Journal of Psychoeducational Assessment

http://jpa.sagepub.com

Performance-Based Assessment in Schools: A Comment on Hojnoski, Morrison, Brown, and Matthews (2006)

Steven R. Smith

Journal of Psychoeducational Assessment 2007; 25; 248 originally published online Jul 12, 2007;
DOI: 10.1177/0734282906297109

The online version of this article can be found at:
http://jpa.sagepub.com/cgi/content/abstract/25/3/248

Published by:
SAGE Publications
http://www.sagepublications.com

Journal of Psychoeducational Assessment
Volume 25 Number 3
September 2007 248-256

10.1177/0734282906297109
http://jpa.sagepub.com
hosted at
http://online.sagepub.com

Performance-Based Assessment in Schools

A Comment on Hojnoski, Morrison, Brown, and Matthews (2006)

Steven R. Smith
University of California, Santa Barbara

This article addresses a 2006 article by Hojnoski, Morrison, Brown, and Matthews on the use of performance-based measurement among school-based practitioners. Their results suggest that many of their survey respondents favor the use of this form of measurement. This line of research is important and addresses an important issue in current clinical practice. However, they offer a critique of this form of assessment, in response to which the author raises four issues. First, there is a difference between *tests* and *techniques*. Second, assessment tools do not make decisions or diagnoses; clinicians do. Third, actuarial prediction and clinical expertise are mutually enhancing. Last, the relationship between science and practice should be bidirectional and integrative. These points are discussed in terms of the utility and appropriateness of performance-based measurement tools and techniques for helping psychologists answer diagnostic, placement, and treatment questions in the school setting.

Keywords: *performance-based tests; projective tests; school-based assessment; child assessment; clinical judgment*

In their article "Projective Test Use Among School Psychologists: A Survey and Critique," Hojnoski, Morrison, Brown, and Matthews (2006) report the results of a survey of school psychologists' use of performance-based (formerly known in the field as *projective*) assessment. Their results reflect that "more than half" (p. 150) of survey respondents find value in a wide range of performance-based assessments, such as sentence completion, drawings, the Rorschach, and story telling. Like prior surveys of clinical child psychologists (Archer & Newsom, 2000; Cashel, 2002) and school psychologists (Shapiro & Heick, 2004), this survey reflects that the use of performance-based measurement is "alive and well" in clinical practice, at least for some referral questions. However, the authors then go on to offer a critique in which they discuss why these measures might continue to be used despite "poor psychometric properties" (p. 155). The authors are to be applauded for addressing an important issue of assessment practice, and their results are illuminating regarding the perceived value of performance-based measurement. Although the survey results are interesting and important, there are some aspects of their critique that bear some discussion. Specifically, there appear to be four areas in which this manuscript might do a disservice to performance-based assessment and to the clinicians who practice this potentially valuable form of clinical work.

Author's Note: Correspondence concerning this article should be addressed to Steven R. Smith, University of California–Santa Barbara, 1100 Phelps Hall, Santa Barbara, CA 93106-9490; e-mail: ssmith@education.ucsb.edu.

