et al., 2002; Ritzler et al., 2002a; 2002b).

[7]For overviews, see Meyer and Archer (2001) and Weiner (2001).

[8]Meyer (Ed.; 1999, 2001).

[9]Meyer et al. (2001).

[10]"Validity coefficients for many psychological tests are indistinguishable from those observed for many medical tests. For instance, when considering validity coefficients in the .30–.50 range, one finds results from the MMPI, Millon Clinical Multiaxial Inventory, Thematic Apperception Test, Rorschach, Hare Psychopathy Checklist, various neuropsychological and cognitive tests, and the impact of psychological assessment feedback on the subsequent well-being of patients. One also finds results from electrocardiograms, mammography, magnetic resonance imaging (MRI), dental radiographs, Papanicolaou (Pap) smears, cardiac fluoroscopy, single photon emission computed tomography, technetium bone scanning, and serum cholesterol level." (Meyer et al., 2001, p. 135).

[11]Meyer and Archer (2001).

[12]MMPI: Hathaway and McKinley (1943); MMPI–2: Butcher, Dahlstrom, Graham, Tellegen, and Kaemmer (1989).

[13]Meyer and Archer (2001, pp. 491–492).

[14]Weiner (1996).

[15]See Hunsley (2003) and Hunsley and Meyer (2003).

[16]See Fowler, Piers, Hilsenroth, Holdwick, and Padawer (2001), Hartmann, Sunde, Kristensen, and Martinussen (2003), Hartmann, Wang, Berg, and Sæther (2003), Janson and Stattin (2003), Meyer (2000), Stokes, Pogge, Powell-Lunder, Ward, Bilginer, and DeLuca (2003), and Sultan, Jebrane, and Heurtier-Hartemann (2002).

[17]See Viglione and Hilsenroth (2001), which summarizes findings described in Viglione (1999).

[18]See Hunter and Schmidt (2004) or Lipsey and Wilson (2001).

[19]At the same time, to ensure each test was adequately represented, the panel included researchers with recognized meta-analytic expertise applied to the Rorschach (Robert F. Bornstein) and the MMPI/MMPI–2 (David T. R. Berry).

[20]Hiller et al. (1999).

[21]Rosenthal et al. (2001).

[22]"In a meta-analytic comparison of criterion-related validity coefficients for the MMPI and for the Rorschach, we found both instruments to have validity effect sizes of substantial magnitude (unweighted mean *r* of .30 and .29 for the MMPI and Rorschach, respectively). Validity estimates for the MMPI and Rorschach were not reliably different from each other, even when studies in which test predictors and criterion variables had common measurement methods were removed from consideration. … The methodological features of this study, including random sampling from the published literature, expert judgments for inclusion of validity evidence, and the use of accepted effect size estimation techniques, lend

greater credibility to these results compared with those from previous efforts." (Hiller et al. 1999, pp. 291–292).

[23]"As noted by Cohen (1988), 'when one looks at the near-maximum correlation coefficients of personality measures with … real-life criteria, the values one encounters fall at the order of … $r = .30$' (p. 81). In other words, validity for these instruments is about as good as can be expected for personality tests." (Hiller et al., 1999. p. 291).

[24]See Meyer (1993, 1997).

[25]See Jørgensen et al. (2000, 2001). DEPI = Depression Index, SCZI = Schizophrenia Index, PTI = Perceptual-Thinking Index, TDI = Thought Disorder Index.

[26]The Code of Ethics of the American Psychological Association (APA, 2002) can serve as a guideline for further action, including, when appropriate, filing an ethical complaint with the APA, relevant state association, Board of Examiners, or other professional association.

[27]Grove and Barden (1999); Grove et al. (2002), and Wood, Nezworski, Stejskal et al. (2001).

[28]Gacono, Evans, and Viglione (2002); Hilsenroth and Stricker (2004) McCann (1998); Ritzler et al. (2002a, 2002b).

REFERENCES

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060–1073.

Archer, R. P. (Ed.). (1999). Perspectives on the Rorschach [Special section]. *Assessment, 6,* 307–351.

Aronow, E. (2001). CS norms, psychometrics, and possibilities for the Rorschach technique. *Clinical Psychology: Science & Practice, 8*, 383–385.

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

Bornstein, R. F., & Masling, J. M. (2005). *Scoring the Rorschach: Seven validated systems*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Minnesota Multiphasic Personality Inventory—2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Exner, J. E. (1995). "Narcissism in the Comprehensive System for the Rorschach": Comment. *Clinical Psychology: Science & Practice, 2*, 200–206.

Exner, J. E. (1996). A comment on "The Comprehensive System for the Rorschach: A critical examination." *Psychological Science, 7*, 11–13.

Exner, J. E. (2001). A comment on "The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach." *Clinical Psychology: Science & Practice, 8*, 386–396.

Exner, J. E. (2003). *The Rorschach: A Comprehensive System* (4th ed.). New York: Wiley.

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

Gacono, C. B. (Ed.). (2002). Forensic psychodiagnostic testing [Special section]. *Journal of Forensic Psychology Practice, 2,* 1–117.

Gacono, C. B., Evans, F. B., & Viglione, D. J. (2002). The Rorschach in forensic practice. *Journal of Forensic Psychology Practice, 2*, 33–54.

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

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment, 13*, 433–448.

Garfield, S. L. (2000a). The Rorschach test in clinical diagnosis. *Journal of Clinical Psychology, 56*, 387–393.

Garfield, S. L. (2000b). "The Rorschach test in clinical diagnosis": A brief commentary. *Journal of Clinical Psychology, 56*, 431–434.

Grønnerød, C. (2003). Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment, 80*, 272–293.

Grønnerød, C. (2004). Rorschach assessment of changes following psychotherapy: A meta-analytic review. *Journal of Personality Assessment, 83*, 256–276.

Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, & Law, 5*, 224–242.

Grove, W. M., Barden, R. C., Garb, H. N., & Lilienfeld, S. O. (2002). Failure of Rorschach-Comprehensive-System-based testimony to be admissible under the Daubert-Joiner-Kumho standard. *Psychology, Public Policy, & Law, 8*, 216–234.

Hamel, M., Gallagher, S., & Soares, C. (2001). The Rorschach: Here we go again. *Journal of Forensic Psychology Practice, 1*, 79–88.

Hartmann, E., Sunde, T., Kristensen, W., & Martinussen, M. (2003a). Psychological measures as predictors of military training performance. *Journal of Personality Assessment, 80*, 87–98.

Hartmann, E., Wang, C. E., Berg, M., & Saether, L. (2003b). Depression and vulnerability as assessed by the Rorschach method. *Journal of Personality Assessment, 81*, 242–255.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* Minneapolis: University of Minnesota Press.

Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) "The scientific status of projective techniques". *Journal of Personality Assessment, 80*, 260–271.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*, 278–296.

Hilsenroth, M. J., & Stricker, G. (2004). A consideration of challenges to psychological assessment instruments used in forensic settings: Rorschach as exemplar. *Journal of Personality Assessment, 83*, 141–152.

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15*, 443–445.

Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science & Practice, 8*, 378–382.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment, 15*, 446–455.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analyses: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Janson, H., & Stattin, H. (2003). Prediction of adolescent and adult from childhood Rorschach ratings. *Journal of Personality Assessment, 81*, 51–63.

Jorgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment, 7*, 259–280.

Jorgensen, K., Andersen, T. J., & Dam, H. (2001). "The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review": Erratum. *Assessment, 8*, 355.

Kinder, B. N. (Ed.). (2001). More data on the current Rorschach controversy [Special series]. *Journal of Personality Assessment, 77,* 1–70.

Lerner, P. M. (2000). A nonreviewer's comment: On the Rorschach and baseball. *Journal of Clinical Psychology, 56*, 439.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27–66.

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

McCann, J. T. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment, 70*, 125–144.

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

Meyer, G. J. (1997). On the integration of personality assessment methods: The Rorschach and MMPI. *Journal of Personality Assessment, 68*, 297–330.

Meyer, G. J. (Ed.). (1999). Special section 1: The utility of the Rorschach for clinical assessment [Special section]. *Psychological Assessment, 11*, 235–302.

Meyer, G. J. (2000). The incremental validity of the Rorschach Prognostic Rating Scale over the MMPI Ego Strength Scale and IQ. *Journal of Personality Assessment, 74*, 356–370.

Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science & Practice, 8*, 389–396.

Meyer, G. J. (Ed.). (2001). Special section 2: The utility of the Rorschach for clinical assessment [Special section]. *Psychological Assessment, 13*, 419–502.