Issues Raised

There is a difference between a test and a technique. The first problematic aspect of this article is that the authors have grouped together several different types of measures under the term *projective*. This is misleading because the Rorschach, Thematic Apperception Test (TAT), sentence completion, and drawing techniques are all very different forms of assessment, with varying degrees of empirical support. For example, the Rorschach (Exner, 2003), which can rightfully be referred to as a *test* (because of the standardized administration and scoring techniques and a normative sample), has substantial research support in child and adolescent populations. Indeed, the Rorschach has been found to be sensitive to a number of issues in child and adolescent assessment, including cognitive functioning (Acklin, 1990; Acklin & Fechner-Bates, 1989; Cruz, Brier, & Reznikoff, 1997; Wood, Krishnamurthy, & Archer, 2003), attention deficit/hyperactivity disorder (Bartell & Solanto, 1995), aggression (Liebman, Porcerelli, & Abell, 2005; A. M. Smith, Gacono, & Kaufman, 1997), posttraumatic stress disorder (Holaday, 2000), thought disorder (S. R. Smith, Baity, Knowles, & Hilsenroth, 2001; Stokes, Pogge, Grosso, & Zaccario, 2001), and treatment outcome (Abraham, Lepisto, Lewis, Schultz, & Finkelberg, 1994). Furthermore, the general reliability and validity of this measure has been supported in a number of studies (Exner, 2003; Ganellen, 1996; Gronnerod, 2003; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; Hilsenroth & Stricker, 2004; Meyer, 2000; Meyer & Archer, 2001; Meyer et al., 2002; Meyer, Mihura, & Smith, 2005; Parker, Hanson, & Hunsley, 1988; Society for Personality Assessment, 2005; Viglione, 1999; Viglione & Hilsenroth, 2001; Viglione & Taylor, 2003; Wagner, Alexander, Roos, & Adair, 1986; Weiner, 1996, 1997; Wood, Nezworski, & Garb, 2003). The Hand Test (Wagner, 1983; Wagner, Rasch, & Marsico, 1991) is also a test with robust psychometric properties and empirical support with child and adolescent populations (Clemence, Hilsenroth, Sivec, & Rasch, 1999; Clemence, Hilsenroth, Sivec, Rasch, & Waehler, 1998; S. R. Smith, Blais, Vangala, & Masek, 2005). Although there are some who question the validity of the Rorschach and other performance-based measures (Garb, 1999; Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Lilienfeld, Wood, & Garb, 2000; Wood & Lilienfeld, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001; Wood, Nezworski, Lilienfeld, & Garb, 2003), the psychometrics of the Rorschach allow it to be subjected to scientific inquiry, and as is the case for all tests, ongoing evaluation is needed to determine not whether such tests are valid but for what.

In contrast, story telling, sentence completion, and drawings are best thought of as *techniques*. As the authors point out, there are standardized scoring systems available for all of these measures. For example, the Social Cognition and Object Relations Scale (Westen, 1995) method of scoring the TAT and other narratives has garnered some empirical support (Ackerman, Clemence, Weatherill, & Hilsenroth, 1999; Ackerman, Hilsenroth, Clemence, Weatherill, & Fowler, 2001; Niec & Russ, 2002), and the Roberts-2 (Roberts & Gruber, 2005) has a very elaborate scoring system. However, as noted in the article (p. 155), some clinicians may use these techniques to generate clinical hypotheses without benefit of standardized administration and scoring. When used in this manner, these tools might not represent tests, per se. However, in the same manner, most clinicians conduct unstructured diagnostic interviews as opposed to psychometrically sound structured interviews of the

clients and students they see. We would not assume that this practice represents a substandard level of care. An assessment battery with various tests and techniques interpreted in the context of client history will be far more valid than one relying on only one form of measurement (Meyer, 2002). The authors' lumping together all of these varied measures under one umbrella of *projectives* "with poor psychometric properties" throws the proverbial baby out with the proverbial bathwater.

The purpose of performance-based assessment is to assess personality. The authors note that many if not most of the respondents use performance-based assessment to help make decisions regarding educational eligibility. They also correctly note that these tests were not designed for this purpose, nor is there a robust literature base that supports using these particular tests in this manner. I would never suggest that tests be used for reasons for which they have shown to be invalid; such practice is potentially unethical. However, other popular self-report measures such as the Minnesota Multiphasic Personality Inventory–Adolescent (Butcher et al., 1992) have not been explicitly validated for this purpose either; performance-based measures are no different in this respect, and it is questionable why they might be singled out in this way. Furthermore, it is important to note that no single test can be used to make important decisions regarding placement, treatment, or diagnosis and that relying on only one or a few forms of data is also a potentially unethical practice. However, the important issue here is that treatment and placement decisions are made by clinicians, not tests. Although tests can aid clinicians in this decision-making process, trained clinicians armed with test data, behavioral observations, history, medical information, and input from parents and teachers are the final decision makers. To suggest otherwise reduces clinicians to the role of technicians who merely administer and score tests. In essence, this is the difference between psychological testing and psychological assessment noted by Handler and Meyer (1998).

Second and more important, the survey did not ask clinicians to describe the test battery that they use to make placement and treatment decisions. Although the clinicians acknowledge using performance-based measurement in this process, we can assume that they use multiple tests (including behavior-rating scales, self-report, and performance-based measures) that provide information about different and important aspects of client functioning. Meyer (2002) cogently argues that multiple tests provide important nonoverlapping pieces of data that increase the validity of clinical decision making. If all tests provide the same type of information in the same form, these data are likely to be highly consistent and reliable, but the breadth of information provided will be substantially narrowed. The present survey is important because it hints that clinicians likely rely on a breadth of test, history, and contextual data to make important decisions regarding placement and treatment. In a psychological assessment battery, for some referral reasons and some clients, performance-based assessments have their place. Unless it can be demonstrated that clinicians using a full battery of tests (including performance-based measures) provide less reliable or less valid decisions regarding placement and treatment, it does not seem wise to remove a tool from their toolboxes.

Last, it is valuable to discuss eligibility issues in the context of the Individuals with Disabilities Education Act, the federal law that mandates appropriate educational accommodations for children with handicapping conditions. These conditions can be either physical or mental in nature, with the latter including learning disabilities, mental retardation,

and emotional disturbance. Certainly, most personality assessment measures may not be useful in determining educational eligibility based on the presence of a learning disability. However, when determining the presence of an emotional disturbance, such tools may likely be useful components of such an evaluation. In the same way that a hammer is not useful for driving a screw, the choice of a particular tool should be contingent on the job at hand. From the description of the survey, it appears that the authors asked respondents about which measures they used in the "determination of eligibility for services" in addition to "diagnostic or initial assessment," "treatment planning," "reevaluation," (p. 148), and so on. If we assume that these diagnoses, determinations, treatment plans, and reevaluations are related to the presence or absence of emotional disturbance, then this suggests that the inclusion of personality assessment measures might be appropriate; if, however, they relate more to the presence or absence of a learning disability or mental retardation, then the use of personality assessment as a primary assessment source might be more problematic. Because it is not obvious from the question asked, we cannot know to what degree these measures are used appropriately for the purposes for which they are intended.

Actuarial prediction and clinical expertise are mutually enhancing. In attempting to describe why school psychologists might continue to use projective assessments, the authors cite the work of Meehl and others on statistical prediction versus clinical judgment. As the authors correctly state, Meehl and others note that statistical judgment and prediction can be more robust than informal and subjective prediction (Dawes, Faust, & Meehl, 1989; Grove & Lloyd, 2006; Grove & Meehl, 1996; Meehl, 1954, 1959, 1967). However, as a psychoanalyst, Meehl also understood that there is value to clinical expertise and that the goals of diagnostic prediction may not be the same as the goals of psychological treatment. Westen and Weinberger (2004, 2005) have written convincingly about the difference between clinical description and statistical prediction. They argue that consistent with Meehl's beliefs, clinical expertise and clinical judgment are far more robust when judgments are made at moderate (rather than high) levels of inference, the clinician has some degree of expertise borne of training and experience, and clinicians make use of tools or conditions that optimize that expertise (Westen & Weinberger, 2005). The combination of clinical experience and test data reduces the error of uninformed clinical judgment. Therefore, clinicians who ethically rely on tests and tools become more experienced, leading to well-informed judgments and predictions.

Indeed, a recent survey of the Fellows of the American Association of School Psychology found that clinical judgment (based on an integration of quantitative and qualitative data) was deemed the most important criterion for determining the presence of a learning disability in children (Schrank, Miller, Caterino, & Desrochers, 2006). The importance of blending clinical and actuarial methods has been noted in medical decision making as well (Tonelli, 2006). Certainly, all tests that are improperly or carelessly used will lead to erroneous conclusions that can have negative consequences for clients; this is as true for the Weschler Intelligence Scale for Children–Fourth Edition (Wechsler, 2004) as it is for the Rorschach. Meehl was not opposed to the notion of clinical expertise but rather to the sloppy application of empirically or theoretically unsupported rules and lay interpretation (in both academia and clinical practice). It casts quite a negative net to suggest that

school psychologists are so lacking in both expertise in human behavior and experience with assessment that their choice of certain tests and techniques may be unethical.