Meyer, G. J. (2004). The reliability and validity of the Rorschach and TAT compared to other psychological and medical procedures: An analysis of systematically gathered evidence. In M. J. Hilsenroth & D. Segal (Vol. Eds.) & M. Hersen (Series Ed.), *Personality assessment. Vol. 2. Comprehensive handbook of psychological assessment* (pp. 315–342). Hoboken, NJ: Wiley.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13*, 486–502.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128–165.

Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment, 69*, 1–38.

Meyer, G. J., & Handler, L. (2000). "The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale": Correction. *Journal of Personality Assessment, 74*, 504–506.

Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C., et al. (2002). An examination of interrater reliability for scoring the Rorschach comprehensive system in eight data sets. *Journal of Personality Assessment, 78*, 219–274.

Meyer, G. J., Mihura, J. L., & Smith, B. L. (2005). The interclinician reliability of Rorschach interpretation in four data sets. *Journal of Personality Assessment, 84*, 296–314.

Nezworski, M. T., & Wood, J. M. (1995). Narcissism in the Comprehensive System for the Rorschach. *Clinical Psychology: Science & Practice, 2*, 179–199.

Ritzler, B., Erard, R., & Pettigrew, G. (2002a). A final reply to Grove and Barden: The relevance of the Rorschach Comprehensive System for expert testimony. *Psychology, Public Policy, & Law, 8*, 235–246.

Ritzler, B., Erard, R., & Pettigrew, G. (2002b). Protecting the integrity of Rorschach expert witnesses: A reply to Grove and Barden (1999) re: The admissibility of testimony under Daubert/Kumho analyses. *Psychology, Public Policy, & Law, 8*, 201–215.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3–25.

Rorschach, H. (1921). *Psychodiagnostik*. Bern, Switzerland: Bircher.

Rosenthal, R., Hiller, J. B., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (2001). Meta-analytic methods, the Rorschach, and the MMPI. *Psychological Assessment, 13*, 449–451.

Stokes, J. M., Pogge, D. L., Powell-Lunder, J., Ward, A. W., Bilginer, L., & DeLuca, V. A. (2003). The Rorschach Ego Impairment Index: Prediction of treatment outcome in a child psychiatric population. *Journal of Personality Assessment, 81*, 11–19.

Sultan, S., Jebrane, A., & Heurtier-Hartemann, A. (2002). Rorschach variables related to blood glucose control in insulin-dependent diabetes patients. *Journal of Personality Assessment, 79*, 122–141.

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11*, 251–265.

Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment, 13*, 452–471.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment, 8*, 206–213.

Weiner, I. B. (2000). Using the Rorschach properly in practice and research. *Journal of Clinical Psychology, 56*, 435–438.

Weiner, I. B. (2001). Advancing the science of psychological assessment: The Rorschach Inkblot Method as exemplar. *Psychological Assessment, 13*, 423–432.

Widiger, T. A. (2001). The best and the worst of us? *Clinical Psychology: Science & Practice, 8*, 374–377.

Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000a). Limitations of the Rorschach as a diagnostic tool: A reply to Garfield (2000), Lerner (2000), and Weiner (2000). *Journal of Clinical Psychology, 56*, 441–448.

Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000b). The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology, 56*, 395–430.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001a). The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science & Practice, 8*, 350–373.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001b). Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations. *Clinical Psychology: Science & Practice, 8*, 397–402.

Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach? Science confronts the controversial inkblot test*. San Francisco: Jossey-Bass.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996a). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*, 3–10.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996b). Thinking critically about the Comprehensive System for the Rorschach: A reply to Exner. *Psychological Science, 7*, 14–17.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). "Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996)": Erratum. *Assessment, 6*, 305.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., & McKinzey, R. K. (2001). Problems of the Comprehensive System for the Rorschach in forensic settings: Recent developments. *Journal of Forensic Psychology Practice, 1*, 89–103.

APPENDIX: Supporting Tables

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001)

Table 2
Examples of Testing and Assessment Validity Coefficients With an Emphasis on Meta-Analytic Results

Predictor and criterion (study and notes)	*r*	*N*
1. Dexamethasone suppression test scores and response to depression treatment (Ribeiro, Tandon, Grunhaus, & Greden, 1993).[a]	.00	2,068
2. Fecal occult blood test screening and reduced death from colorectal cancer (Towler et al., 1998).	.01	329,642
3. Routine umbilical artery Doppler ultrasound and reduced perinatal deaths in low-risk women (Goffinet, Paris-Llado, Nisand, & Bréart, 1997; the authors also examined the impact of routine umbilical artery ultrasound on 13 other measures of successful outcome. The average effect size across these other criteria was $r = -.0036$ [*ns* from 6,373 to 11,375], with the largest correlation in the expected direction being .0097 [for Apgar scores at 5 minutes]).	.01	11,375
4. Routine ultrasound examinations and successful pregnancy outcomes (Bucher & Schmidt, 1993; outcomes considered were live births [$r = .0009$], no induced labor [$r = .0176$], no low Apgar scores [$r = -.0067$], no miscarriages [$r = .0054$], and no perinatal mortality [$r = .0168$]).	.01	16,227
5. MMPI Ego Strength scores and subsequent psychotherapy outcome (Meyer & Handler, 1997; this meta-analysis considered only studies in which the Ego Strength scale was used along with the Rorschach PRS).	.02	280
6. Routine umbilical artery Doppler ultrasound and reduced perinatal deaths in high-risk women (Alfirevic & Neilson, 1995; the authors also examined the impact of routine umbilical artery ultrasound on 19 other measures of successful outcome. The average effect size across these other criteria was $r = .018$ [*ns* from 476 to 7,474]).	.03	7,474
7. Denial/repressive coping style and development of breast cancer (McKenna, Zevon, Corn, & Rounds, 1999; weighted effect size computed from the study data in their Table 1).	.03	12,908
8. Triple marker[b] prenatal screening of maternal serum and identification of Trisomy 18 (Yankowitz, Fulton, Williamson, Grant, & Budelier, 1998).[c]	.03	40,748
9. Impact of geriatric medical assessment teams on reduced deaths (data combined from the meta-analysis by Rubenstein, Stuck, Siu, & Wieland, 1991, and the following more recent studies: Boult et al., 1994; Büla et al., 1999; Burns, Nichols, Graney, & Cloar, 1995; Englehardt et al., 1996; Fabacher et al., 1994; Fretwell et al., 1990; Germain, Knoeffel, Wieland, & Rubenstein, 1995; Hansen, Poulsen, & Sørensen, 1995; Harris et el., 1991; Karppi & Tilvis, 1995; Naughton, Moran, Feinglass, Falconer, & Williams, 1994; Reuben et la., 1995; Rubenstein, Josephson, Harker, Miller, & Wieland, 1995; Rubin, Sizemore, Loftis, & de Mola, 1993; Silverman et al., 1995; Siu et al., 1996; Thomas, Brahan, & Haywood, 1993; and Trentini et al., 1995; only the latest available outcome data were used for each sample).	.04	10,065
10. MMPI depression profile scores and subsequent cancer within 20 years (Persky, Kempthorne-Rawson, & Shekelle, 1987).[c]	.05	2,018
11. Ventilatory lung function test scores and subsequent lung cancer within 25 years (Islam & Schottenfeld, 1994).[c]	.06	3,956
12. Rorschach Interaction Scale scores and subsequent cancer within 30 years (Graves, Phil, Mead & Pearson, 1986; scores remained significant predictors after controlling for baseline smoking, serum cholesterol, systolic blood pressure, weight, and age).[c]	.07	1,027
13. Unique contribution of an MMPI high-point code (vs. other codes) to conceptually relevant criteria (McGrath & Ingersoll, 1999a, 1999b).	.07	8,614
14. MMPI scores and subsequent prison misconduct (Gendreau, Goggin, & Law, 1997).	.07	17,636
15. Beck Hopelessness Scale scores and subsequent suicide (data combined from Beck, Brown, Berchick, Stewart, & Steer, 1990; and Beck, Steer, Kovacs, & Garrison, 1985).[c]	.08	2,123
16. MMPI elevations on Scales F, 6, or 8 and criminal defendant incompetency (Nicholson & Kugler, 1991).	.08	1,461
17. Extraversion test scores and success in sales (concurrent and predictive; data combined from Barrick & Mount, 1991, Table 2; Salgado, 1997, Table 3; and Vinchur, Schippman, Switzer, & Roth, 1998 [coefficients from their Tables 2 and 3 were averaged, and the largest *N* was used for the overall sample size]).	.08	6,004
18. Attention and concentration test scores and residual mild head trauma (Binder, Rohling, & Larrabee, 1997).	.09	622
19. In cervical cancer, lack of glandular differentiation on tissue biopsy and survival past 5 years (Heatley, 1999; this study reported two meta-analyses. The other one found that nuclear DNA content was of no value for predicting cancer progression in initially low-grade cervical intraepithelial neoplasia).	.11	685