The relationship between science and practice should be bidirectional and integrative. The authors also posit that the continued use of performance-based assessment may reflect "a lack of knowledge, inadequate professional training . . . , and habit" (p. 156). They go on to speculate that because many of their survey respondents were trained a decade or more ago, when "the use of projectives was more of a focus than it is currently" (p. 156), they might cling to these assessments. It is true that test-based assessment is less focal in graduate training than it was in decades past. But this reduction is generally cast as a bad development among practitioners and has resulted in a significant rift between academic training programs and the demands of the clinical marketplace (Clemence & Handler, 2001). Results of surveys of predoctoral internship directors consistently suggest that they see personality-assessment skills as vital components of professional practice in psychology (Stedman, Hatch, & Schoenfeld, 2000). Surveys also suggest that internship directors find that many if not most of their trainees are inadequately trained in assessment (Stedman, Hatch, & Schoenfeld, 2001b, 2002). Particularly important to many of these survey respondents is experience with performance-based techniques, including the Rorschach and TAT (Durand, Blanchard, & Mindell, 1988; Stedman et al., 2000; Stedman et al., 2001b; Stedman, Hatch, & Schoenfeld, 2001a; Watkins, 1991).

However, the authors make the assumption that that which is taught in academia should be held in higher esteem than those techniques and tools that practicing clinicians find to be most helpful in their work day after day. It seems that this discussion would be better advanced by acknowledging the clinical utility of these measures and respecting the experience of clinicians. Certainly, clinicians need to be fully versed in the research literature related to the measures they use, but it is important not only that clinicians alter their practices based on the work of academics but also that training programs teach the methods that are the most clinically useful. The tools of science should be used to understand the circumstances under which these measures might be the most useful. The relationship between science and practice should be bidirectional, and the wisdom of both should be respected.

It should be noted that the rift between academia and practice may give rise to an important ethical dilemma. That is, if assessment trainers believe that a particular form of assessment is empirically unsupported, they might feel that the teaching of these methods supports unethical practice. The ethics code of the American Psychological Association (2002) states that psychologists should refrain from using tests that are outdated or obsolete for the particular assessment question. This is the case for all assessments, regardless of purpose or technique, and certainly, instructors who believe that teaching performance-based tests represents the promotion of obsolete tests should be concerned about their ethical obligations (as should we all). For performance-based tests (including the Rorschach and Hand Test, among others), there is substantial evidence of utility for some assessment questions, whereas for performance-based techniques, the same quantity of evidence may be lacking. However, to quote the old saying, absence of evidence is not evidence of absence. Therefore, it seems fair for assessment instructors to teach the strengths, limitations, and particular uses of those measures that have been found to be useful in clinical settings but at the same time to acknowledge that much more

work needs to be done to examine when such measures may and may not be the most useful. To provide a balanced exposure to those measures that seem to be supported by ample clinical evidence in the absence of research evidence seems both ethically appropriate and responsible from a student-training and student-preparation perspective.

Conclusion

In sum, the article by Hojnoski et al. addresses an important aspect of clinical practice. Their work suggests that performance-based measurement continues to be held in high esteem among school psychologists. Such a finding is not surprising given that these techniques are among those most commonly used in child clinical and school settings (Archer & Newsom, 2000; Cashel, 2002; Shapiro & Heick, 2004; Yalof & Abraham, 2007). Yet Hojnoski et al. suggest that their results indicate that practitioners need additional education about the limits of performance-based assessment because their use of performance-based assessment represents scientifically unsupported practice. They take this position despite the fact that a review of the extant literature on a number of performance-based tests suggests considerable reliability and validity. A more accurate and informed position is that both academics and practitioners should be fully educated about and aware of the strengths and limitations of all forms of assessment and should not either blindly dismiss or embrace a particular test or technique. Furthermore, the results of this study suggest that academic training programs should provide more, not less, training in personality assessment techniques because these measures clearly are useful in clinical practice.

References

Abraham, P. P., Lepisto, B. L., Lewis, M. G., Schultz, L., & Finkelberg, S. (1994). An outcome study: Changes in Rorschach variables of adolescents in residential treatment. *Journal of Personality Assessment, 62*, 505-514.

Ackerman, S. J., Clemence, A. J., Weatherill, R., & Hilsenroth, M. J. (1999). Use of the TAT in the assessment of DSM-IV Cluster B personality disorders. *Journal of Personality Assessment, 73*, 422-442.

Ackerman, S. J., Hilsenroth, M. J., Clemence, A. J., Weatherill, R., & Fowler, J. C. (2001). Convergent validity of Rorschach and TAT scales of object relations. *Journal of Personality Assessment, 77*, 295-306.

Acklin, M. W. (1990). Personality dimensions in two types of learning-disabled children: A Rorschach study. *Journal of Personality Assessment, 54*, 67-77.

Acklin, M. W., & Fechner-Bates, S. (1989). Rorschach developmental quality and intelligence factors. *Journal of Personality Assessment, 53*, 537-545.

American Psychological Association. (2002). *Ethical principals of psychologists and code of conduct*. Washington, DC: Author.

Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment, 7*, 227-235.

Bartell, S. S., & Solanto, M. V. (1995). Usefulness of the Rorschach Inkblot Test in assessment of attention deficit hyperactivity disorder. *Perceptual and Motor Skills, 80*, 531-541.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., et al. (1992). *Minnesota Multiphasic Personality Inventory–Adolescent (MMPI-A): Manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practice and the impact of managed care. *Professional Psychology: Research and Practice, 33*, 446-453.

Clemence, A. J., & Handler, L. (2001). Psychological assessment on internship: A survey of training directors and their expectations for students. *Journal of Personality Assessment, 76*, 18-47.

Clemence, A. J., Hilsenroth, M. J., Sivec, H. J., & Rasch, M. A. (1999). Hand Test AGG and AOS variables: Relation with teacher rating of aggressiveness. *Journal of Personality Assessment, 73*, 334-344.

Clemence, A. J., Hilsenroth, M. J., Sivec, H. J., Rasch, M. A., & Waehler, C. A. (1998). Use of the Hand Test in the classification of psychiatric inpatient adolescents. *Journal of Personality Assessment, 71*, 228-241.

Cruz, E. B., Brier, N. M., & Reznikoff, M. (1997). An examination of the relationship between form level rating on the Rorschach and learning disability status. *Journal of Psychology: Interdisciplinary and Applied, 131*, 167-174.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

Durand, V. M., Blanchard, E. B., & Mindell, J. A. (1988). Training in projective testing: Survey of clinical training directors and internship directors. *Professional Psychology: Research and Practice, 19*, 236-238.

Exner, J. E. (2003). *The Rorschach: A comprehensive system* (4th ed.). New York: John Wiley.

Ganellen, R. J. (1996). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: A review. *Journal of Personality Assessment, 67*, 219-243.

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.

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

Grove, W. M., & Lloyd, M. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology, 115*, 192-194.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, & Law, 2*, 293-323.

Handler, L., & Meyer, G. J. (1998). The importance of teaching and learning personality assessment. In L. Handler & M. J. Hilsenroth (Eds.), *Teaching and learning personality assessment* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum.

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*(2), 141-152.

Hojnoski, R. L., Morrison, R., Brown, M. C., & Matthews, W. J. (2006). Projective test use among school psychologists: A survey and critique. *Journal of Psychoeducational Assessment, 24*, 145-159.

Holaday, M. (2000). Rorschach protocols from children and adolescents diagnosed with posttraumatic stress disorder. *Journal of Personality Assessment, 75*, 143-157.

Liebman, S. J., Porcerelli, J., & Abell, S. C. (2005). Reliability and validity of Rorschach aggression variables with a sample of adjudicated adolescents. *Journal of Personality Assessment, 81*, 33-39.

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.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Lanham, MD: Rowan & Littlefield.

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology, 13*, 106-128.

Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 594-599). New York: McGraw-Hill.

Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment, 75*, 46-81.

Meyer, G. J. (2002). Implications of information-gathering methods for a refined taxonomy of psychopathology. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 69-105). Washington, DC: American Psychological Association.

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., 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.

Niec, L. N., & Russ, S. W. (2002). Children's internal representations, empathy and fantasy play: A validity study of the SCORS-Q. *Psychological Assessment, 14*, 331-338.

Parker, K. C., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103*, 367-373.

Roberts, G. E., & Gruber, C. P. (2005). *Roberts-2 manual*. Los Angeles: Western Psychological Services.

Schrank, F. A., Miller, J. A., Caterino, L. C., & Desrochers, J. (2006). American Academy of School Psychology survey on the independent educational evaluation for a specific learning disability: Results and discussion. *Psychology in the Schools, 43*, 771-780.