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
20. Negative emotionality test scores and subsequent heart disease (Booth-Kewley & Friedman, 1987; data were derived from their Table 7, with negative emotionality defined by the weighted effect for anger/hostility/aggression, depression, and anxiety).	.11	(*k* = 11)
21. Triple marker[b] prenatal screening of maternal serum and identification of Down's syndrome (Conde-Agudelo & Kafury-Goeta, 1998; results were reported across all ages).	.11	194,326
22. General cognitive ability and involvement in automobile accidents (Arthur, Barrett, & Alexander, 1991).	.12	1,020
23. Conscientiousness test scores and job proficiency (concurrent and predictive; data combined from Barrick & Mount, 1991, Table 3; Mount, Barrick, & Stewart, 1998; Salgado, 1998, Table 1; and Vinchur et al., 1998 [coefficients from their Tables 2 and 3 were averaged, and the largest *N* was used for the overall sample size]).	.12	21,650
24. Platform posturography and detection of balance deficits due to vestibular impairment (Di Fabio, 1996).	.13	1,477
25. General intelligence and success in military pilot training (Martinussen, 1996).	.13	15,403
26. Self-report scores of achievement motivation and spontaneous achievement behavior (Spangler, 1992; coefficient derived from the weighted average of the semioperant and operant criterion data reported in Spangler's Table 2).	.15	(*k* = 104)
27. Graduate Record Exam Verbal or Quantitative scores and subsequent graduate GPA in psychology (E. L. Goldberg & Alliger, 1992).	.15	963
28. Low serotonin metabolites in cerebrospinal fluid (5-HIAA) and subsequent suicide attempts (Lester, 1995).	.16	140
29. Personality tests and conceptually meaningful job performance criteria (data combined from Robertson & Kinder, 1993; Tett, Jackson, & Rothstein, 1991; and Tett, Jackson, Rothstein, & Reddon, 1994; we used the single scale predictors from Robertson & Kinder [their Table 3] and the confirmatory results from Table 1 in Tett et al., 1994).	.16	11,101
30. Implicit memory tests and differentiation of normal cognitive ability from dementia (Meiran & Jelicic, 1995).	.16	1,156
31. MMPI Cook–Medley hostility scale elevations and subsequent death from all causes (T. Q. Miller, Smith, Turner, Guijarro, & Hallet, 1996; data were drawn from their Table 6).	.16	4,747
32. Motivation to manage from the Miner Sentence Completion Test and managerial effectiveness (Carson & Gilliard, 1993; results were averaged across the three performance criterion measures of managerial success. Because the three criterion measures were not independent across studies, the *N* reported is the largest *N* used for any single criterion).	.17	2,151
33. Extraversion and subjective well-being (DeNeve & Cooper, 1998).	.17	10,364
34. MRI T_2 hyperintensities and differentiation of affective disorder patients from healthy controls (Videbech, 1997; data from Videbech's Tables 1 and 2 were combined, but only those statistics used by the original author are included here).	.17	1,575
35. Test anxiety scales and lower school grades (Hembree, 1988; reported effect is the average effect size for the course grade and GPA data from Hembree's Table 1. Participants were assumed to be independent across studies).	.17	5,750
36. High trait anger assessed in an interpersonal analogue and elevated blood pressure (Jorgensen, Johnson, Kolodziej, & Schreer, 1996; data come from the "Overall" column of their Table 4).	.18	(*k* = 34)
37. Reduced blood flow and subsequent thrombosis or failure of synthetic hemodialysis graft (Paulson, Ram, Birk, & Work, 1999).	.18	4,569
38. MMPI validity scales and detection of known or suspected underreported psychopathology (Baer, Wetter, & Berry, 1992; weighted average effect size was calculated from data reported in their Table 1 for all studies using participants presumed to be underreporting).	.18	328
39. Dexamethasone suppression test scores and subsequent suicide (Lester, 1992).	.19	626
40. Short-term memory tests and subsequent job performance (Verive & McDaniel, 1996).	.19	17,741
41. Depression test scores and subsequent recurrence of herpes simplex virus symptoms (Zorrilla, McKay, Luborsky, & Schmidt, 1996; effect size is for prospective studies).	.20	333
42. Four preoperative cardiac tests and prediction of death or MI within 1 week of vascular surgery (Mantha et al., 1994; the four tests considered were dipyridamole-thallium scintigraphy, ejection fraction estimation by radionuclide ventriculography, ambulatory ECG, and dobutamine stress ECG. The authors concluded no test was conclusively superior to the others).	.20	1,991
43. Scholastic Aptitude Test scores and subsequent college GPA (Baron & Norman, 1992).[c]	.20	3,816

(*table continues*)

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
44. Self-reported dependency test scores and physical illness (Bornstein, 1998; weighted effect size was calculated from the retrospective studies reported in Bornstein's Table 1 [Studies 3, 5, 7, 8, 13, and 19] and the prospective studies listed in Bornstein's Table 2 [Studies 1–4]).	.21	1,034
45. Dexamethasone suppression test scores and psychotic vs. nonpsychotic major depression (Nelson & Davis, 1997; effect size calculated from the weighted effects for the individual studies in their Table 1).	.22	984
46. Traditional ECG stress test results and coronary artery disease (Fleischmann, Hunink, Kuntz, & Douglas, 1998; results were estimated from the reported sensitivity and specificity in conjunction with the base rate of coronary artery disease and the total independent *N* across studies).	.22	5,431
47. Graduate Record Exam Quantitative scores and subsequent graduate GPA (Morrison & Morrison, 1995).	.22	5,186
48. TAT scores of achievement motivation and spontaneous achievement behavior (Spangler, 1992; coefficient was derived from the weighted average of the semioperant and operant criterion data in Spangler's Table 2).	.22	(*k* = 82)
49. Isometric strength test scores and job ratings of physical ability (Blakley, Quiñones, & Crawford, 1994).	.23	1,364
50. Single serum progesterone testing and diagnosis of ectopic pregnancy (Mol, Lijmer, Ankum, van der Veen, & Bossuyt, 1998; following the original authors, we used only the 18 prospective or retrospective cohort studies listed in their Table III).	.23	6,742
51. Cognitive multitask performance test scores and subsequent pilot proficiency (Damos, 1993).	.23	6,920
52. WISC distractibility subscales and learning disability diagnoses (Kavale & Forness, 1984; the effect sizes from this meta-analysis are likely to be underestimates because the authors computed the average effect for individual test scales rather than the effect for a composite pattern).	.24	(*K* = 54)
53. Fetal fibronectin testing and prediction of preterm delivery (Faron, Boulvain, Irion, Bernard, & Fraser, 1998; data were aggregated across low- and high-risk populations and across designs with single or repeated testing for all studies using delivery before 37 weeks as the criterion).	.24	7,900
54. Decreased bone mineral density and lifetime risk of hip fracture in women (Marshall, Johnell, & Wedel, 1996; the results were restricted to those from absorptiometry using single or dual energy, photon, or X-ray; quantitative CT; quantitative MRI; or ultrasound scanning. The overall effect was estimated from their Table 3 using a total lifetime incidence of 15%; the effect would be smaller if the lifetime risk incidence was lower [e.g., if the incidence were 3%, the effect would be *r* = .13]. Total *N* was derived from the *n* for each study in their Table 1 reporting the incidence of hip fractures).	.25	20,849
55. General intelligence test scores and functional effectiveness across jobs (Schmitt, Gooding, Noe, & Kirsch, 1984; data were obtained from their Table 4).	.25	40,230
56. Internal locus of control and subjective well-being (DeNeve & Cooper, 1998).	.25	8,481
57. Integrity test scores and subsequent supervisory ratings of job performance (Ones, Viswesvaran, & Schmidt, 1993; effect size was taken from the "predictive-applicant" cell of their Table 8).	.25	7,550
58. Self-reported dependency test scores and dependent behavior (Bornstein, 1999; coefficient was derived from all results listed in Bornstein's Table 1 as reported in his footnote 8).	.26	3,013
59. Self-efficacy appraisals and health-related treatment outcomes (Holden, 1991).	.26	3,527
60. Elevated Jenkins Activity Survey scores and heart rate and blood pressure reactivity (Lyness, 1993; the effect size reflects the average reactivity for heart rate, systolic blood pressure, and diastolic blood pressure as reported in Lyness's Table 6. It was assumed that overlapping studies contributed to each of these criterion estimates, so *k* was estimated as the largest number of effect sizes contributing to a single criterion measure).	.26	(*k* = 44)
61. Combined internal, stable, and global attributions for negative event outcomes and depression (Sweeney, Anderson, & Bailey, 1986; only the finding that dealt with the composite measure of attributions and negative outcome was included. Coefficients were lower for positive outcomes and for single types of attributions [e.g., internal]).	.27	5,788
62. Neuroticism and decreased subjective well-being (DeNeve & Cooper, 1998).	.27	9,777
63. Screening mammogram results and detection of breast cancer within 2 years (Mushlin, Kouides, & Shapiro, 1998).	.27	192,009

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
64. Microbiologic blood culture tests to detect bloodstream infection from vascular catheters (Siegman-Igra et al., 1997; only results from studies without criterion contamination were summarized [see Siegman-Igra et al., 1997, pp. 933–934]).	.28	1,354
65. C-reactive protein test results and diagnosis of acute appendicitis (Hallan & Åsberg, 1997; mean weighted effect size was derived from data in their Table 1, excluding two studies that did not use histology as the validating criteria and one study that did not report the prevalence of appendicitis).	.28	3,338
66. Graduate Record Exam Verbal scores and subsequent graduate GPA (Morrison & Morrison, 1995).	.28	5,186
67. Hare Psychopathy Checklist scores and subsequent criminal recidivism (Salekin, Rogers, & Sewell, 1996; only effects for predictive studies were summarized).	.28	1,605
68. Short-term memory tests and subsequent performance on job training (Verive & McDaniel, 1996).	.28	16,521
69. Cranial ultrasound results in preterm infants and subsequent developmental disabilities (Ng & Dear, 1990).	.29	1,604
70. Serum CA-125 testing and detection of endometriosis (Mol, Bayram, et al., 1998).	.29	2,811
71. Neuropsychological test scores and differentiation of patients with multiple sclerosis (Wishart & Sharpe, 1997).	.29	(*k* = 322)
72. For women, ECG stress test results and detection of coronary artery disease (Kwok, Kim, Grady, Segal, & Redberg, 1999; our *N* was obtained from their Table 1. It differs from the *N* reported by the authors [3,872 vs. 3,721], though it is not clear what would account for this difference. Although the article also examined the thallium stress test and the exercise ECG, there was not sufficient data for us to generate effect sizes for these measures).	.30	3,872
73. YASR total problems and psychiatric referral status (receiving treatment vs. not; Achenbach, 1997; effect size was estimated from data in Part 1 of Achenbach's Table 7.5. Because the percentages listed in this table were too imprecise to accurately generate effect size estimates, all possible 2 × 2 tables that would match the given percentages were generated. Subsequently, the effect size was obtained from those 2 × 2 tables that also produced odds ratios that exactly matched the odds ratios reported in the text. When rounded to two decimal places, all appropriate 2 × 2 tables produced the same effect size. The effect size compares the self-reports of young adults in treatment with the self-reports of demographically matched controls who were not receiving treatment).[c]	.30	1,142
74. Fecal leukocyte results and detection of acute infectious diarrhea (Huicho, Campos, Rivera,. & Guerrant, 1996; results are reported for the most studied test [*K* = 19]. For the remaining tests, effect sizes could be generated for only two small studies of fecal lactoferrin, and the average results for occult blood tests were lower [*r* = .26; *K* = 7]).	.30	7,132
75. Neuropsychological test scores and differentiation of learning disabilities (Kavale & Nye, 1985; we report the results for neuropsychological functioning because it was studied most frequently).	.30	(*K* = 394)
76. Continuous performance test scores and differentiation of ADHD and control children (Losier, McGrath, & Klein, 1996; overall sample weighted effect was derived by combining the omission and commission data reported in their Tables 7 and 8).	.31	720
77. Effects of psychological assessment feedback on subsequent patient well-being (coefficient combined the follow-up data reported in Finn & Tonsager, 1992; and Newman & Greenway, 1997).[c]	.31	120
78. Expressed emotion on the CFI and subsequent relapse in schizophrenia and mood disorders (Butzlaff & Hooley, 1998).	.32	1,737
79. CT results and detection of aortic injury (Mirvis, Shanmuganathan, Miller, White, & Turney, 1996; from the information provided, an effect size could not be computed for two studies included in this meta-analysis).	.32	3,579
80. Screening mammogram results and detection of breast cancer within 1 year (Mushlin, Kouides, & Shapiro, 1998; overall effect size includes studies that combined mammography with clinical breast examination).	.32	263,359
81. Halstead–Reitan Neuropsychological Tests and differentiation of impaired vs. control children (Forster & Leckliter, 1994; the reported weighted effect size is slightly inflated because some observations were based on group differences relative to the control group standard deviation [rather than the pooled standard deviation]. When possible, effect sizes were computed directly from the data reported in their Tables 1 and 2. The reported *N* indicates the total number of independent observations across studies).	.33	858

[*table continues*]

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
82. CT results for enlarged ventricular volume and differentiation of schizophrenia from controls (Raz & Raz, 1990).	.33	(*k* = 53)
83. Long-term memory test scores and diagnosis of multiple sclerosis (Thornton & Raz, 1997; effect size was obtained from their Table 2 with the outlier study excluded).	.33	(*K* = 33)
84. Hare Psychopathy Checklist scores and subsequent violent behavior (Salekin, Rogers, & Sewell, 1996; only effects for predictive studies were summarized).	.33	1,567
85. Alanine aminotransferase results and detection of improved liver function in hepatitis C patients (Bonis, Ioannidis, Cappelleri, Kaplan, & Lau, 1997; data reflect the criterion of any histologically identified improvement).	.34	480
86. Rorschach scores and conceptually meaningful criterion measures (data combined from Atkinson, 1986, Table 1 [*K* = 79]; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999, Table 4 [*K* = 30]; and K. P. Parker, Hanson, & Hunsley, 1988, Table 2 [*K* = 14]. Hiller et al. expressed concern that Atkinson's and K. P. Parker et al.'s effect size estimates may have been inflated by some results derived from unfocused *F* tests [i.e., with >1 *df* in the numerator]. However, Atkinson excluded effects based on *F*, and K. P. Parker et al.'s average effect size actually increased when *F* test results were excluded. Recently, Garb, Florio, & Grove, 1998, conducted reanalyses of K. P. Parker et al.'s data. Although these reanalyses have been criticized [see K. P. Parker, Hunsley, & Hanson, 1999], if the results from Garb et al.'s first, second, or third analysis were used in lieu of those from K. P. Parker et al., the synthesized results reported here would change by −.0096, −.0036, or −.0007, respectively, for the Rorschach and by .0203, .0288, or .0288, respectively, for the MMPI [see Entry 100, this table]).	.35	(*K* = 122)
87. Papanicolaou Test (Pap smear) and detection of cervical abnormalities (Fahey, Irwig, & Macaskill, 1995; overall weighted effect calculated from data reported in their Appendix 1).	.36	17,421
88. Conventional dental X-rays and diagnosis of biting surface cavities (occlusal caries; Ie & Verdonschot, 1994; the overall weighted effect was derived from all the studies listed in their Table 1. In each case, the original citations were obtained, and raw effect sizes were calculated from the initial study).	.36	5,466
89. Incremental contribution of Rorschach PRS scores over IQ to predict psychotherapy outcome (Meyer, 2000).	.36	290
90. Rorschach or Apperceptive Test Dependency scores and physical illness (Bornstein, 1998; weighted effect size was calculated from the retrospective studies reported in Bornstein's Table 1 [Studies 1, 11, 14–16, and 18]. No prospective studies used these types of scales as predictors).	.36	325
91. Assessment center evaluations and job success (data combined from Schmitt, Gooding, Noe, & Kirsch, 1984; and Gaugler, Rosenthal, Thornton, & Bentson, 1987; the overall effect size was derived from the sample weighted average reported in each study. Although Schmitt et al.'s study was conducted earlier than Gaugler et al.'s, they relied on a larger *N*. Because each meta-analysis undoubtedly relied on some common studies, the *N* reported here is from Schmitt et al.).	.37	15,345
92. Competency screening sentence-completion test scores and defendant competency (Nicholson & Kugler, 1991).	.37	627
93. MCMI–II scale score and average ability to detect depressive or psychotic disorders (Ganellen, 1996; each individual study contributed one effect size averaged across diagnostic criteria and type of predictor scales [single vs. multiple scales]. Results were averaged across analyses reported in different publications using the same sample. Although Ganellen reported larger effect sizes for studies that used multiscale predictors, these studies relied on unreplicated multivariate predictor equations. As such, multiscale predictors were averaged with hypothesized, single-scale predictors).[c]	.37	575
94. MMPI scale scores and average ability to detect depressive or psychotic disorders (Ganellen, 1996; see Entry 93, this table).[c]	.37	927
95. Rorschach Apperceptive Test Dependency scores and dependent behavior (Bornstein, 1999; coefficient was derived from all results listed in Bornstein's Table 1 as reported in his footnote 8).	.37	1,808
96. Accuracy of home pregnancy test kits in patients conducting testing at home (Bastian, Nanda, Hasselblad, & Simel, 1998; results derived from the pooled "effectiveness score," which was described and thus treated as equivalent to Cohen's *d*. Also, findings were very different when tests were evaluated using researcher-assisted volunteers rather than actual patients [*r* = .81; *N* = 465]).	.38	155

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
97. Sperm penetration assay results and success with in vitro fertilization (Mol, Meijer, et al., 1998).	.39	1,335
98. Endovaginal ultrasound in postmenopausal women and detection of endometrial cancer (Smith-Bindman et al., 1998; effect size was derived from the authors' pooled results [their Table 2] using their recommended cutoff of 5 mm to define endometrial thickening).	.39	3,443
99. MMPI Validity scales and detection of underreported psychopathology (primarily analogue studies; Baer, Wetter, & Berry, 1992; weighted average effect size calculated from data in their Table 1).	.39	2,297
100. MMPI scores and conceptually meaningful criterion measures (data combined from Atkinson, 1986, Table 1; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999, Table 4; and K. P. Parker, Hanson, & Hunsley, 1988, Table 2. See also Entry 86, this table).	.39	(*K* = 138)
101. Neuropsychologists' test-based judgments and presence/absence of impairment (Garb & Schramke, 1996; coefficient was calculated from the accuracy of judgments relative to base rates [see Garb & Schramke, 1996, pp. 143, 144–145]).	.40	2,235
102. Prostate-specific antigen and estimated detection of prostate cancer for men aged 60–70 (Aziz & Barathur, 1993).	.40	4,200
103. Short-term verbal learning and differentiation of major depression from controls (Veiel, 1997; although the author reported many effect sizes, we report the variable that was studied most often).	.41	(*K* = 10)
104. CT results and detection of lymph node metastases in cervical cancer (Scheidler, Hricak, Yu, Subak, & Segal, 1997; an effect size could not be computed for one study included in this meta-analysis).	.41	1,022
105. Dissociative Experiences Scale scores and detection of MPD or PTSD vs. controls (Van IJzendoorn & Schuengel, 1996; we assumed the *N*s for both criterion diagnoses were not independent, so the reported *N* is that for the largest analysis).	.41	1,705
106. Colposcopy and detection of normal/low-grade SIL vs. high-grade SIL/cancer of the cervix (Mitchell, Schottenfeld, Tortolero-Luna, Cantor, & Richards-Kortum, 1998; effect sizes were calculated from data reported in their Table 3).	.42	2,249
107. Cortical tuber count on MRI and degree of impaired cognitive development in tuberous sclerosis (M. Goodman et al., 1997).	.43	157
108. Conventional dental X-rays and diagnosis of between-tooth cavities (approximal caries; Van Rijkom & Verdonschot, 1995; this is an unweighted effect size for all studies that used a "strong" validity criterion [i.e., microradiography, histology, or cavity preparation]).	.43	(*K* = 8)
109. Cardiac fluoroscopy and diagnosis of coronary artery disease (Gianrossi, Detrano, Colombo, & Froelicher, 1990).	.43	3,765
110. Serum chlamydia antibody levels and detection of fertility problems due to tubal pathology (Mol et al., 1997; only the results for the optimal predictor assays and optimal criterion measures are presented).	.44	2,131
111. Rorschach PRS scores and subsequent psychotherapy outcome (Meyer & Handler, 1997, 2000).	.44	783
112. Digitally enhanced dental X-rays and diagnosis of biting surfaces cavities (Ie & Verdonschot, 1994; the overall weighted effect size was derived from all the studies listed in their Table 1. In each case, the original citations were obtained, and raw effect sizes were calculated from the initial study).	.44	2,870
113. WAIS IQ and obtained level of education (Hanson, Hunsley, & Parker, 1988).	.44	(*k* = 9)
114. MMPI Validity scales and detection of known or suspected malingered psychopathology (data combined from Berry, Baer, & Harris, 1991; and Rogers, Sewell, & Salekin, 1994; the average weighted effect size was calculated from data presented in Tables 1 and 2 of Berry et al. and Table 1 of Rogers et al. for participants presumed or judged to be malingering disturbance).	.45	771
115. D-dimer blood test results and detection of deep vein thrombosis or pulmonary embolism (Becker, Philbrick, Bachhuber, & Humphries, 1996; results are reported for only the 13 [of 29] studies with stronger methodology).	.45	1,652
116. Exercise SPECT imaging and identification of coronary artery disease (Fleischmann, Hunink, Kuntz, & Douglas, 1998; results were estimated from the reported sensitivity and specificity in conjunction with the base rate of coronary artery disease and the total independent *N* across studies).	.46	3,237
117. Antineutrophil cytoplasmic antibody testing and detection of Wegener's granulomatosis (Rao et al., 1995; sensitivity for each study was estimated from their Figure 1).	.47	13,562

(*table continues*)

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
118. Technetium bone scanning results and detection of osteomyelitis (bone infection; Littenberg, Mushlin, & the Diagnostic Technology Assessment Consortium, 1992).	.48	255
119. Clinical examination with routine lab tests and detection of metastatic lung cancer (Silvestri, Littenberg, & Colice, 1995).	.48	1,593
120. Lecithin/sphingomyelin ratio and prediction of neonatal respiratory distress syndrome (Petersen, Smith, Okorodudu, & Bissell, 1996; the most frequently studied predictor test was reported).	.50	1,170
121. Sensitivity of total serum cholesterol levels to changes in dietary cholesterol (Howell, McNamara, Tosca, Smith, & Gaines, 1997).	.50	(*k* = 307)
122. Memory recall tests and differentiation of schizophrenia from controls (Aleman, Hijman, de Haan, & Kahn, 1999; effect size is for studies with demographically matched comparison participants).	.50	2,290
123. CBCL parent report of total problems and psychiatric referral status (receiving treatment vs. not; Achenbach, 1991b; raw data to generate this effect size were obtained from Thomas M. Achenbach [personal communication, February 5, 1999]. Coefficient compares parent ratings of children in treatment to parent ratings of demographically matched control children not receiving treatment).[c]	.51	4,220
124. WAIS IQ subtests and differentiation of dementia from controls (H. Christensen & Mackinnon, 1992; effect computed from data presented in their Tables 1 and 2. The reported *N* is for the largest sample across the individual subtest comparisons).	.52	516
125. Single serum progesterone testing and diagnosis of any nonviable pregnancy (Mol, Lijmer, et al., 1998; following the original authors, we used only the 10 prospective cohort studies listed in their Table II).	.52	3,804
126. MRI results and detection of ruptured silicone gel breast implants (C. M. Goodman, Cohen, Thornby, & Netscher, 1998; these authors found that mammography [*r* = .21, *N* = 381] and ultrasound [*r* = .42, *N* = 541] were less effective than MRI).	.53	382
127. Association of Hachinski ischemic scores with postmortem classification of dementia type (Moroney et al., 1997; effect size computed from their Figure 1 using continuous scores and the Alzheimer's, mixed, and multiinfarct group classifications on a continuum).	.55	312
128. MRI results and detection of lymph node metastases in cervical cancer (Scheidler, Hricak, Yu, Subak, & Segal, 1997; an effect size could not be computed for one study included in this meta-analysis).	.55	817
129. Cognitive tests of information-processing speed and reasoning ability (Verhaeghen & Salthouse, 1997).	.55	4,026
130. MRI results and differentiation of dementia from controls (Zakzanis, 1998; PET and SPECT findings from this meta-analysis were slightly less valid or based on smaller samples, so are not reported. Neuropsychological findings were not used because D. Christensen, Hadzi-Povlovic, & Jacomb, 1991, reported a more extensive meta-analysis).	.57	374
131. WAIS IQ scores and conceptually meaningful criterion measures (K. P. Parker, Hanson, & Hunsley, 1988, Table 2; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999, expressed concern about K. P. Parker et al.'s results because some effect sizes came from unfocused *F* tests [i.e., >1 *df* in the numerator], though the overall effect increases when these results are excluded).	.57	(*K* = 39)
132. Exercise ECG results and identification of coronary artery disease (Fleischmann, Hunink, Kuntz, & Douglas, 1998; results were estimated from the reported sensitivity and specificity in conjunction with the base rate of coronary artery disease and the total independent *N* across studies).	.58	2,637
133. Ultrasound results and identification of deep venous thrombosis (Wells, Lensing, Davidson, Prins, & Hirsh, 1995).	.60	1,616
134. Neuropsychologists' test-based judgments and presence/localization of impairment (Garb & Schramke, 1996; effect size calculated from the accuracy of judgments relative to base rates [see Garb & Schramke, 1996, pp. 143, 144–145]).	.60	1,606
135. Long-term verbal memory tests and differentiation of dementia from depression (H. Christensen, Griffiths, MacKinnon, & Jacomb, 1997; effect data taken from their Table 4).	.61	(*K* = 32)
136. CT results and detection of metastases from head and neck cancer (Merrit, Williams, James, & Porubsky, 1997; *N* was obtained from the original studies).	.64	517

Supporting Table From a Review of Psychological and Medical Test Validity by Meyer et al. (2001) (Continued)

Table 2 (*continued*)

Predictor and criterion (study and notes)	*r*	*N*
137. Neuropsychological tests and differentiation of dementia from controls (D. Christensen, Hadzi-Pavlovic, & Jacomb, 1991; the effect size was derived from studies explicitly stating that dementia had been diagnosed independent of the neuropsychological test results [see D. Christensen et al., 1991, p. 150]).	.68	(*k* = 94)
138. Immunoglobulin-G antiperinuclear factor scores and detection of rheumatoid arthritis (Berthelot, Garnier, Glémarec, & Flipo, 1998).	.68	2,541
139. MMPI Validity scales and detection of malingered psychopathology (primarily analogue studies; data combined from Berry, Baer, & Harris, 1991; and Rogers, Sewell, & Salekin, 1994; average weighted effect size calculated from Tables 1 and 2 of Berry et al. and Table 1 of Rogers et al.).	.74	11,204
140. MMPI basic scales: booklet vs. computerized form (Finger & Ones, 1999; the alternate forms reliability coefficients for each scale were weighted by sample size [ns from 508 to 872], and the average *N* is reported).	.78	732
141. Thoracic impedance scores and criterion measures of cardiac stroke volume and output (Fuller, 1992; only data from methodologically "adequate" studies were included. The mean weighted correlation for each criterion measure was weighted by the number of studies contributing to the mean and then averaged across all criterion measures. Because Fuller [1992, p. 105] cryptically stated that studies were excluded unless there was "concurrence of measurement between the two instruments being compared," it is possible that relevant studies were omitted when the findings did not support the hypothesis).	.81	(*K* = 24)
142. Creatinine clearance test results and kidney function (glomerular filtration rate; Campens & Buntinx, 1997; results for measured and estimated [by the Cockroft–Gault formula] creatinine clearance were pooled. The *N* reported in our table is slightly inflated because it was impossible to identify the specific *n* for two of the studies that used both measures).	.83	2,459
143. Duplex ultrasonography results and identification of peripheral artery disease (de Vries, Hunink, & Polak, 1996; weighted effect size derived from data in their Table 2 using patient samples. The reported *N* refers to the number of observations; some patients were tested multiple times).	.83	4,906
144. Finger or ear pulse oximetry readings in patients and arterial oxygen saturation (L. A. Jensen, Onyskiw, & Prasad, 1998).	.84	4,354

Note. ADHD = attention-deficit hyperactivity disorder; CBCL = Child Behavior Checklist; CFI = Camberwell Family Interview; CT = computed tomography; ECG = electrocardiogram; GPA = grade point average; IQ = intelligence quotient; *k* = number of effect sizes contributing to the mean estimate; *K* = number of studies contributing to the mean estimates; MCMI-II = Millon Clinical Multiaxial Inventory—2nd Edition; MMPI = Minnesota Multiphasic Personality Inventory; MPD = multiple personality disorder; MRI = magnetic resonance imaging; PET = positron emission tomography; PRS = Prognostic Rating Scale; PTSD = posttraumatic stress disorder; SIL = squamous intraepithelial lesions; SPECT = single photon emission computed tomography; TAT = Thematic Apperception Test; WAIS = Wechsler Adult Intelligence Scale; WISC = Wechsler Intelligence Scale for Children; YASR = Young Adult Self-Report.
[a] The actual effect was a statistically nonsignificant value of −.013 (i.e., in the direction of opposite of prediction). [b] Triple marker refers to the joint use of alpha-fetoprotein, human chorionic gonadotropin, and unconjugated estriol. [c] These results are not from meta-analyses and were not identified through our systematic literature search.

From "Psychological Testing and Psychological Assessment: A Review of Evidence and Issues," by G. J. Meyer, S. E. Finn, L. D. Eyde, G. G. Kay, K. L. Moreland, R. R. Dies, et al. 2001, *American Psychologist, 56*, pp. 136–143. Copyright 2001 by the American Psychological Association. Reprinted with permission.

Supporting Tables of Meta-Analytic Evidence from Meyer and Archer (2001)

Table 2
Summary of Meta-Analytic Results Examining the Global Validity of the Rorschach, MMPI, and WAIS

Study and level of aggregation	Description	Effects (k)/ Samples (K)	*N*	Summary mean *r*		
				Rorschach	MMPI	WAIS
Atkinson (1986)	1930–1980, any journal, any Rorschach scale, no method confound	k = 276	?	.36		
Hypothesis level	1960–1980, any journal, any MMPI scale, MMPI not criterion	k = 237	?		.40	
Parker et al. (1988)[a]	1970–1981 in JPA/JCP, 9 Rorschach scales, any criterion, no χ^2	K = 13	872	.37		
Citation level	1970–1981 in JPA/JCP, 14 MMPI scales, any criterion	K = 66	10,776		.43	
	1970–1981 in JPA/JCP, 14 WAIS scales, any criterion	K = 39	5,795			.57
Garb et al. (1998)[b]	Rorschach, same studies as Parker et al. but including χ^2	K = 18	1,302	.29		
Citation level	MMPI, same studies as Parker et al.	K = 66	10,776		.48	
	Rorschach, as above but no method confound, no χ^2	K = 10	656	.36		
	MMPI, as above but no method confound	K = 36	5,640		.55	
Current analysis	1970–1981 in JPA/JCP, any Rorschach scale, any criterion	k = 286	(24,952)	.33		
Hypothesis level	1970–1981 in JPA/JCP, any MMPI scale, any criterion	k = 727	(72,509)		.22	
Hypothesis level	Rorschach, as above but no method confound	k = 247	(22,597)	.27		
	MMPI, as above but no method confound	k = 296	(37,048)		.25	
	1970–1981 in JPA/JCP, any WAIS scale, no method confound	k = 104	(10,122)			.36
Sample level	1970–1981 in JPA/JCP, any Rorschach scale, any criterion	K = 44	4,855	.27		
	1970–1981 in JPA/JCP, any MMPI scale, any criterion	K = 103	15,105		.36	
Sample level	Rorschach, as above but no method confound	K = 43	4,807	.28		
	MMPI, as above but no method confound	K = 58	11,531		.30	
	1970–1981 in JPA/JCP, any WAIS scale, no method confound	K = 25	3,593			.33
Hiller et al. (1999)	1977–1997 in any journal, any Rorschach scale, any criterion	K = 30	1,713	.29		
Citation level	1977–1997 in any journal, any MMPI scale, any criterion	K = 30	4,920		.30	
	Rorschach, as above but no method confound[c]	K = 30	1,713	.29		
	MMPI, as above but no method confound[d]	K = 27	4,454		.29	

Note. *N*s in parentheses are nonindependent totals. MMPI = Minnesota Multiphasic Personality Inventory; WAIS = Wechsler Adult Intelligence Scale; JPA = *Journal of Personality Assessment;* JCP = *Journal of Clinical Psychology.*

[a] *N* obtained from Parker et al.'s (1988) data set. Average effect sizes were computed from their Table 2, using the corrected mean reported in Parker, Hunsley, and Hanson (1999).

[b] *N* was obtained from Parker et al.'s (1988) data set based on study inclusion information provided by Howard N. Garb.

[c] No studies used Rorschach scales as criterion variables, which was the definition of monomethod results for Atkinson (1986), Garb et al. (1998), and our reanalysis of Parker et al. (1988). However, Hiller et al. (1999) conducted an analysis that excluded other "projective" tests as criteria. They found Rorschach validity to be slightly higher than what we report here (r = .30, K = 27, N = 1,509).

[d] Results are from Table 9 in Hiller et al. (1999). We believe two studies should have been excluded from this analysis. If so, the unconfounded validity of the MMPI would be lower (r = .26, K = 25, N = 4,357).

Supporting Tables of Meta-Analytic Evidence from Meyer and Archer (2001) (Continued)

Table 3
Results From Focused Meta-Analyses Comparing the Rorschach to Alternative Predictors of the Same Criterion

Study and criterion/predictor scale	No. of samples	*N*	Mean *r*	
			Rorschach	Other
Bornstein (1999): Observed dependent behavior				
Rorschach Oral Dependence Scale	21	1,320	.37	
TAT Dependency Scale	4	125		.34
Blacky Picture Test Oral Dependence Scale	6	323		.50
MMPI Dependency Scale	5	320		.20
MCMI Dependency Scale	9	720		.17
EPPS Succorance Scale	9	485		.35
IDI Dependency Scale	9	424		.33
Meyer and Handler (1997, 2000) and Meyer (2000): Psychotherapy outcome				
Baseline Rorschach PRS	17	624	.45	
Baseline MMPI Ego Strength Scale	5	280		.02
Baseline IQ	6	246		.15
Incremental validity of Rorschach PRS over IQ	8	290	.36	
Romney (1990): Relatives of schizophrenic patients vs. relatives of controls				
Rorschach Communication Deviance	3	230	.22	
Lovibond Object Sorting Test Thought Processes[a]	5	464		.23
All Non-Rorschach Tests[b] of Thought Processes	11	872		.23
Bornstein (1998b): Physical illness (retrospective designs)				
Rorschach Oral Dependence Scale	2	56	.72	
Dependency by Thematic.Story	4	269		.29
Dependency by DSM Interview	2	200		.09
Dependency by Self-Report Questionnaire	6	539		.18

Note. TAT = Thematic Apperception Test; MMPI = Minnesota Multiphasic Personality Inventory; MCMI = Millon Clinical Multiaxial Inventory; EPPS = Edwards Personal Preference Schedule; IDI = Interpersonal Dependency Inventory; PRS = Prognostic Rating Scale; DSM = *Diagnostic and Statistical Manual of Mental Disorders.*

[a] Romney (1990) reported results for a study he conducted using two predictor variables. We obtained an effect size for just the Lovibond Object Sorting Test based on data reported in Catts, McConaghy, Ward, Fox, and Hadzi-Pavlovic (1993). With slightly different inclusion and exclusion criteria, the meta-analysis by Catts et al. reported nearly identical validity for the Lovibond Test in relation to the same criterion ($r = .24$, $k = 7$, $N = 534$).

[b] Tasks included proverbs, object sorting, verbal associates, repertory grid, the TAT, and observation of structured interactions.

Supporting Tables of Meta-Analytic Evidence from Meyer and Archer (2001) (Continued)

Table 4

Summary Effect Sizes (r) From Focused and General Meta-Analyses Examining the Validity of the Rorschach, MMPI, and IQ Tests

Predictor and criterion	Rorschach	MMPI	IQ	*N*
1. MMPI Ego Strength scores and subsequent psychotherapy outcome		.02		280
2. Unique contribution of an MMPI high point code (vs. other codes) to relevant criteria[a]		.07		8,614
3. MMPI scores and subsequent prison misconduct		.07		17,636
4. MMPI elevations on Scales F, 6, or 8 and criminal defendant incompetency		.08		1,461
5. MMPI Scale 8 and differentiation of schizophrenic vs. depressed disorders		.12		2,435
6. Lower general cognitive ability and involvement in automobile accidents			.12	1,020
7. General intelligence and success in military pilot training			.13	15,403
8. Rorschach DEPI and detection of depressive diagnosis	.14			994
9. MMPI Scale 2 and differentiation of neurotic vs. psychotic disorders		.14		6,156
10. MMPI Scale 8 and differentiation of neurotic vs. psychotic disorders		.14		6,156
11. Baseline IQ and subsequent psychotherapy outcome			.15	246
12. MMPI Cook–Medley Hostility Scale elevations and subsequent death from all causes		.16		4,747
13. MMPI validity scales and detection of known or suspected underreported psychopathology		.18		328
14. MMPI Dependency Scale and dependent behavior		.20		320
15. Rorschach to detect thought disturbance in relatives of schizophrenic patients	.22			230
16. WISC Distractibility subscales and learning disability diagnoses			.24	(K = 54)
17. General intelligence test scores and functional effectiveness across jobs			.25	40,230
18. General validity of Rorschach studies without method confounds	.29			6,520
19. General validity of MMPI studies without method confounds		.29		15,985
20. MMPI Scale 2 and differentiation of schizophrenic vs. depressed disorders		.31		2,435
21. General validity of Rorschach hypotheses without method confounds	.32			(k = 523)
22. General validity of MMPI hypotheses (includes some method confounds)		.32		(k = 533)
23. General validity of WAIS studies without method confounds			.33	3,593
24. MMPI Scale 2 or Depression Scale and detection of depressive diagnosis		.35		2,905
25. Incremental contribution of Rorschach PRS scores over IQ to predict treatment outcome	.36			290
26. General validity of WAIS hypotheses without method confounds			.36	(k = 104)
27. Rorschach Oral Dependence Scale and dependent behavior	.37			1,320
28. MMPI validity scales to detect underreported psychopathology (primarily analog studies)		.39		2,297
29. MMPI Scale 8 and differentiation of psychiatric patients vs. controls		.42		23,747
30. Rorschach SCZI and detection of psychotic diagnosis	.44			717
31. MMPI Scale 2 and differentiation of psychiatric patients vs. controls		.44		23,747
32. WAIS IQ and obtained level of education			.44	(k = 9)
33. Rorschach PRS scores and subsequent psychotherapy outcome	.45			624
34. MMPI validity scales and detection of known or suspected malingered psychopathology		.45		771
35. Rorschach X+% and differentiation of clinical/target group from controls	.46			1,517
36. WAIS IQ subtests and differentiation of dementia from normal controls			.52	516
37. MMPI validity scales and detection of malingered psychopathology (primarily analog studies)		.74		11,204
38. MMPI basic scales: Booklet vs. computerized form		.78		732

Note. Table entries are from Meyer et al. (2001), except as follows: 5, 9, 10, 20, 29, and 31 are from Zalewski (1989); 8 and 30 are from Jørgensen et al. (2000); 11, 14, 15, and 27 are from Table 3; 18, 19, 21–23, and 26 are from Table 2; 35 is from Meyer (2001); and 24 is from Gross, Keyes, and Greene (2000). MMPI = Minnesota Multiphasic Personality Inventory; DEPI = Depression Index; WISC = Wechsler Intelligence Scale for Children; WAIS = Wechsler Adult Intelligence Scale; SCZI = Schizophrenia Index; PRS = Prognostic Rating Scale. K = number of samples; k = number of effects.

[a] The design in this research should produce results more akin to incremental validity than univariate validity.

From "The Hard Science of Rorschach Research: What Do We Know and Where Do We Go?" by G. J. Meyer and R. P. Archer, 2001, *Psychological Assessment, 13*, p. 493. Copyright 2001 by the American Psychological Association. Reprinted with permission.

Supporting Table of Incremental Validity Evidence from Viglione and Hilsenroth's (2001) Structured Review of the 1977–1997 Literature

Table 4
Studies Included in Viglione (1999) With Findings Consistent With Rorschach Incremental Validity

Author and year	Findings
Archer & Gordon (1988)	Optimal overall correct classification (OCC hit rate) of individual inpatients with schizophrenia for Rorschach Schizophrenia (SCZI = 5) = .80. Optimal OCC for Minnesota Multiphasic Personality Inventory (MMPI) *Sc* scale ($Sc \geq 75$) = .76. Utilizing traditional cutoff scores, OCC rates were as follows: SCZI ≥ 4, OCC = .69; $Sc \geq 65$, OCC = .48; and $Sc \geq 70$, OCC = .60.
Archer & Krishnamurthy (1997)	In classification of depression in adolescents with a stepwise discriminant function analysis, Rorschach variables *Vista* and *Afr* added $R^2 = .05$ beyond combined $R^2 = .14$ for MMPI-A scales *A-DEP* and *Ma*. These four variables had highest positive predictive power over any single variable or combination of variables.
Bornstein et al. (1997)	Rorschach Dependency scores significantly correlated with both number of significant interpersonal events ($r = .84$) and impact ratings of these events ($r = -.38$, $p < .01$). Self-report measure of dependency was not significantly related to either ($r = -.11$ and $r = .16$, respectively).
Cooper et al. (1991)	The Rorschach Defense scales provided unique prediction of outcome GAF (Global Assessment of Functioning, Health–Sickness) ratings in regression equations beyond initial GAF and borderline personality self-report scale.
Hilsenroth et al. (1995)	Rorschach variables were able to significantly differentiate ($p = .008$) those patients prematurely terminating from psychotherapy vs. those continuing in treatment, whereas the MMPI-2 was unable to do so ($p = .56$). Specifically Rorschach scores from the interpersonal–relational cluster had a mean effect size (ES) of .57, while the MMPI-2 content scale Negative Treatment Indicators had an ES of −.14.
Holzman et al. (1974)	The classification of a recent schizophrenia diagnosis (hospitalized less than 6 months) and deviant eye tracking was greater (65% accuracy) utilizing Rorschach data alone than a clinical team diagnosis (58%).
O'Connell et al. (1989)	Rorschach data (Thought Disorder Index [TDI]) predicted the development of psychotic and psychotic-like symptoms over a 2–3 year period in a sample of Axis II and affective disorder patients over and above information from clinical interview on lifetime prevalence of psychotic and psychotic-like symptoms (combined $R^2 = .21$, TDI–beta = .32, $p < .03$), schizotypal symptoms (combined $R^2 = .42$, TDI–beta = .32, $p < .009$), or schizotypal and borderline symptoms (combined $R^2 = .51$, TDI–beta = .31, $p < .006$). Initial TDI scores also demonstrated clinical utility in prediction of psychotic and psychotic-like symptoms at follow-up.
Perry & Braff (1994)	Human Experience variable component of the Ego Impairment Index (EII) significantly related to neuropsychological markers of schizophrenia ($r = -.42$, $p < .01$; $r = -.37$, $p < .025$; $r = -.35$, $p < .025$), whereas thought disorder scales based on clinical interview (Schedule for Positive Symptoms and Schedule for Negative Symptoms) were not ($p > .05$).
Perry & Viglione (1991)	Rorschach EII predicted outcome Beck Depression Inventory (BDI; $p < .0002$) and Carroll Rating Scale for Depression ($p < .01$) scores in depressed patients treated with tricyclic antidepressants beyond variance accounted for by gender and baseline scores on BDI and EII. Other demographic variables were also considered but did not affect outcome.
Russ (1980)	Rorschach measures of adaptive regression (*AR*) and defensive effectiveness (*DE*) were significantly related to academic achievement, independent of IQ (*AR*: $r = .45$, $p < .01$; *DE*: $r = .40$, $p < .01$, respectively).
Russ (1981)	Rorschach measure of *AR* was significantly related to reading and overall academic achievement independent of IQ ($r = .51$, $p < .001$, and $r = .47$, $p < .001$). Index AR scores were significantly predictive of reading achievement 1 year later ($r = .29$, $p < .05$).
Shapiro et al. (1990)	Rorschach Depression Index significantly differentiated sexually abused African American girls from controls ($p < .005$). Children's Depression Inventory scores for sample were not significantly different from controls ($p > .05$), consistent with incremental validity of the Rorschach relative to self-reported depression. The groups did differ on the Internalization scale of the Child Behavior Checklist ($p < .0001$).
Skelton et al. (1995)	The dependent variable was a ratio of Rorschach TDI over a TDI derived from the Wechsler. This ratio was 2.46 times higher in a group of 25 identity-disordered adolescents than it was among 35 conduct-disordered and oppositional-defiant adolescents ($p < .01$).

From "The Rorschach: Facts, Fictions, and Future" by D. J. Viglione and M. J. Hilsenroth, 2001, *Psychological Assessment, 13*, p. 458. Copyright 2001 by the American Psychological Association. Reprinted with permission.

Supporting Tables of Meta-Analytic Evidence from Hiller et al. (1999)

Table 4
Meta-Analytic Summary of MMPI and Rorschach Studies

Statistic	MMPI (n = 30)	Rorschach (n = 30)
Central tendency (r)		
Unweighted M	.30	.29
Weighted M	.37	.26
Mdn	.22	.29
Significance		
Stouffer's Z	19.60	9.85
One sample t	5.22	4.70
Variability (r)		
Range	1.02	1.09
s	.26	.26
χ^2 for heterogeneity	630.86	112.68
Confidence interval for r[a]		
95%	.19–.40	.17–.39
99%	.15–.43	.13–.43
99.9%	.11–.46	.09–.46

Note. MMPI = Minnesota Multiphasic Personality Inventory.
[a] Confidence intervals are based on the number of studies, not the number of participants.

From "A Comparative Meta-Analysis of Rorschach and MMPI Validity" by J. B. Hiller, R. Rosenthal, R. F. Bornstein, D. T. R. Berry, and S. Brunell-Neuleib, 1999, *Psychological Assessment, 11*, p. 286. Copyright 1999 by the American Psychological Association. Reprinted with permission.

Table 9
Comparisons Between MMPI and Rorschach Studies

Comparison	MMPI		Rorschach		Unweighted mean effect size		Fixed-effects analysis		Random-effects analysis		
	No. of studies	n	No. of studies	n	MMPI	Rorschach	Z[a]	p[b]	t[a]	p[b]	r[a]
Global	30	4,920	30	1,713	.30	.29	0.30	.76	0.14	.89	.02
Excluding monomethod studies	27	4,454	27	1,509	.29	.30	−0.39	.70	−0.19	.85	−.03
Excluding monomethod studies and psychiatric diagnoses	21	3,785	23	1,306	.26	.32	−1.39	.16	−0.68	.50	−.10
Objective criterion variables only	14	1,281	13	572	.20	.37	−2.68	.007	−1.39	.18	−.27
Psychiatric diagnoses only	6	669	4	203	.37	.18	2.33	.02	1.18	.27	.39
Observer ratings only	4	1,804	4	193	.17	.28	−1.01	.31	−0.78	.46	−.30
Self-report measures only	3	466	6	416	.39	.23	2.67	.008	0.64	.54	.24

Note. MMPI = Minnesota Multiphasic Personality Inventory.
[a] Positive values indicate larger effect sizes for MMPI studies, whereas negative values indicate larger effect sizes for Rorschach studies. [b] Two-tailed.

From "A Comparative Meta-Analysis of Rorschach and MMPI Validity" by J. B. Hiller, R. Rosenthal, R. F. Bornstein, D. T. R. Berry, and S. Brunell-Neuleib, 1999, *Psychological Assessment, 11*, p. 289. Copyright 1999 by the American Psychological Association. Reprinted with permission.

Supporting Table of Meta-Analytic Evidence from Rosenthal et al. (2001)

Table 1
Weighted and Unweighted Mean and Median Effect-Size Correlations for 30 Minnesota Multiphasic Personality Inventory (MMPI) and 30 Rorschach Studies (Panel A), and for Trimmed 24 MMPI and 24 Rorschach Studies (Panel B)

	MMPI			Rorschach		
	M	*Mdn*	Row *M*	*M*	*Mdn*	Row *M*
			Panel A			
Weighted	.37	.35	.36	.26	.29	.28
Unweighted	.30	.22	.26	.28	.28	.28
Column *M*	.34	.29	.31	.27	.29	.28
			Panel B			
Weighted	.33	.35	.34	.24	.26	.25
Unweighted	.28	.22	.25	.24	.28	.26
Column *M*	.31	.29	.30	.24	.27	.26

Note. To address a concern raised by Garb et al. (2001), we reduced the validity coefficient for one Rorschach study from $r = .47$ to $r = .10$. All other Rorschach and MMPI coefficients were identical to those reported.by Hiller et al. (1999).

From "Meta-Analytic Methods, the Rorschach, and the MMPI" by R. Rosenthal, J. B. Hiller, R. F. Bornstein, D. T. R. Berry, and S. Brunell-Neuleib, 2001, *Psychological Assessment, 13*, p. 450. Copyright 2001 by the American Psychological Association. Reprinted with permission.

Board of Trustees
Society for Personality Assessment
6109 H Arlington Boulevard
Falls Church, VA 22044
Email: manager@spaonline.org