Shapiro, E. S., & Heick, P. E. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools, 41*, 551-561.

Smith, A. M., Gacono, C. B., & Kaufman, L. (1997). A Rorschach comparison of psychopathic and nonpsychopathic conduct disordered adolescents. *Journal of Clinical Psychology, 53*, 289-300.

Smith, S. R., Baity, M. R., Knowles, E. S., & Hilsenroth, M. J. (2001). Assessment of disordered thinking in children and adolescents: The Rorschach perceptual-thinking index. *Journal of Personality Assessment, 77*, 447-463.

Smith, S. R., Blais, M. A., Vangala, M., & Masek, B. J. (2005). Exploring the Hand Test with medically ill children and adolescents. *Journal of Personality Assessment, 85*, 82-91.

Society for Personality Assessment. (2005). The status of the Rorschach in clinical and forensic practice: An official statement by the board of trustees of the Society for Personality Assessment. *Journal of Personality Assessment, 85*, 219-237.

Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2000). Preinternship preparation in psychological testing and psychotherapy: What internship directors say they expect. *Professional Psychology: Research and Practice, 31*, 321-326.

Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2001a). The current status of psychological assessment training in graduate and professional schools. *Journal of Personality Assessment, 77*, 398-407.

Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2001b). Internship directors' valuation of preinternship preparation in test-based assessment and psychotherapy. *Professional Psychology: Research and Practice, 32*, 421-424.

Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2002). Preinternship preparation of clinical and counseling students in psychological testing, psychotherapy, and supervision: Their readiness for medical school and nonmedical school internships. *Journal of Clinical Psychology in Medical Settings, 9*, 267-271.

Stokes, J. M., Pogge, D. L., Grosso, C., & Zaccario, M. (2001). The relationship of the Rorschach schizophrenia index to psychotic features in a child psychiatric sample. *Journal of Personality Assessment, 76*, 209-228.

Tonelli, M. R. (2006). Integrating evidence into clinical practice: An alternative to evidence-based approaches. *Journal of Evaluation in Clinical Practice, 12*, 248-256.

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 13*, 452-471.

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

Viglione, D. J., & Taylor, N. (2003). Empirical support for interrater reliability of Rorschach Comprehensive System scoring. *Journal of Clinical Psychology, 59*, 111-121.

Wagner, E. E. (1983). *The Hand Test manual* (Rev. ed.). Los Angeles: Western Psychological Services.

Wagner, E. E., Alexander, R. A., Roos, G., & Adair, H. (1986). Optimum split-half reliabilities for the Rorschach: Projective techniques are more reliable than we think. *Journal of Personality Assessment, 50*, 107-112.

Wagner, E. E., Rasch, M. A., & Marsico, D. S. (1991). *Hand Test manual supplement: Interpreting child and adolescent responses*. Los Angeles: Western Psychological Services.

Watkins, C. E. (1991). What have surveys taught us about the teaching and practice of psychological assessment? *Journal of Personality Assessment, 56*, 426-437.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children manual* (4th ed.). San Antonio: Psychological Corporation.

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

Weiner, I. B. (1997). Current status of the Rorschach Inkblot method. *Journal of Personality Assessment, 68*, 5-19.

Westen, D. (1995). *Social cognition and object relations scale: Q-sort for projective stories (SCORS-Q)*. Unpublished manuscript, Cambridge Hospital and Harvard Medical School, Cambridge, MA.

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist*, *59*, 595-613.

Westen, D., & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology, 61*, 1257-1276.

Wood, J. M., Krishnamurthy, R., & Archer, R. P. (2003). Three factors of the comprehensive system for the Rorschach and their relationship to Wechsler IQ scores in an adolescent sample. *Assessment, 10*, 259-265.

Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment, 6*, 341-349.

Wood, J. M., Nezworski, M. T., & Garb, H. N. (2003). What's right with the Rorschach? *Scientific Review of Mental Health Practice, 2*, 142-146.

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

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

Yalof, J., & Abraham, P. P. (2007). Personality assessment in schools. In S. R. Smith & L. Handler (Eds.), *The clinical assessment of children and adolescents: A practitioner's handbook* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum.