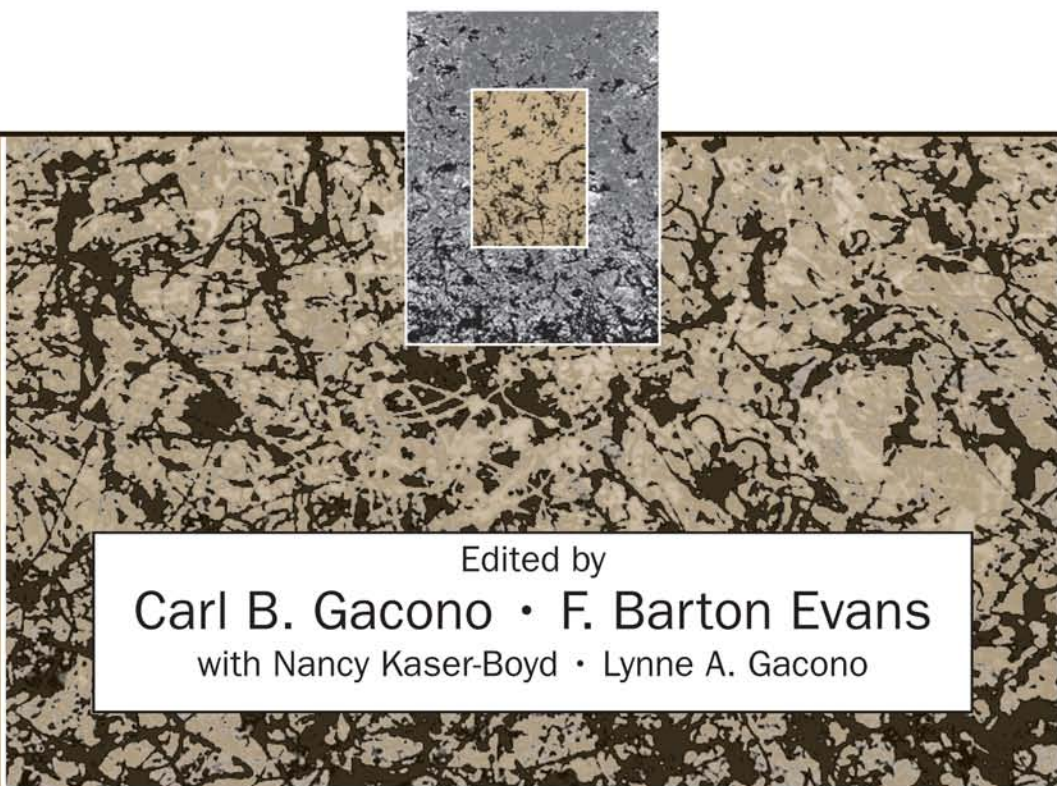




The Handbook of
**Forensic Rorschach
Assessment**



Edited by
Carl B. Gacono • F. Barton Evans
with Nancy Kaser-Boyd • Lynne A. Gacono

The Handbook of
**Forensic Rorschach
Assessment**

The LEA Series in Personality and Clinical Psychology

Irving B. Weiner, Editor

-
- Bornstein/Masling (Eds.)** • *Scoring the Rorschach: Seven Validated Systems*
- Calhoun/Tedeschi** • *Facilitating Posttraumatic Growth: A Clinician's Guide*
- Dana (Ed.)** • *Handbook of Cross-Cultural and Multicultural Personality Assessment*
- Eisen/Quas/Goodman (Eds.)** • *Memory and Suggestibility in the Forensic Interview*
- Exner (Ed.)** • *Issues and Methods in Rorschach Research*
- Frederick/McNeal** • *Inner Strengths: Contemporary Psychotherapy and Hypnosis for Ego-Strengthening*
- Gacono (Ed.)** • *The Clinical and Forensic Assessment of Psychopathy: A Practitioner's Guide*
- Gacono/Meloy** • *The Rorschach Assessment of Aggressive and Psychopathic Personalities*
- Ganellen** • *Integrating the Rorschach and the MMPI-2 in Personality Assessment*
- Goodman/Rudorfer/Maser (Eds.)** • *Obsessive-Compulsive Disorder: Contemporary Issues in Treatment*
- Hall/Bernieri (Eds.)** • *Interpersonal Sensitivity: Theory and Measurement*
- Handler/Hilsenroth (Eds.)** • *Teaching and Learning Personality Assessment*
- Hy/Loevinger** • *Measuring Ego Development, Second Edition*
- Kelly** • *The Assessment of Object Relations Phenomena in Adolescents: TAT and Rorschach Measures*
- Kelly** • *The Psychological Assessment of Abused and Traumatized Children*
- Kohnstamm/Halverson/Mervielde/Havill (Eds.)** • *Parental Descriptions of Child Personality: Developmental Antecedents of the Big Five?*
- Loevinger (Ed.)** • *Technical Foundations for Measuring Ego Development: The Washington University Sentence Completion Test*
- McCallum/Piper (Eds.)** • *Psychological Mindedness: A Contemporary Understanding*
- Meloy/Acklin/Gacono/Murray/Peterson (Eds.)** • *Contemporary Rorschach Interpretation*
- Needleman** • *Cognitive Case Conceptualization: A Guidebook for Practitioners*
- Nielsen/Johnson/Albert Ellis Institute** • *Counseling and Psychotherapy With Religious Persons: A Rational Emotive Behavior Therapy Approach*
- Nolen-Hoeksema/Larson** • *Coping With Loss*
- Rosowsky/Abrams/Zwiegl (Eds.)** • *Personality Disorders in the Elderly: Emerging Issues in Diagnosis and Treatment*
- Sarason/Pierce/Sarason (Eds.)** • *Cognitive Interference: Theories, Methods, and Findings*
- Silverstein** • *Self Psychology and Diagnostic Assessment: Identifying Selfobject Functions Through Psychological Testing*
- Taylor (Ed.)** • *Anxiety Sensitivity: Theory, Research, and Treatment of the Fear of Anxiety*
- Tedeschi/Park/Calhoun (Eds.)** • *Posttraumatic Growth: Positive Changes in the Aftermath of Crisis*
- Van Hasselt/Hersen (Eds.)** • *Handbook of Psychological Treatment Protocols for Children and Adolescents*
- Weiner** • *Principles of Rorschach Interpretation*
- Wong/Fry (Eds.)** • *The Human Quest for Meaning: A Handbook of Psychological Research and Clinical Applications*
- Zillmer/Harrower/Ritzler/Archer** • *The Quest for the Nazi Personality: A Psychological Investigation of Nazi War Criminals*

The Handbook of
**Forensic Rorschach
Assessment**

Edited by

Carl B. Gacono • F. Barton Evans

with Nancy Kaser-Boyd • Lynne A. Gacono

Cover design by Kathryn Houghtaling.

Lawrence Erlbaum Associates
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Lawrence Erlbaum Associates
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

© 2008 by Taylor & Francis Group, LLC

Lawrence Erlbaum Associates is an imprint of Taylor & Francis Group, an Informa business

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8058-5823-5 (Hardcover)

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

*This book is dedicated to
Paul Lerner and John Exner, Jr.*



Portrait of Hermann Rorschach
by Carl B. Gacono, PhD

CONTENTS

PREFACE	XI
<i>Carl B. Gacono and F. Barton Evans</i>	
ACKNOWLEDGMENTS	XXI
CONTRIBUTORS	XXIII
ABOUT THE EDITORS	XXV
PART I: SCIENTIFIC AND LEGAL FOUNDATIONS	
1 ESSENTIAL ISSUES IN THE FORENSIC USE OF THE RORSCHACH	3
<i>Carl B. Gacono, F. Barton Evans, and Donald J. Viglione</i>	
2 AN OVERVIEW OF RORSCHACH PSYCHOMETRICS FOR FORENSIC PRACTICE	21
<i>Donald J. Viglione and Greg J. Meyer</i>	
3 ADMISSIBILITY OF THE RORSCHACH	55
<i>Joseph T. McCann and F. Barton Evans</i>	
4 THE AUTHORITY OF THE RORSCHACH: AN UPDATE	79
<i>J. Reid Meloy</i>	
5 RORSCHACH ASSESSMENT OF MALINGERING AND DEFENSIVE RESPONSE SETS	89
<i>Ronald J. Ganellen</i>	
6 PRESENTING AND DEFENDING RORSCHACH TESTIMONY	121
<i>Irving B. Weiner</i>	

PART II: FORENSIC APPLICATIONS

- 7 THE USE OF THE RORSCHACH INKBLOT METHOD IN TRIAL COMPETENCY EVALUATIONS 141
B. Thomas Gray and Marvin W. Acklin
- 8 THE RORSCHACH TEST AND FORENSIC PSYCHOLOGICAL EVALUATION: PSYCHOSIS AND THE INSANITY DEFENSE 157
Marvin W. Acklin
- 9 DANGEROUSNESS RISK ASSESSMENT 175
B. Thomas Gray, J. Reid Meloy, and Michael T. Jumes
- 10 DEATH PENALTY AND MITIGATION 195
Nancy Kaser-Boyd
- 11 USE OF THE RORSCHACH IN FORENSIC TREATMENT PLANNING 211
Carl B. Gacono, Michael T. Jumes, and B. Thomas Gray
- 12 THE RORSCHACH IN CHILD CUSTODY AND PARENTING PLAN EVALUATIONS: A NEW CONCEPTUALIZATION 233
F. Barton Evans and Benjamin M. Schutz
- 13 RORSCHACH ASSESSMENT OF PSYCHOLOGICAL TRAUMA 255
Nancy Kaser-Boyd and F. Barton Evans
- 14 RORSCHACH ASSESSMENT IN TORT AND EMPLOYMENT LITIGATION 279
Bruce L. Smith
- 15 THE RORSCHACH IN FITNESS FOR DUTY EVALUATIONS 301
James F. Gormally

PART III: FORENSIC REFERENCE GROUPS

- 16 THE RORSCHACH AND ANTISOCIAL PERSONALITY DISORDER 323
Carl B. Gacono, Lynne A. Gacono, and F. Barton Evans
- 17 A RORSCHACH UNDERSTANDING OF ANTISOCIAL AND PSYCHOPATHIC WOMEN 361
Ted B. Cunliffe and Carl B. Gacono

18	A RORSCHACH UNDERSTANDING OF PSYCHOPATHS, SEXUAL HOMICIDE PERPETRATORS, AND NONVIOLENT PEDOPHILES <i>Carl B. Gacono, J. Reid Meloy, and Michael R. Bridges</i>	379
19	INMATES IN PRISON PSYCHIATRIC TREATMENT: A MULTIMETHOD DESCRIPTION <i>Myla H. Young, Philip S. Erdberg, and Jerald Justice</i>	403
20	SOME CONSIDERATIONS FOR THE RORSCHACH ASSESSMENT OF FORENSIC PSYCHIATRIC OUTPATIENTS <i>Lynne A. Gacono and Carl B. Gacono</i>	421
21	CHILD CUSTODY LITIGANTS: RORSCHACH DATA FROM A LARGE SAMPLE <i>Jacqueline Singer, Carl F. Hoppe, S. Margaret Lee, Nancy W. Olesen, and Marjorie G. Walters</i>	445
PART IV: SPECIAL TOPICS		
22	BATTERED WOMAN SYNDROME: ASSESSMENT-BASED EXPERT TESTIMONY <i>Nancy Kaser-Boyd</i>	467
23	THE RORSCHACH AND IMMIGRATION EVALUATIONS <i>F. Barton Evans</i>	489
24	THE USE OF THE RORSCHACH IN PROFESSIONAL FITNESS TO PRACTICE EVALUATIONS <i>Scott C. Stacey, Peter Graham, and George I. Athey</i>	505
25	THE USE OF THE RORSCHACH IN POLICE PSYCHOLOGY: SOME PRELIMINARY THOUGHTS <i>Peter A. Weiss, William U. Weiss, and Carl B. Gacono</i>	527
26	THE RORSCHACH ASSESSMENT OF AGGRESSION: THE RORSCHACH EXTENDED AGGRESSION SCORES <i>Carl B. Gacono, Lynne A. Gacono, J. Reid Meloy, and Matthew R. Baity</i>	543
27	MULTIMETHOD ASSESSMENT AS A FORENSIC STANDARD <i>Philip S. Erdberg</i>	561

APPENDIX A: THREE CRITIQUES OF "WHAT'S WRONG WITH THE RORSCHACH?" <i>Hale Martin, Carl B. Gacono, F. Barton Evans, and J. Reid Meloy</i>	567
AUTHOR INDEX	579
SUBJECT INDEX	597

PREFACE

Carl B. Gacono

Private Practice, Austin, TX

Barton Evans

Private Practice, Bozeman, MT

The Rorschach Inkblot Method (RIM) has a long and noble tradition within the field of personality assessment. The number of RIM research studies and scholarly citations, second only to the MMPI/MMPI-2, speak to the amount of interest in the test's usage. The development of the Comprehensive System (Exner, 1997, 2003) has anchored the RIM within the mainstream of empirical personality assessment instruments, making it acceptable to use in forensic assessment (Gacono, Evans, & Viglione, 2002).

The RIM provides an open-structured, performance-based cognitive perceptual problem-solving task that is quite different from self-report measures. As research has demonstrated (Ganellen, 1994, 1996), it is difficult to manipulate by conscious effort to underreport or overreport psychological difficulties. It is this unique element, which adds to the RIM's value in forensic assessment (Gacono, Evans, & Viglione, 2002; Ganellen, 1994, 1996; Ganellen, Wasyliw, Haywood, & Grossman, 1996; Grossman, Wasyliw, Benn, & Gyoerkoe, 2002). The refinement of a variety of relevant forensic groups (Gacono & Meloy, 1994; see chaps. 16–21, this vol.) has increased its usefulness as a forensic assessment tool.

Concurrent to these psychometric advances for the Rorschach,¹ psychology has witnessed a disturbing increase in psychological journals offering “pseudo-debates” concerning the relevance of various assessment methods or psychological tests to clinical and forensic practice (Gacono, Loving, & Bodholdt, 2001; Meyer, 1999, 2000; Weiner, 2001). We use the word “disturbing” because unqualified individuals, often academics who do not practice psychological assessment, have elevated themselves to a seemingly

¹The position paper of the Board of the Society for Personality Assessment (2005), also known as the Rorschach White Paper, in which they carefully reviewed the scientific literature on the Rorschach, concluded, “This statement affirms that the Rorschach possesses reliability and validity similar to that of the generally accepted personality assessment instruments and it's responsible use in personality assessment is appropriate and justified.”

expert role through their association in the literature with legitimate experts in psychological assessment. As Weiner (2001, p. 7) stated, “We’ve got some people who have come along and are raising criticisms ... who have never published any Rorschach research of their own and know very little about how to use the Rorschach in practice. They seem to be on some kind of crusade to bad-mouth the instrument.”

Dedicated researchers and practitioners have responded by producing a substantial body of new validating research, as well as bringing a wealth of clinical/forensic experience about the inestimable value of the RIM in delivering mental health and forensic services as diagnostic consultants (Gacono, Evans, Jumes, & Loving, 2002; Meyer, 2000; Wiener, 2001). Persistent detractors have seldom demonstrated the same level of scientific rigor by weighing all available evidence, discriminating between compelling and questionable research findings, and drawing conclusions on the basis of a balanced and open-minded determination of where the facts lie (Gacono & Evans, 2004; Wiener, 2001). Rather, their rhetoric and tactics have been likened to advocacy. “An irony in this situation is the fact that contemporary Rorschach critics, while waving the banner of scientific legitimacy, are pursuing slash-and-burn tactics that have far more in common with advocacy than with science” (Weiner, 2001, p. 7). The result is that these “pseudo-debates” have an “armchair” quality about them (Hare, 1998, p.188)² that does little to further scientific exploration (Meyer, 2000).

The lack of applied knowledge on the part of assessment detractors, who are often not qualified as clinical or forensic experts, has resulted in a distorted picture of the issues. Their flawed or superficial understanding of essential applied theoretical and methodological issues directly impacts the manner in which the “debated” issues are presented (Gacono & Evans, 2004). More akin to politics than science, the authors demonstrate a very selective inattention to the literature, ignoring the well-designed studies demonstrating scientific robustness of the Rorschach, while emphasizing and summarizing any study they can find that seems to suggest something negative about the instrument (Meyer, 2000; Weiner, 2001). Although their “straw man” arguments sound reasonable to the naive reader (including well meaning academics who depend on peer-reviewed journals), the erroneous conclusions actually provide little useful information to either the practitioner who struggles with the ethical application of assessment measures or researchers facing the daunting work of Rorschach research. Additionally, these attacks on psychological assessment and psychological testing weaken rather than strengthen the public’s view of applied psychology.

Like it or not, these “pseudo-debates,” and the associated literature containing articles that do not meet the rigors of good science, eventually find their way into the court room and provide another source of distraction in an already difficult work arena. Forensic psychologists find themselves in the embarrassing and awkward position of having to ed-

²Concerning the “arm chair” quality Gacono and Bodholdt (2001) noted, “We would extend by drawing attention to the occasional recourse to rhetorical devices, including the straw man, and selective abstraction of a backwater of supposed conclusions, which on careful reading, run counter to the prevailing tone, substance or conclusion of the source-proper” (p. 66).

ucate the courts about inaccuracies on the part of these so-called experts.³ Even psychologists very well qualified to educate the court and defend psychological assessment methods against these pseudo-debates often find themselves in lengthy and tedious cross-examinations designed to diminish the impact of their findings, sometimes by numbing juries and judges with highly technical “scientific” debate and thus losing the point of the forensic evaluation.

The aforementioned issues require active intervention by the forensic practitioner (Gacono et al., 2001). Forensic psychologists must educate their peers and others concerning what we do and how it works. Biased and scientifically unsupported attacks on one assessment instrument detract from the practice of forensic psychology as well as the practice of psychological assessment in general. It is in this spirit that we offer *The Handbook of Forensic Rorschach Assessment*. While focusing primarily on presenting guidelines for using the Rorschach in forensic practice, these chapters will assist evaluators, more broadly, in critically evaluating the advantages and disadvantages of psychological testing within the context of the forensic examination.

FORENSIC PSYCHOLOGICAL ASSESSMENT

A sophisticated and applied knowledge of psychological assessment and psychological testing, an understanding of psycholegal issues and the rules of evidence, and experience with forensic populations are essential to understanding the role of the Rorschach in forensic practice, including offering informed commentary on its efficacy. In order to evaluate the utility of the RIM in forensic assessment, one must understand what psychological assessment is and is not. First, psychological assessment does not equal psychological testing! Viewing the two as synonymous demonstrates a lack of understanding of the much broader scope of forensic work (Gacono et al., 2001; Gacono & Bodholdt, 2001). Such a misconception detracts from forensic psychologists’ unique contribution to assessment, and encourages a de-emphasis of well-balanced and in-depth clinical training in graduate psychology programs. Psychological assessment is more broadly defined as a process that “integrate[s] the results of several carefully selected tests with relevant history, information and observation ... enabl[ing] the sophisticated clinician to form an accurate, in-depth understanding of the patient; formulate the most appropriate and cost-effective treatment plan; and later, monitor the course of intervention.” (Meyer et al., 1998).

Alternately:

Assessment is a *process* of deduction, selective inquiry, and also inference ... rooted in a knowledge of developmental psychology, personality and individual differences, statistics and measurement, with knowledge of limits (e.g. in prediction), in cognitive science, ethics, abnormal

³Some thoughts on how psychologists can handle challenges: Voir dire should be utilized to challenge the qualifications of these people related to their licensing status, their actual practice of psychological assessment (do they see and assess people), their knowledge of forensic practice and guidelines, their advanced standing in any recognized professional personality assessment group (i.e., fellow status in the Society for Personality Assessment), and so forth, before considering the testimony.

psychology including dynamics and defenses ... Assessment forms the cornerstone of the “forensic mind-set”—one that is data based, utilizing test data, observation, interviewing, and multi-sources of substantiated historical information in *forming, testing, and modifying hypotheses* ... Assessment is a multifaceted, ongoing, interactive process (Gacono, 2000, pp. 194–195)

Forensic psychologists are always aware that psychological testing is only one component of psychological assessment, that no single data source can accurately assess the complexity inherent in forensic assessment, and that personality tests are not designed to directly assess psycholegal issues (see Otto, 2001).⁴

Having clarified the psycholegal issue and clearly understanding the limitations of any one method of obtaining information, the forensic examiner will find Monahan and Steadman’s (1994) risk assessment model useful as a guide for directing specific assessment methods (Gacono, 2002a, 2000b, 2000c). Monahan et al. (2001) emphasized the need for gathering data using multiple methods from multiple domains:

1. Dispositional factors (including anger, impulsivity, psychopathy, and personality disorders).
2. Clinical or psychopathological factors (including diagnosis of mental disorder, alcohol or substance abuse, and the presence of delusions, hallucinations, or violent fantasies).
3. Historical or case history variables (including previous violence, arrest history, treatment history, history of self-harm, as well as social, work, and family history).
4. Contextual factors (including perceived stress, social support, and means for violence).

After the psychologist’s role in assessing the relevant psycholegal issue is clearly defined, the forensic psychologist must determine which, if any, of the previous domains provide information needed to address the referral questions (psycholegal issue). Next, the forensic psychologist chooses reliable and valid methods and instruments for obtaining and organizing the data from the relevant domains.⁵ Finally, valid results from the assessment methods are integrated into opinions that aid the trial of fact in addressing the psycholegal issue. This four-step process can be summarized as:

⁴Paradoxically, psychological assessment and psychological testing have been de-emphasized during a time when their usefulness has been clearly articulated (Meyer et al., 1998). In clinical settings, the de-emphasis of assessment has been rationalized as (a) too “costly” without a balanced accounting of the costs involved when it is ignored; and/or (b) “too time intensive” (actually it is the skill level of the clinician that prolongs the administration, scoring, and interpretation of the data—not inherent qualities of the test). There is also the practice of utilizing easily administered paper-and-pencil tests, which require minimal contact with the patient, with little consideration for the actual purpose of the evaluation (little if any relationship between the test and referral question). Of no surprise, these poorly conceived “window dressing” assessment protocols result in findings of little value; a finding that is subsequently used to justify the de-emphasis of formal assessment.

⁵The usefulness of psychological tests depends on the individual assessment context. For example, when assessing competency to stand trial, routine cases require semistructured interview questions to ascertain the individuals’ understanding of their current legal situation, their ability to cooperate with counsel, and so forth; in a given case, personality testing may not be necessary. However, in the case of an identified psychopath (PCL–R ≥ 30) suspected of malingering Schizophrenia, the evaluation of malingering may necessitate administration of the SIRS (Rogers, 1986), observation of ward behavior, and assessment of thought disorder with the Rorschach. The same logic can be applied to other forensic issues including the assessment (not diagnosis) of psychopathy.

1. Establish referral questions/psycholegal issue.
2. Determine domains of information relevant to 1.
3. Choose assessment methods for addressing domains identified in 2.
4. Valid data is integrated into forensic opinions.

In the process of choosing assessment methods for forensic evaluations, we emphasize the need for using multiple assessment methods, such as review of collateral materials and records, clinical and semistructured interviewing, standardized psychological testing and so forth. For example, some methods, such as the PCL–R (Hare, 2003) and other semistructured interviews, are useful for collecting and quantifying certain dispositional and historical variables, whereas other methods such as the RIM and MMPI–2 add to understanding certain clinical and dispositional variables.

Although psycholegal issues suggest the relevant domains needed to be assessed, with the chosen domains guiding the psychologist’s selection of assessment methods, the Federal Rules of Evidence and other legal standards guide the admissibility of psychological testimony. These standards require that expert testimony be relevant to the psycholegal issue, be of assistance to the fact finder, provide information beyond the understanding of a layperson, and not be overly prejudicial (Otto, 2001). Admissibility standards require that assessment methods, including psychological tests, must have *relevance* to the forensic issue and, as such, determine what constitutes a sound assessment strategy (McCann, 1998).

FORENSIC ASSESSMENT: ASSESSING HISTORICAL, DISPOSITIONAL, CLINICAL, AND CONTEXTUAL FACTORS

To better illustrate sound forensic assessment strategy, we use the following approach for assessing violence risk as an illustrative example. A first step in risk assessment involves a thorough review of historical and collateral information, including documentation relating to history of violence (including sexual assault), previous offenses, weapon use, and other factors relevant to specific risk concerns. This data provides historical and contextual information essential to forming an opinion. Next, the forensic psychologist assesses clinical and dispositional factors, including contemporary data on critical mental status markers, acute paranoid ideation, and delusions that require third-party corroboration from a review of treatment records, staff interviews, and other pertinent sources. Also, the evaluator reviews specific antecedents and consequents surrounding previous violent acts (contextual and dispositional) as well as the mode or type of violence (affective versus predatory). The evaluatee’s violence history is further clarified through interviewing concerning cognitive, affective, and behavioral patterns prior to, during, and consequent to violent episodes, as well as any current situational or dynamic factors that could be influenced by immediate intervention.⁶ In addition to relevant historical, dispositional, clinical, and contextual factors, victim characteristics should also be assessed, such as age, gender, and the circumstances under which predation occurred.

⁶Record review and clinical interview allow identification of specific person-context factors (e.g., medication noncompliance, alcohol or drug use, level of supervision or custody) expected to mitigate or amplify more immediate risk of re-offense, including violent re-offense.

Subsequent to assessing history and mental status through a review of records and other documentation, and conducting forensic and collateral interviews, the forensic psychologist's opinions can be further enhanced by using an established actuarial risk assessment instrument such as the Violence Risk Appraisal Guide (VRAG), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998), and/or, the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997). Historical, clinical, dispositional, and contextual variables are quantified through the completion of such assessment procedures. For example, the VRAG and SORAG (Quinsey et al., 1998) are two protocols that produce a violence prediction probability estimate based on an empirically derived algorithm utilizing demographic, historical, and clinical variables, with a significant contribution made by the patient's psychopathy level assessed by the PCL-R (Hare, 1991, 2003). Although this actuarial data establishes an essential basis for forming opinions, the forensic psychologist must consider the limitations of primarily static, unchangeable data that are acquired through these methods (see Zamble & Quinsey, 1997, concerning the problems of "tombstone" predictors).

Forensic evaluation using the previous assessment domains provides a sound basis for case conceptualization. In addition, personality testing refines our understanding of dispositional or clinical factors such as impulsivity, levels of anger and hostility, presence of thought disorder, problems with affect regulation, and methods of coping with emotions (Gacono et al., 2001; Gacono & Meloy, 1994, 2002). Standardized psychological testing adds to understanding important similarities and differences among individuals to an extent not possible using only risk assessment guides and PCL-R scores that provide primarily nomothetic comparisons (Gacono, 1998). Combining historical information, risk assessment guide scores, PCL-R scores, and personality testing data allow the forensic psychologist to provide opinions emphasizing specific, individualized context-person dynamics—that is, under what circumstances a given patient is more likely to perpetrate a certain type of violence toward a particular type of victim. Such specificity allows the rigor of scientific knowledge to blend with the art of clinical insight to provide a uniquely comprehensive and human view of the individual being assessed. The forensic psychologist operates from the assumption that assessment is a multifaceted, ongoing, interactive process involving a continuous process of forming, testing, and modifying hypotheses.

CONCLUSIONS

In many jurisdictions, psychologists are called on to articulate how conclusions were derived (see *Daubert v. Merrell Dow Pharmaceuticals*, 1993). Under *Daubert* guidelines, the trier of fact evaluates the probative value of the forensic assessment using four criteria: the underlying theory or technique can and has been tested, the methodology employed has been subjected to scrutiny via peer review and publication, rates of error and classification obtained when using the technique are known and acceptable, and the degree to which the technique is accepted within the scientific community. Whereas these criteria are particularly relevant to specific methods of gathering data (individual tests),

the use of the Monahan and Steadman (1994) domains for organizing global assessment strategies offers an especially comprehensive model for articulating the overall assessment process.

Additionally, it is of critical importance that forensic psychologists are cognizant of the nature and limitations of their data. For example, some data are static, whereas others are dynamic (changeable). Test scores should be considered in terms of error rates (i.e., within the context of a range of scores) and compared to corresponding normative data. Group comparisons (nomothetic data) may be only inferentially relevant to an individual case. In this manner, nomothetic data provides a starting point for interpretation; however, individual differences, teased out through multimethods and multilevels of assessment (collecting assessment data from multiple domains), are necessary to forming sound opinions.

Furthermore, the forensic psychologist has a special duty to understand how psychological tests work. This knowledge is essential for interpreting and reconciling “apparent” discrepancies among tests. Particularly relevant to the forensic context is the fact that results from self-report measures such as the MMPI–2 and MCMI–III may measure how people accurately see themselves or alternatively how they would like to appear others (McCann, 2001). As such, a given profile may be heavily influenced by the forensic context and not yield an accurate measure of the existing psychopathology. In this regard, forensic psychologists must consider the potential impact of the response style to the assessment process (Bannatyne, Gacono, & Greene, 1999). This critical issue is best assessed through a battery of tests that access different aspects of personality. In this manner, the Rorschach contributes uniquely to forensic practice (Gacono et al., 2001) in its demonstrated resistance to response style influence (Ganellen, 1994, 1996; Ganellen et al., 1996; Grossman et al., 2002).

All the chapters in this text purposely include at least one author who is a full or part-time licensed practicing forensic psychologist. The authors bring their practical expertise, guided by a systematic approach to personality assessment and knowledge of the good science, to their presentation of the Rorschach use in specific forensic contexts. In Part I (chaps. 1–6), the reader is provided with essential information related to the scientific and legal basis of the Rorschach. These chapters provide essential elements for preparing for informed court testimony. In Part II, models are presented for using the Rorschach in typical forensic evaluations (chaps. 7–15) involving both criminal (competency, insanity, risk assessment) and civil (custody, personal injury, fitness for duty). Part III contains updated references samples for various forensic populations, which are to be used as part of the ever-growing Comprehensive System database (chaps. 16–21).

In Part IV, the psychologist will find useful models for the Rorschach use in specialized areas of forensic practice (chaps. 22–25), such as with battered women, immigration court assessment, assessing impaired professionals, and working within the field of police psychology. Additionally, chapter 27 discusses how Rorschach findings are integrated with other assessment methods. It is our hope (Gacono, Evans, Kaser-Boyd, & Gacono) that this text will provide psychologists with a comprehensive resource to guide their forensic practice. It is within this context that we dedicate this text to John E. Exner and Paul Lerner, who passed away during the past year.

ACKNOWLEDGMENTS

A version of this chapter, "Introduction to a Special Series: Forensic Psychodiagnostic Testing," was previously published in *Journal of Forensic Psychology Practice*, 2(3), 1–10, 2002.

REFERENCES

- Bannatyne, L., Gacono, C., & Greene, R. (1999). Differential patterns of responding among three groups of chronic, psychotic, forensic outpatients. *Journal of Clinical Psychology*, 55(12), 1553–1565.
- Daubert v. Merrell Dow Pharmaceuticals*, 516 US 869 (1995).
- Exner, J. E. (1997, 2003). *The Rorschach: A Comprehensive System: Basic foundations and principles of interpretation* (4th ed.). Hoboken, NJ: Wiley.
- Gacono, C. (1998). The use of the Psychopathy Checklist–Revised (PCL–R) and Rorschach in treatment planning with Antisocial Personality Disordered patients. *International Journal of Offender Therapy and Comparative Criminology*, 42(1), 49–64.
- Gacono, C. (2000). Suggestions for the implementation and use of the Psychopathy Checklists in forensic and clinical practice. In C. B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 175–202). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gacono, C. B. (2002a). Guest editorial: Why there is a need for the personality assessment of offenders. *International Journal of Offender Therapy and Comparative Criminology*, 46, 271–273.
- Gacono, C. B. (2002b). Introduction to a special series: Psychological testing in forensic settings. *International Journal of Offender Therapy and Comparative Criminology*, 46, 274–280.
- Gacono, C. B. (2002c). Introduction to a special series: Forensic psychodiagnostic testing. *Journal of Forensic Psychology Practice*, 2, 1–10.
- Gacono, C., & Bodholdt, R. (2001). The Role of the Psychopathy Checklist–Revised (PCL–R) in violence risk and threat assessment. *Journal of Threat Assessment*, 1(4), 65–79.
- Gacono, C., & Evans, B. (2004). Entertaining reading but not science: Book review of What's wrong with the Rorschach. *International Journal of Offender Therapy & Comparative Criminology*, 48(2), 253–257.
- Gacono, C., Evans, B., & Viglione, D. (2002). The Rorschach in forensic practice. *Journal of Forensic Psychology Practice*, 2(3), 33–53.
- Gacono, C., Evans, B., Jumes, M., & Loving, J. (2002). The Psychopathy Checklist in forensic practice: PCL–R testimony. *Journal of Forensic Psychology Practice*, 2(3), 11–32.
- Gacono, C., Loving, J., & Bodholdt, R. (2001). The Rorschach and psychopathy: Toward a more accurate understanding of the research findings. *Journal of Personality Assessment*, 77(1), 16–38.
- Gacono, C., & Meloy, R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gacono, C., & Meloy, R. (2002). Assessing antisocial and psychopathic personalities. In J. Butcher (Ed.), *Clinical foundations of personality assessment* (2nd ed., pp. 361–375). New York: Oxford University Press.
- Ganellan, R. J. (1994). Attempting to conceal psychological disturbance: MMPI defensive response sets and the Rorschach. *Journal of Personality Assessment*, 63, 423–437.
- Ganellan, R. J. (1996). Exploring MMPI–Rorschach relationships. *Journal of Personality Assessment*, 67, 529–542.
- Ganellen, R. J., Wasyliw, O. E., Haywood, T. W., & Grossman, L. S. (1996). Can psychosis be malingered on the Rorschach: An empirical study. *Journal of Personality Assessment*, 66, 65–80.

- Grossman, L. S., Wasylw, O. E., Benn, A. F., & Gyoerkoe, K. L. (2002). Can sex offenders who minimize on the MMPI conceal psychopathology on the Rorschach. *Journal of Personality Assessment*, 78, 484–501.
- Hare, R. (1998). The Hare PCL–R: Some issues concerning its use and misuse. *Legal and Criminal Psychology*, 3, 99–119.
- Hare, R. (1991). *Hare Psychopathy Checklist Revised (PCL–R) technical manual*. North Tonawanda, NY: Mutli-Health Systems.
- Hare, R. (2003). *Hare Psychopathy Checklist Revised (PCL–R) technical manual* (2nd ed.). North Tonawanda, NY: Mutli-Health Systems.
- McCann, J. T. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment*, 70(1), 125–144.
- McCann, J. (2001). Guidelines for the forensic application of the MCMI–III. *Journal of Forensic Psychology Practice*, 2.
- Meloy, R., & Gacono, C. (2000). Assessing psychopathy: psychological testing and report writing. In C. B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 231–250). Mahwah, NJ: Lawrence Erlbaum Associates.
- Meyer, G. (1999). Introduction to the special series on the utility of the Rorschach for clinical assessment. *Psychological Assessment*, 11(3), 235–239.
- Meyer, G. (2000). On the science of Rorschach research. *Journal of Personality Assessment*, 75(1), 46–81.
- Meyer, G., Finn, S., Eyde, L., Kay, G., Kubiszyn, R., Moreland, K., Eisman, E., & Dies, R. (1998). *Benefits and costs of psychological assessment in healthcare delivery: Report of the Board of Professional Affairs Psychological Assessment Work Group, Part I*. Washington, DC: American Psychological Association Practice Directorate.
- Monahan, J., & Steadman, H. (1994). Toward a rejuvenation of risk assessment research. In J. Monahan & H. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment* (pp. 1–17). Chicago: University of Chicago Press.
- Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford.
- Otto, R. (2001). Use of the MMPI–2 in forensic settings. *Journal of Forensic Psychology Practice*, 2.
- Quinsey, V., Harris, G., Rice, M., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR–20 assessing risk for violence version 2*. Burnaby, British Columbia Mental Health Law and Policy Institute, Simon Fraser University.
- Weiner, I. (2001). Assessment advocacy—report of the SPA coordinator. *SPA Exchange*, 12(1), 7.
- Zamble, E., & Quinsey, V. (1997). *The criminal recidivism process*. New York: Cambridge University Press.

ACKNOWLEDGMENTS

I would like to acknowledge, and thank, the individual authors for contributing their areas of Rorschach expertise to this volume. Their particular virtuosity and insight is evident in the exceptional quality of the text. A special acknowledgment is reserved for my wife, Lynne Gacono, for not only her emotional support and patience, but also for her tireless focus in reviewing and editing manuscripts. If not for her, this text would not have been finished.

I wish to thank Steve Rutter and Nicole Buchmann, production editors at Lawrence Erlbaum Publishers, for their patience in keeping me on track and offering encouragement throughout this project. Also, a special thanks to friends at LEA, Michele McEvoy and Lawrence (Larry) Erlbaum. Larry, as always, provided valued guidance and perspective unique among publishers.

It is with much affection that this book is dedicated to Paul M. Lerner and John Exner, Jr. They have not only served as significant role models, but have been two of my most valued friends, colleagues, and mentors.

—CBG

The contributors to this book have each offered an important piece of the overall mosaic of the forensic use of the Rorschach. Far from being outside the mainstream, these authors together demonstrate in very real and practical ways that the Rorschach is alive, well, and making valuable contributions to the practice of forensic psychology.

My special appreciations are many. Endlessly energetic, relentlessly honest, creative, and brilliant, Carl Gacono is simply the very best collaborator with whom I have ever written. Deepest thanks go to dear friend Benjamin Schutz, who introduced me to the practice of forensic psychology, who has been my major forensic guide, and in many ways who is the inspiration for this book. Particular thanks also go to Bruce Copeland, a long-time friend, colleague, and source of support on all things forensic and fly fishing. Sincere appreciation goes to our responsive and vastly competent editors Steve Rutter and Nicole Buchmann at LEA for their support, help, and encouragement. Words cannot begin to describe my gratitude to Larry Erlbaum, a man wise beyond his many years, who proves over and over again that good relationships are the heart of good business.

On a more personal note, were it not for the sudden and devastating losses of two friends and lions of the Rorschach, Paul Lerner and John Exner, I would have dedicated this book to the following three people. I personally dedicate this book, along with my

deepest love and appreciation, to my daughter Rebecca Evans and my son Samuel Evans, whose caring and honesty over the years have taught me more than they could possible imagine, and to Judy Maris, whose unwavering devotion to putting the head and the heart together inspires me like no other.

—*FBE*

CONTRIBUTORS

Marvin W. Acklin, Private Practice, Honolulu, HI
George I. Athey, Jr., PhD, Heritage Mental Health Clinic, Topeka, KS
Matthew R. Baity, PhD, Massachusetts General Hospital, Boston, MA
Michael Bridges, PhD, Temple University, Philadelphia, PA
Ted B. Cunliffe, PhD, California Department of Corrections and Rehabilitation,
Susanville, CA
Philip S. Erdberg, PhD, Private Practice, Corte Madera, CA
F. Barton Evans, PhD, Private Practice, Bozeman, MT
Carl B. Gacono, PhD, ABAP, Clinical and Forensic Psychology, Austin, TX
Lynne A. Gacono, PhD, Austin State Hospital, Austin, TX
Ronald J. Ganellen, Northwestern University Medical School, Chicago, IL
James F. Gormally, PhD, Private Practice, Silver Spring, MD
Peter Graham, PhD, Acumen Assessments, Inc., Lawrence, KS
B. Thomas Gray, PhD, ABPP, North Texas State Hospital, Vernon, TX
Carl F. Hoppe, PhD, Private Practice, Beverly Hills, CA
Michael T. Jumes, PhD, North Texas State Hospital, Vernon, TX
Jerrald Justice, Private Practice, Walnut Creek, CA
Nancy Kaser-Boyd, PhD, Geffen School of Medicine, UCLA, Los Angeles, CA
Margaret Lee, PhD, Private Practice, Greenbrae, CA
Hale Martin, PhD, Denver University, Denver, CO
Joseph McCann, PsyD, JD, UHS Hospitals, Binghamton, NY
J. Reid Meloy, PhD, ABPP, University of California, San Diego, CA
Greg J. Meyer, PhD, University of Toledo, Toledo, OH
Nancy W. Olesen, PhD, Private Practice, San Rafael, CA
Benjamin M. Schutz, PhD, Private Practice, Springfield, VA
Jacqueline Singer, PhD, Private Practice, Sonoma, CA
Bruce L. Smith, PhD, University of California, Berkley, CA
Scott C. Stacy, PsyD, Acumen Assessments, Inc., Lawrence, KS
Donald J. Viglione, Alliant International University, San Diego, CA
Marjorie G. Walters, PhD, Private Practice, San Rafael, CA
Irving B. Weiner, PhD, ABPP, ABAP, University of South Florida, Tampa, FL
Peter A. Weiss, PhD, Interfaith Medical Center, Brooklyn, NY

William U. Weiss, PhD, University of Evansville, IN
Myla H. Young, Private Practice, Walnut Creek, CA

ABOUT THE EDITORS

Carl B. Gacono, PhD, is a licensed psychologist who maintains a clinical and forensic private practice in Austin, Texas. Formerly the Assessment Center Director at Atascadero State Hospital and, later, the Chief Psychologist at the Federal Correctional Institution, Bastrop, Texas, he has over 20 years of correctional and institutional experience. Dr. Gacono is the author of *The Clinical and Forensic Interview Schedule for the Hare Psychopathy Checklist: Revised* and *Screening Version*, senior author of *The Rorschach Assessment of Aggressive and Psychopathic Personalities*, coeditor of *Contemporary Rorschach Interpretation*, editor of *The Clinical and Forensic Assessment of Psychopathy: A Practitioner's Guide*, and has authored or coauthored over 80 scientific articles and book chapters. He is the 1994 recipient of the Samuel J. And Anne G. Beck Award for excellence in early career research, the 2000 Walter G. Klopfer Award, a member of the American Board of Assessment Psychology, and a Fellow of the Society for Personality Assessment. Dr. Gacono is sought as an expert in the area of criminal behavior, psychopathy, treatment of conduct and antisocial personality disorders, and clinical, forensic, and research applications of the Rorschach and Psychopathy Checklists.

F. Barton Evans, PhD, is a licensed psychologist who maintains a clinical and forensic private practice in Bozeman, MT. He has practiced psychology for 30 years in wide variety of settings, including Yale University School of Medicine and Georgetown University. Dr. Evans is Clinical Professor of Psychiatry and Behavior Science at George Washington University Medical School. He is the author of a book, *Harry Stack Sullivan: Interpersonal Theory and Psychotherapy*, and has authored or coauthored over 20 scientific articles and book chapters. He is a Fellow of the Society for Personality Assessment, an officer of its Board of Trustees, and chaired its ad hoc committee responsible for the development of "The Status of the Rorschach in Clinical and Forensic Practice," also known as the Rorschach White Paper. Dr. Evans is sought as an expert in the areas of psychological issues in immigration matters, parenting plan evaluations, psychological trauma, and forensic applications of the Rorschach.

Nancy Kaser-Boyd, PhD, is a licensed psychologist who has maintained a clinical and forensic practice in Los Angeles for 25 years. She is Associate Clinical Professor in the Department of Psychiatry and Biobehavioral Sciences at UCLA Medical School, where she teaches the use of psychological tests to evaluate difficult clinical cases (e.g., facti-

tious disorders, complicated mood disorders, effects of trauma, personality disorders) and forensic issues (e.g., dangerousness, insanity, malingering). Dr. Kaser-Boyd is coauthor of *Essentials of Rorschach Assessment*, and numerous book chapters and journal articles on evaluating trauma and forensic topics. Dr. Kaser-Boyd completed a postdoctoral fellowship at the University of Southern California Institute of Psychiatry and Law. She is member of the American Board of Assessment Psychology, and a Fellow in the Society for Personality Assessment. A member of two Los Angeles County Superior Court Panels (Criminal and Dependency), her expertise is sought in cases of severe child neglect and abuse, in family homicide, and in psychological assessments of forensic issues where she has used the Rorschach to illustrate critical mental states.

Lynne A. Gacono, PhD, is a licensed psychologist in both private and public practice. Dr. Gacono is the chairperson for the Manifest Dangerousness Committee at Austin State Hospital. She is also a member of the Texas State Forensic Committee. Her research interests and publications have involved the use of the Rorschach and MMPI–2 in assessing forensic psychiatric patients. Dr. Gacono maintains a private clinical and forensic practice in Austin, Texas, where she works with women’s issues, eating disorders, addictions, chronic pain, and physical conditions.

THE HANDBOOK OF FORENSIC RORSCHACH ASSESSMENT

PART

I

SCIENTIFIC AND LEGAL FOUNDATIONS

1

ESSENTIAL ISSUES IN THE FORENSIC USE
OF THE RORSCHACH

Carl B. Gacono

Private Practice, Austin, TX

F. Barton Evans III

Private Practice, Bozeman, MT

Donald J. Viglione

Alliant International University

The Rorschach is one of the most widely used, openly accepted, and frequently requested tests in forensic psychology practice (Meloy, 1991; Piotrowski, 1996; Weiner, Exner, & Sciara, 1996). It consistently meets the rigors of forensic scrutiny (Weiner et al., 1996). Weiner et al. (1996) found that in 7,934 recent federal and state court cases, in which psychologists presented Rorschach data, 6 cases were challenged and in just one the testimony was excluded. Uniformly, where challenges are sustained, it has not been the Rorschach's psychometric properties (Viglione & Meyer, chap. 2, this vol.) that have been impeached, but rather the psychologist's interpretations (Meloy, chap. 4, this vol.; Meloy, Hanson, & Weiner, 1997). Discredited interpretations were either too broad (used to prove a crime was committed), too narrow (aiding in diagnosis formulation without linking them to the forensic issue), or irrelevant to the legal issue before the court. Accordingly, these three caveats suggest potential pitfalls in Rorschach testimony.

Despite its favorable status and a substantial body of literature attesting to its reliability and validity (Viglione & Meyer, chap. 2, this vol.; Weiner, 1996), the Rorschach has been targeted for attack by a small group of psychologists. In fact, as with Neuropsychological assessment, it is this very acclaim and acceptance that has induced method critics to attack the test (Board of Professional Affairs, 1998; Wood & Lilienfeld, 1999). Under the guise of "good" science, the rhetoric and tactics of these detractors has been likened instead to "advocacy" or politics (Weiner, 2001a, p. 7). Weiner noted that, instead of a neutral, dispassionate, and thorough review of the literature, these critics ignored methodologically sound studies, while citing studies that emphasized negative

aspects of the Rorschach (Meyer, 2000; Weiner, 2001a).¹ Such a practice ignores the very dictates of good science.

These attacks on the Rorschach are largely polemical and editorial. As Atkinson (1986) suggested, upon completing a meta-analysis with favorable results for the Rorschach, the “deprecation of the Rorschach is a sociocultural, rather than scientific, phenomenon” (p. 244). These attacks rest on a misunderstanding of clinical and forensic practice and a misrepresentation of the empirical literature.

The most vexing problem about these attacks is the manner in which ethical and moralistic language is used as a vehicle for conveying polarized positions as moral imperatives. Historically, this moralistic point of view about the Rorschach and performance based testing occurred episodically since the 1920s and these modern-day biased Rorschach criticisms are the latest incarnation of a longer tradition (Exner, 2003; Viglione & Rivera, 2003). These moral pretensions make claims about the righteousness of science and its empirical foundations, but offer no new empirical data and eschew comprehensive and balanced syntheses of the empirical based Rorschach literature. Under careful examination, the misleading tactics of these detractors becomes evident (Gacono & Evans, 2004; Martin, 2003). As stated by Gacono and Bodholdt (2001, p. 66), “We would extend by drawing attention to the occasional recourse to rhetorical devices, including the straw man, and selective abstraction of a backwater of supposed conclusions, which on careful reading, run counter to the prevailing tone, substance or conclusion of the source-proper.”

Beyond any intentional biases, the conclusions of these detractors suggest that they do not understand how the Rorschach actually works (Gacono & Evans, 2004; Martin, 2003; see appendix B).² As noted by Weiner (1995, p. 73), “Those who currently believed the Rorschach is an unscientific or unsound test with limited utility have not read the relevant literature of the last 20 years; or, having read it, they have not grasped its meaning.”

Well-trained psychologists with a sophisticated understanding of psychological constructs use the Rorschach to derive information beyond what is available from diagnosis, self-report tests, and interview. From a broader perspective, the Rorschach plays a valuable role in the description of the complex interaction among psychological, biological, environmental, and behavioral domains (Viglione & Perry, 1991), and therefore is frequently relied on in forensic evaluations (Ackerman & Ackerman, 1996, 1997).

¹Summaries and meta-analyses of Rorschach research (Atkinson, Quarrington, Alp, & Cyr, 1986; Bornstein, 1996, 1999; Gacono, Loving, & Bodholdt, 2001; Meyer & Archer, in press; Hiller, Rosenthal, Bornstein, Berry, & Brunell Neuleib, 1999; Parker, Hanson, & Hunley, 1988; Viglione, 1999) reveal that Rorschach variables are associated with many criteria relevant to forensic contexts and incremental validity beyond interview, diagnosis, self-report test, and ability tests. Also, respondents’ efforts to minimize problems and to present themselves in a positive light do not greatly influence Rorschach and projective test variables in contrast to self-report tests (Brems & Johnson, 1991; Bornstein, Rossner, Hill, & Stepnian, 1994; Ganellen, 1994; Harder, 1984; Shedler, Erdberg, & Haroian, 1993).

²Voire dire can be utilized to challenge the qualifications of these individuals regarding their licensing status, their practice of psychological assessment (do they see and assess people), their advanced standing in any recognized professional personality group (i.e., fellow status in the Society for Personality Assessment), and so forth. In some cases, testimony might be encouraged in order to expose their lack of competence and discredit their ethical violations (i.e., practicing in areas beyond their expertise, training, and competence).

Weiner (1996) observed that critics had ignored 20 years of empirical support for the Rorschach. Taking up this challenge, Viglione (1999) reviewed 20 years of empirical investigations of the Rorschach in five major journals and found that the available data supported its validity and utility in a variety of areas. An excellent example of Weiner's observation of critics' ignoring the empirical data can be found in Grove and Barden's (1999) work advocating for the inadmissibility of the Rorschach under *Daubert/Kumho*. Ostensibly offering a "scientific" review of the evidence on reliability and validity of the instrument, these authors systematically ignored the numerous positive reliability and validity studies. Given the persistent bias against "Projective tests,"³ Masling (1997) questioned whether data supporting the Rorschach would change the mind of the critics. He attributes some of this rigidity, politicization, and bias to the fact that former students, emboldened by their teachers, have become critics of Projective tests. This bias against the Rorschach has existed since the 1920s among American academic psychology departments and, despite the research in support of the Rorschach, a reading of the recent criticisms suggests that Masling's speculations are true.

Like it or not, these biased and unscientific opinions find their way into the courtroom through published articles and even books. Although they do little to promote scientific study (Meyer, 2000) or provide useful information to the trier of fact, such biased criticisms present an avenue for opposing attorneys to challenge when they search for possible weaknesses in psychological testimony. Forensic psychologists using the Rorschach need to prepare for this additional challenge (see chaps. 1–6 this vol.). This chapter summarizes key issues from several essential articles that will aid the examiner in preparing for forensic testimony.

ESSENTIAL ARTICLES

One positive outcome of these pseudo-debates⁴ has been the publication of a plethora of excellent, scientific articles that serve as a guide for forensic Rorschach usage. The following summarizes four articles that provide a starting point for anchoring the contents of this handbook. We recommend that forensic psychologists thoroughly review the original sources.

Rorschach Testimony

Meloy (1991) provides an excellent overview for using the Rorschach in court. He emphasized six major points:

1. The training of the examiner.

³The correct term for describing the Rorschach is a "performance-based test." It has been demonstrated that the Rorschach is a cognitive perceptual task that involves only limited amounts of "projection."

⁴We use the term "pseudo-debate" as, for the most part, Rorschach detractors have presented "legal briefs" rather than scientific reviews (Weiner, 2001b). Legal briefs are designed to present convincing arguments and evidence to support these arguments (presenting alternative evidence is the responsibility of the opposing party). A scientific review, on the other hand, is expected to present an unbiased account of the literature and to report contradictory evidence when it exists" (Barrett & Morris, 1993, pp. 201–202). Inviting individuals with little working knowledge of the test to debate has been counterproductive. By nature of the resultant publications, it has elevated unqualified professionals to the role of "Rorschach expert" and given them a voice that offers little to Rorschach science.

2. Ensuring accurate scoring and administration.
3. Having a thorough knowledge of the validity data.
4. Not overinterpreting the data.
5. Avoiding the use of psychological jargon.
6. Use of the Rorschach data in court.

Training, Scoring, and Administration. Unlike most self-report measures, the Rorschach requires extensive course work and supervised practice to become proficient with its basic administration and scoring. In addition to course work in graduate school, achieving scoring proficiency requires advanced training with the Exner Comprehensive System (CS),⁵ as well as supervision on administration and scoring. Because of the higher standards of expertise required in Forensic assessment, psychologists holding themselves out as experts on the Rorschach CS should have experience scoring 50 to 100 protocols at a minimum. Rorschach users understand that proficient scoring alone does not equal Rorschach mastery, including skillful interpretation (Gacono, DeCato, Brabender, & Goertzel, 1997).

We cannot emphasize enough the importance of proper administration. Whereas reliable scoring is seldom an issue—research has found most Rorschach variables achieve an acceptable level of agreement—inadequate inquiry can be. Specifically, because scoring is based on recorded verbalizations that are directly linked to inquiry, high levels of scoring agreement can be achieved without “proper” inquiry. For example, CBG consulted on a case where the defense hypothesized that their client committed an act of affective, rather than predatory, violence. Although the floating MMPI–2 protocol was consistent with this hypothesis, Rorschach results yielding +3 *D* and +3 *AdjD* and a *RIAP* stating that “this person has a sturdier tolerance for stress than most people and he is unlikely to experience problems with controls ...” were not. Whereas a careful examination of the protocol revealed only two nonsignificant content scoring errors, close inspection of the responses recorded during administration indicated that the examiner failed to inquire whenever the presence of shading or inanimate movement was suggested. The absence of inquiry concerning these key variables, which contribute significantly to the *D* and *AdjD*, occurred in over 12 responses. When the Structural Summary was recalculated including these “missed” *shading and inanimate movement* responses, the resultant –3 *D* and –1 *AdjD* was more consistent with both the MMPI–2 findings and the defense attorney’s hypothesis. This underscores a critical point, namely, scoring reliability should only be examined after it has been determined that the test was properly administered and inquired.

Validity. Whereas we discuss Rorschach validity at greater length in the next section, we concur with Meloy (1991) concerning the need for the forensic Rorschach examiner to not only be familiar with the validity research related to specific CS variables, but also be conversant with relevant forensic comparative data and the implications for forensic practice. These forensic reference groups (also see chaps. 16–21) have included

⁵Due to the extensive research and normative data, the Exner Comprehensive System has been consistently endorsed by forensic psychologists as the only system that can meet the standards of forensic scrutiny (McCann, 1998).

specific data available for: forensic psychiatric patients (Bannatyne, Gacono, & Greene, 1999; Gacono & Meloy, 1994; Gacono & Gacono, chap. 20, this vol.; Nieberding et al., 2003; Young, Justice, Erdberg, & Gacono, 2000), sex offenders (Bridges, Wilson, & Gacono, 1998; Cohan, 1998; Gacono & Meloy, 1994; Gacono, Meloy, & Bridges, 2000; Gacono, Meloy, & Bridges, chap. 18, this vol.), conduct disordered children and adolescents (Gacono, Gacono, & Evans, chap. 16, this vol.; Gacono & Meloy, 1994; Loving & Russell, 2000; Smith, Gacono, & Kaufman, 1995; Weber, Meloy, & Gacono, 1992), and antisocial and psychopathic males and females (Cunliffe, 2002; Cunliffe & Gacono, 2005; Cunliffe & Gacono, chap. 17, this vol.; Gacono & Meloy, 1994; Gacono, Meloy, & Berg, 1992; Gacono et al., 2000; Gacono, Gacono, & Evans, chap. 16, this vol.).

Interpretations, Psychological Jargon, and Rorschach Data in Court. Psychologists' interpretations should be firmly anchored in validity research. Equally important is the manner in which Rorschach information is presented (Weiner, chap. 6, this vol.). A common trap for forensic psychologists is for attorneys on cross-examination to elicit responses to questions about reliability and validity that mimic dissertation defenses or presentations at scientific meetings. Although extraordinary technical detail about statistics, methodology, and research are important survival tactics in such settings, this kind of response in court will often bore the judge or jury to tears. This reaction is exactly what the attorney intends. Paradoxically, by going down the Alice in Wonderland rabbit hole of detailed, technical exposition of kappa, chance corrected reliabilities, and base rate sensitivities, vivid and cogent Rorschach findings can be lost and the credibility of the Forensic examiner to the trier of fact can be seriously eroded. It is incumbent on forensic psychologists to prepare counsel calling them to testify for such challenges, to use plain language in the courtroom, and to prepare short, precise, confidently stated, and understandable explanations of complex topics suitable to the average lay person.

Recent research lends support for this position. Krause and Sales (2001) found that clinical expert opinion was more persuasive than actuarial expert opinion in mock jurors' ratings about dangerousness determination and that adversarial procedures such as cross-examination or competing expert testimony had less impact on clinical expert opinion. Additionally, in a large national survey of state court judges, Gatowski et al. (2001) found that, although judges strongly endorse (91%) the utility of *Daubert/Kumho* guidelines, their clear understanding particularly behind the scientific concepts of falsifiability (5%) and error rate (4%) was quite low. Together these studies suggest that scientific expert opinion is likely to be most persuasive when it is rich with clinical examples and where it does not overwhelm the trier of fact with arcane scientific exegesis.

As with all psychological tests, Meloy (1991) reaffirmed the ethical requirement of protecting the security of psychological tests and measures, as essential to forensic Rorschach use. Examiners should avoid photocopying copy written materials, such as location charts, and/or taking the actual Rorschach cards to court. Forensic lore includes the story of the examiner who takes the Rorschach cards to court, only to have them taken away while testifying and passed among the judge and jury. Subsequently, a rendition of the examinee's verbatim responses is used to discredit the test.

Rorschach Validity

As noted by Weiner (1996), addressing the current attacks on the Rorschach requires an understanding of the validity of psychological tests. Tests are inferential measures, and as such, their correlation with other tests provides only limited information concerning their validity. Several tests may correlate with each other, but not with observable behavior (Weiner, 1996). Correlations with observed thoughts, feelings, and actions provide more powerful information than correlations with other assessment instruments. Furthermore, the validity of any test can only be addressed by specifying the purposes for which they are more or less valid, that is, valid for what? (Weiner, 1996).

Key meta-analytic studies have confirmed that conceptual, theory-based studies demonstrate substantially higher validity coefficients for Rorschach variables than research undertaken without a theoretical or empirical rationale (Viglione & Hilsenroth, 2001; Weiner, 2001a). Conceptual validation studies of the Rorschach indicate adequate validity values equivalent to those found for the MMPI: "The MMPI and Rorschach are valid, stable, and reliable under certain circumstances. When either test is used in the manner for which it was designed and validated, its psychometric properties are likely to be adequate for either clinical or research purposes" (Parker, Hanson, & Hunsley, 1988, p. 373). Furthermore, there are at least six well-designed, original meta-analyses addressing Rorschach validity, which all find empirical support for the validity of the test (Atkinson et al., 1986; Bornstein, 1996, 1999; Hiller et al., 1999; Meyer & Archer, 2001; Parker et al., 1988).

As Weiner et al. (1996) and Meloy et al. (1997) found, validity of the Rorschach was most often questioned when interpretations were too broad, too narrow, or irrelevant to the legal issue before the court. All of these challenges fall squarely into concerns about test validity. The question for personality assessors is not whether the Rorschach is valid, rather to answer the question, "What can the Rorschach do for you?" (i.e., What is it valid for?; Weiner, 1999). These general cautions suggest an approach to Rorschach interpretation based on the specific psycholegal questions asked for the forensic assessment. Whereas Exner's (2000) interpretative search strategy is the most sophisticated method available for mining the rich data yield available in the Rorschach CS for clinical interpretation, not all CS variables are relevant to the psycholegal questions to be assessed or are empirically sturdy enough to draw definitive conclusions. It is incumbent on the Forensic psychologist to understand the construct validity research behind the CS variables and sets of variables relevant to the legal questions.

On a related note, a test must be reliable in order to be valid. Careful reading of the Rorschach literature is essential to defending the test's reliability and for revealing when a critic is only using studies that support their biases or to reveal inconsistencies and misinterpretation of existing data (Viglione & Meyer, chap. 2, this vol.). For example, in commentaries on the research, the only empirical data cited (Garb et al., 2001; Wood & Lilienfeld, 1999) to support claims of unreliability in Rorschach scoring are 8 reliability coefficients (2%) selected from 403 reported by Acklin, McDowell, Verschell, and Chan

(2000). Three of these 8 intraclass correlation coefficients (ICC) were less than .60, and mistakenly described as “low” (Garb et al., 2001, p. 436).⁶

In another criticism of Rorschach’s reliability, the critics failed to include 16 studies that produced findings counter to their arguments (see Meyer, 1997a, 1997b).⁷ Rorschach reliability coefficients compare quite favorably to those for other tests (Acklin et al., 2000; Exner, 1993; McDowell & Acklin, 1996; Meyer, 1997a; Meyer et al., 2002; Viglione, 1999; Viglione & Hilsenroth, 2001). Well-trained raters can score both high and low base rate CS variables with good (Kappa and ICC > .60) to excellent (> .75–.80) reliability (Garb, 1998; Shrout & Fliess, 1979). The majority of the CS variables that are central to the interpretive variables are also reliable (Exner, 1993). Those variables without test–retest data are, for the most part, not central to interpretation or contained in other variables for which we have test–retest coefficients (Viglione & Hilsenroth, 2001).

The most glaring example related to interrater reliability is the manner in which Wood, Nezworski, Lilienfeld, and Garb (2003) misrepresent how it is computed. As noted by Gacono and Evans (2004, p. 234):

The straw person rhetoric is no more evident than in the authors’ discussion of interrater reliability (pp. 227–228). The authors provide a method for computing interrater reliability and then indicate why it is faulty. Although their presentation sounds and is plausible, it is based on a disquietingly inaccurate portrayal of how percentage agreement is computed ... which is methodologically completely and unequivocally erroneous. From a scientific perspective, what is disturbing about this example is that, even when presented with information to the contrary, the authors have stuck with their inaccurate beliefs and continue to use them as a basis for supporting their arguments.

Finally, studies reporting less than acceptable temporal consistency reliability for the CS have serious methodological flaws. These flawed studies include research with a nonstandard form of administration (Schwartz, Mebane, & Malony, 1990); a dissertation involving only 17 older adults with an unspecified test–retest interval (Erstad, 1996); and a sample of Schizophrenic patients with various illness courses, medications, hospitalizations, psychotherapies and test–retest intervals ranging from 1 to 18 years (averaging 6.4 years; Adair & Wagner, 1992).

The great majority of individual CS variables are coded with good or excellent interrater reliability (e.g., 95% in Acklin et al., 2000). Only a few very low base rate variables, typically of little interpretive significance (e.g., *MQ none*) occasionally produce poor re-

⁶In previous research conducted by Garb (1998), he, in agreement with experts in the psychometric literature (Cicchetti, 1981; Landis & Koch, 1977; Shrout & Fliess, 1979), defined coefficients greater than .60 as indicating “good” reliability. Additionally, among the 403 coefficients in Acklin et al. (2000), a range of over and underestimates of the true reliability would be expected due to standard errors of measurement. Standard error of measurement are especially large with low base rate variables in small samples as those included in Acklin et al. (2000).

⁷Meyer (1997a) conducted a meta-analysis of these studies and found a mean estimated Kappa of .86 and a mean percent agreement of 92% for the variables.

liability coefficients, but they have not yet been analyzed with sufficiently large samples given their minuscule base rates (Viglione & Meyer, chap. 2, this vol.).

Admissibility

McCann (1998) analyzed the admissibility of the Rorschach in court using both legal and professional standards (McCann, 1998). He systematically evaluated the Rorschach using two major sources of admissibility: expert evidence and psychometric evidence. First, McCann outlines the three most important legal tests for admissibility: the *Federal Rules of Evidence* (1992), the *Frye* test (*United States v. Frye*, 1923), and the *Daubert* standard (*Daubert v. Merrell Dow Pharmaceuticals, Inc.* 1993). Secondly, he summarizes the standards for forensic use of psychological tests set forth by two important professional articles: Heilbrun (1992) and Marlowe (1995).

Using both legal tests and professional standards, McCann ambitiously derives nine standards, including publication and peer review, standard administration and norms, reliability, validity, rate of error, general acceptance, relevance and helpfulness, falsifiability, and response style interpretation. McCann's standards go beyond legal standards, offering a more stringent test for acceptability of psychological tests in forensic settings. Using these higher standards, McCann analyzes the admissibility of the Rorschach in legal and forensic settings. In summary, McCann (1998) concludes:

An analysis of the current clinical and research status of the Rorschach reveals that it meets professional and legal standards for admissibility of psychometric evidence and expert testimony. This conclusion rests on the foundation of a large body of literature that exists for the Exner Comprehensive System because this method of administration and scoring is standardized, has documented psychometric characteristics, and has been the primary subject of Rorschach research over the past 20 years. . . . Legal admissibility does not require that all scientific issues be completely resolved to the satisfaction of all members of the professional community. Expert testimony must rest on methodology that is generally accepted, testable, standardized, relevant, and helpful. In all of these respects, the Rorschach is an appropriate methodology to utilize in forensic evaluation. (pp. 140–141)

Addressing the Critics

Gacono, Loving, and Bodholdt (2001) provide an alternate format for reviewing the Rorschach's status, improving Rorschach research, and addressing biased criticisms. Rather than encouraging pseudodebates that do little to contribute to assessment improvements, Gacono et al. (2001) offer guidelines for evaluating Rorschach research related to Antisocial Personality and Psychopathy, and recommend that other experts do the same in their conceptual areas. They provided five areas of conceptual understanding that are essential to responsible criticisms of the psychopathy/ Rorschach literature:

1. Antisocial Personality Disorder (ASPD; American Psychiatric Association, 1994) and Psychopathy are related but distinct constructs, differing from each other along important historical, theoretical, and definitional lines.

2. Psychopathy may be conceptualized both in dimensional terms (i.e., along a continuum of severity) and in categorical terms (i.e., as a taxon or discrete syndrome), and that applying one of these approaches versus the other to PCL–R scores affects research findings;

3. Psychopathy may manifest in varying forms across various populations, for example across gender or throughout development from youth into adulthood.

4. Personality testing is only one facet of both psychological assessment and diagnosis. It contributes to the assessment of the dimensional aspects of Psychopathy. Research findings must be adapted for clinical application.

5. While methodological limitations inherent to certain Rorschach/Psychopathy studies may limit our ability to generalize these particular findings to other settings, they in no way invalidate the compendium of well-designed studies as certain rather persistent Rorschach detractors would have us believe. (p. 17)

Seemingly plausible sounding arguments have been presented by critics that are actually quite biased and hold little weight. For example, critics may present ASPD and Psychopathy as synonymous, lower cut-off scores for designating Psychopathy as a category, equate diagnosis with assessment, and so on (Gacono, Bodholdt, & Loving, 2001; Gacono & Evans, 2004; Gacono & Gacono, 2006). When a study includes these errors, either the critic is being naive or they are consciously promoting their own biases. In either case, ignoring any of the previous five principles provides a foothold for impeaching Rorschach critics.

In addition to outlining five key conceptual issues, Gacono et al. (2001) offered four essential criteria for reviewing the methodology of individual Psychopathy/Rorschach studies:

1. CD and ASPD are comprised of heterogeneous groups of individuals. Studies that treat Psychopathy as a taxon must validate groups with an appropriate measure (e.g. the [Psychopathology Checklist–Revised] PCL–R with adults, the PCL:YV for adolescents [or a PCL:YV prepublication—modified version of the PCL–R]) and use the accepted cut-off scores ($PCL-R \geq 30$).

2. Studies need to account for (control or delineate) the limitations imposed by factors such as gender, sexual deviance, concurrent Axis I psychosis, age, IQ, testing setting, and legal status. These factors can influence the production of certain Rorschach variables.

3. R (number of responses) must be considered. Increased R is found in certain sex offender groups (Bridges, et al., 1998; Gacono et al., 2000), whereas low R is typical among many criminal groups (Viglione, 1999). Thus, R can act as a moderator influencing the relationship between Rorschach variables and criterion variables. Research should investigate this hypothesis by controlling R and examining the relationship between Rorschach variables and criterion constructs at different levels of R (e.g., $R = 14$ – 17 , etc.).

4. Response style must be considered (Bannatyne et al., 1999). Variables and styles such as R , Λ , $Extratensiveness$, and $Introversiveness$ can impact the production of

certain Rorschach variables (Exner, 1995), contributing to seemingly discrepant findings among studies.

When a reader understands these five conceptual and four methodological criteria, they know “what to look for” in sorting pseudodebates from legitimate scientific debate. It is essential to avoid being seduced by the plausible sound of conclusions based on studies that fail to consider these issues (e.g., Brody & Rosenfeld, 2000). These conclusions are easily impeached with closer scrutiny.

For example, two recent articles addressing forensic issues (Dawes, 1999; Grove & Barden, 1999) presented a biased and unbalanced picture of the Rorschach with obvious glaring deficiencies in the authors’ claims. Both of these publications fail to incorporate large bodies of empirical evidence from refereed journals demonstrating the validity of the Rorschach. (e.g., Blais, Hilsenroth, Castlebury, Fowler, & Baity, 2001; Meyer, 1997a, 1997b; 2000; Meyer & Archer, 2001; Shedler, Mayman, & Manis, 1993; Viglione, 1999; Viglione & Hilsenroth, 2001). Dawes (1999, p. 301) referred to the “deficiency” of the Rorschach in relation to the criterion of reasonable certainty employed in forensic applications. His argument is based on some elusive and illogical juxtaposition of incremental validity statistical analysis with the forensic notion that expert testimony should be “incremental” in the forensic sense of aiding the trier of facts. In making his sweeping conclusions, Dawes ignores much of the available information and research support from incremental and criterion validity for the Rorschach. In the worst case, Dawes misunderstands the Rorschach itself. For example, the requirement that the Ego Impairment Index show incremental validity over its own subcomponent in a small sample, reveals a lack of awareness of the structure of the test. Such lack of expertise is more vividly demonstrated by a description in a previous attack on the Rorschach: “Six of the Rorschach cards are black or various shades of gray, and the remaining four are colored” (Dawes, 1994, p. 146). Anyone who is familiar with the test would know that five cards are all black and gray, two are black and gray with shades of red, and three are chromatically colored without prominent black and grey features.

These unscientific and prejudiced attacks on the Rorschach obfuscate forensic issues and mislead jurists and practitioners. Most useful to the forensic psychologist would be additional Rorschach articles that, following the lead of Gacono et al. (2001), offer conceptual and methodological guidelines for evaluating other relevant bodies of literature central to Rorschach usage.

THE RORSCHACH’S ROLE IN FORENSIC ASSESSMENT

Criticisms of the Rorschach have observed that reliability and validity in the research lab is different from what might be called “field” reliability and validity (i.e., whether the test is reliable and valid as used by practitioners; e.g., Hunsley & Bailey, 1999). Following this lead, Weiner (1999) demonstrated that field incremental validity only actually manifests within a single case when it adds some unique information to an assessment. The Rorschach does this. In addition to its scores and ratios, the Rorschach provides an additional opportunity to observe the client’s behavior in response to novel and complex stim-

uli (their performance). Observing the clients “performance” provides an added dimension to more traditional face-to-face interviewing, both essential components of most forensic evaluations.

The Rorschach also adds incrementally to the information obtained by self-report measures such as the MMPI–2. It is well-established that Rorschach scores are minimally, if at all, correlated with self-report scores in any direct or one-dimensional fashion (Archer & Krishnamurthy, 1993a, 1993b). Ironically, this difference supports the utility of the test. By measuring different aspects of personality functioning and/or measuring personality in a different way, the Rorschach’s incremental validity is ensured.

Additionally, a benefit of the Rorschach rests in its ability to partially bypass an individual’s volitional controls. Research supports the view that the Rorschach is most useful in contexts in which the respondent may be unwilling or unable to engage in the examination fully. Such opportunities for the Rorschach would include employment, criminal forensic, and custody evaluations that involve adversarial relationship components.

Under these adversarial conditions, response manipulation can skew or even invalidate reliable findings from self-report (Bannatyne et al., 1999). As a result, self-report data may provide little usable information in forensic contexts, beyond acknowledging distorted response style. For example, research supports the practitioners’ decision to embrace the Rorschach in custody evaluations in the context of the limitations of self-report instruments. Bathurst, A. W. Gottfried, and A. E. Gottfried (1997) found that 508 custody litigation participants produced defensive MMPI–2s. Mean *L*, *F*, and *K* elevations were 56, 45, and 59, with 20% producing profiles with $L \geq 65$ and 25% with $K \geq 65$. Bagby, Nicholson, Buis, Radovanovic, and Fidler (1999) reported similar results, with MMPI–2 means for 115 custody litigants of 62, 48, and 58. They noted that 52% of their subjects produced either *L* or $K \geq 65$, and that Wiggins Social Desirability Scale and the Superlative Scale identified 74% as underreporting. Consistently, these data demonstrate that the MMPI–2 produce a limited yield in custody evaluations with 40% to 75% of litigants’ results providing indeterminate and noncontributory findings. Bathurst et al. (1997) conclude, “It was not possible to determine from the MMPI–2 per se whether this approach is an overestimate of mental health in a psychologically healthy population or an attempt by psychologically disturbed individuals to conceal symptomatology” (p. 209). Bagby et al. commented that the answer to the Bathurst question “necessitates the collection of extra test data” (p. 28).

Additionally, Haywood, Grossman, Kravitz, and Wasyliw (1994) found that alleged child molesters produced even more defensive findings, with 75% of their subjects minimizing pathology according to a standard MMPI criterion. Haywood et al. also found that sexual cognitive distortions on a specific self-report instrument (Cognition Scale; Abel, Becker, & Cunningham-Rather, 1984) were similarly confounded by defensive responding. Defensiveness and constriction are consistent with findings with Anti-social offenders (Gacono & Meloy, 1994) and other forensic populations (Bannatyne et al., 1999; Nieberding et al., 2003), which further indicated the limited usefulness for self-reported tests in many, if not most, forensic settings (Gacono & Meloy, 2002). The utility of self-report scales is limited in that they do not distinguish between healthy and defensive functioning (Shedler et al., 1993) and emphasize characteristics mediated by

social convention and role behavior rather than implicit motives and tacit traits (McClelland, Koestner, & Weinberger, 1989).

“What the Rorschach can do for you?” is most relevant in custody evaluations. Defensive response styles on the MMPI–2 (see Baer & Miller, 2002; Bagby et al., 1999; Bathhurst et al., 1997; Medhoff 1999; Posthuma & Harper 1998) and the MCMI (Lampel, 1999; Halon, 2001) are common in these contexts. Participants in custody battles understandably may not be willing or able to portray themselves accurately with self-report measures, making the Rorschach more valuable. This may be the reason that psychologists conducting custody evaluations often use the Rorschach for both for adults and children (Ackerman & Ackerman, 1997).

Whereas the Rorschach is often used in conjunction with other assessment methods in addressing a variety of forensic issues, it is least likely to be requested for determinations of competency to stand trial, and most likely to be sought in answering questions of sanity at the time of the crime, treatability, or dangerousness (Rogers & Cavanaugh, 1983). These trends do not eliminate its usefulness to any given forensic evaluation (see Acklin, chap. 8, this vol.; Gacono & Evans, 2004, introduction, this vol.; Gray & Acklin, 2007). For example, in the case of an identified psychopath (PCL–R ≥ 30) suspected of malingering Schizophrenia, the evaluation of competency to stand trial would be aided by a strategy that incorporates the Rorschach, the SIRS (Structured Interview of Reported Symptoms), the PCL–R, and collateral data. Whereas the assessment of malingering may necessitate administration of the SIRS (Rogers, Bagby, & Dickens, 1986), an observation of ward behavior, the assessment of thought disorder (with the Rorschach), and an evaluation of Psychopathy level would add weight to the examiner’s conclusions. First and foremost, the context of the evaluation suggests high motivation for feigning a mental illness. Second, ward observations might reveal inconsistencies in the patient’s presentation, such as interacting in a normal manner when the patient doesn’t realize staff is observing. Next, the Rorschach’s unusual thinking, perceptual accuracy, and reality testing indices may be inconsistent for psychosis, but consistent with character disordered or nonpatient samples. Finally, the elevated PCL–R total score along with a substantiated history of lying and conning and manipulation (PCL–R items 4 & 5 = 2s) add additional data to the hypothesis that this patient is malingering. Knowing the research and literature as suggested earlier in this chapter would lead one to know that information is also available about malingering Schizophrenia on the Rorschach (Netter & Viglione, 1994). Thus, within multiple assessment methods, the Rorschach also has multiple impacts on the evaluation outcomes. The same logic can be applied to other forensic and clinical issues.

It is important for the forensic examiner to develop a systematic strategy for inclusion of Rorschach variables in a forensic assessment. Suggested guidelines for assessing the validity of CS variables are outlined next. These guidelines are ordered from most to least relevant in terms of cogency for the forensic evaluation:

1. Are there Rorschach variables directly relevant to the legal issue before the court?
2. Where Rorschach variables are directly relevant, are there norms and construct validity research on the target forensic population?

3. Where Rorschach variables are not directly relevant to the legal issue before the court, are there CS variables with sufficient validity to be informative about pertinent personality attributes and behavioral propensities?

First, are there Rorschach variables directly relevant to the legal issue before the court? Clearly, the Rorschach is not going to directly answer the question of whether someone is competent to stand trial such as measures like Grisso's (1988) Competency Assessment Procedure. On the other hand, although not a direct measure of a legal or forensic issue, the Rorschach can quite nicely lend important data when the individual's mental health is at issue before the court (Acklin, chap. 8, this vol.; Gray & Acklin, chap. 7, this vol.). For example, the Rorschach can provide important information about whether or not an individual's thinking and perception is indicative of an underlying psychotic process.

Second, where Rorschach variables are directly relevant, are there norms and construct validity research on the target forensic population? Clearly, Exner's (2003) non-forensic, patient and nonpatient samples, and Gacono and Meloy's (1994) Forensic samples provide essential populations for comparison. Part III of this volume adds to these Forensic samples (Cunliffe & Gacono, chap. 17, this vol.; Gacono, Gacono, & Evans, chap. 16, this vol.; Gacono & Gacono, chap. 20, this vol.; Singer et al., chap. 21, this vol.).

Third, where Rorschach variables are not directly relevant to the legal issue before the court, are there CS variables with sufficient validity to be informative about pertinent personality attributes and behavioral propensities? For example, in a custody case in which one parent is alleged to have chronic difficulties with anger management, CS variables involving affect modulation, anger, and aggression, and their durability over time (such as Exner's, 1993, Test-retest studies) are likely to be relevant. Lastly, are there CS variables that due to their singular nature or to the problems of response style on other psychological tests or clinical interview shed some light on the question at hand? These variables may not have the strongest validity research (e.g., Sx Content variable) or may be in an experimental stage, namely, acceptance in the Comprehensive System (e.g., Armstrong's Trauma Content Index, Armstrong & Lowenstein 1990; Gacono & Meloy, 1994, Aggressive Content Scales; Gacono, Gacono, Meloy, & Baity, 2005; chap. 26, this vol.). Clearly, the Forensic psychologist will rely incrementally less on those interpretations, which either have little bearing on the specific legal question or, alternatively, lack strong validity research for a given CS variable.

Forensic psychologists are always cognizant of the fact that the Rorschach should be part of a multimethod approach to assessment utilizing such methods as self-report measures, clinical and structured interviews, and collateral information. All of these sources of data are needed to assess the historical, clinical, dispositional, and contextual variables required in a Forensic Psychological Assessment (Gacono, 2002; Preston & Liebert, 1990). When properly administered, scored, and interpreted, however, the Rorschach is an invaluable tool that adds incrementally to the practitioner's armament. Despite the recent re-emergence of attacks on the Rorschach, the solid research of the past

20 years, including an analysis of the test's acceptance in the courtroom, suggests that the Rorschach will continue to be widely used in clinical and forensic work for years to come.

ACKNOWLEDGMENT

This chapter is a modified version of "Introduction to a Special Series: Forensic Psychodiagnostic Testing," published in the *Journal of Forensic Psychology Practice*, 2(3), 1–10, (2002).

REFERENCES

- Abel, G. G., Becker, J. V., & Cunningham-Rathner, J. (1984). Complications, consent, and cognitions in sex between children and adults. *International Journal of Law and Psychiatry*, 7, 89–103.
- Ackerman, M. J., & Ackerman, M. C. (1996). Child custody practices: A 1996 survey of psychologists. *Family Law Quarterly*, 30, 565–586.
- Ackerman, M. J., & Ackerman, M. C. (1997). Child custody evaluation practices: A survey of experienced professionals (revisited). *Professional psychology: Research and Practice*, 28, 137–145.
- Acklin, M. W., McDowell II, C. J., Verschell, M. S., & Chan, D. (2000). Inter-observer agreement, intra-observer reliability, and the Rorschach comprehensive system. *Journal of Personality Assessment*, 74, 15–47.
- Adair, H. E., & Wagner, E. E. (1992). Stability of unusual verbalizations on the Rorschach for outpatients with schizophrenia. *Journal of Clinical Psychology*, 48(2), 250–256.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Armstrong, J., & Lowenstein, R. (1990). Characteristics of patients with multiple personality and dissociative disorders on psychological testing. *Journal of Nervous and Mental Disease*, 174, 448–454.
- Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology/Psychologie Canadienne*, 27(3), 238–247.
- Atkinson, L., Quarrington, B., Alp, I. E., & Cyr, J. J. (1986). Rorschach validity: An empirical approach to the literature. *Journal of Clinical Psychology*, 42, 360–362.
- Archer, R. P., & Krishnamurthy, R. (1993a). Combining the Rorschach and the MMPI in the assessment of adolescents. *Journal of Personality Assessment*, 60, 132–140.
- Archer, R. P., & Krishnamurthy, R. (1993b). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment*, 61, 277–293.
- Baer, R., & Miller, J. (2002). Underreporting of psychopathology on the MMPI–2: A meta-analytic review. *Psychological Assessment*, 14, 16–26.
- Bagby, R. M., Nicholson, R. A., Buis, T., & Radovanovic, H. (1999). Defensive responding on the MMPI–2 in family custody and access evaluations. *Psychological Assessment*, 11, 24–28.
- Bannatyne, L. A., Gacono, C. B., & Greene, R. (1999). Differential patterns of responding among three groups of chronic psychotic forensic outpatients. *Journal of Clinical Psychology*, 55(12), 1553–1565.
- Barrett, G., & Morris, S. (1993). The American Psychological Associations' amicus curiae brief in *Price Waterhouse v. Hopkins*: The values of science versus the values of the law. *Law and Human Behavior*, 17, 201–216.
- Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI–2 in child custody litigation. *Psychological Assessment*, 7, 419–423.
- Blais, M. A., Hilsenroth, M. J., Castlebury, F., Fowler, J. C., & Baity, M. R. (2001). Predicting DSM–IV Cluster B personality disorder criteria from MMPI–2 and Rorschach data: A test of incremental validity. *Journal of Personality Assessment*, 76, 150–168.

- Board of Professional Affairs (1998). Awards for distinguished professional contributions: John Exner. *American Psychologist*, 53, 391–392.
- Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency scale: 1967–1995. *Psychological Assessment*, 8, 200–205.
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, 11(1), 48–57.
- Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment*, 63, 363–386.
- Brems, C., & Johnson, M. E. (1991). Subtle–Obvious scales of the MMPI: Indicators of profile validity in a psychiatric population. *Journal of Personality Assessment*, 56(3), 536–544.
- Bridges, M., Wilson, J., & Gacono, C. B. (1998). A Rorschach investigation of defensiveness, self-perception, interpersonal relations, and affective states in incarcerated pedophiles. *Journal of Personality Assessment*, 70, 365–385.
- Brody, & Rosenfeld, (2002). Object Relations in Criminal Psychopaths. *International Journal of Offender Therapy and Comparative Criminology*, 46, 4.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Cohan, R. (1998). *A comparison of sex offenders against minors and rapists of adults on selected Rorschach variables*. Unpublished doctoral dissertation, Wright Institute, Berkeley, CA.
- Cunliffe, T. (2002). *A Rorschach investigation of incarcerated female psychopaths*. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.
- Cunliffe, T., & Gacono, C. (2005). A Rorschach investigation of incarcerated female offenders with antisocial personality disorder. *International Journal of Offender Therapy and Comparative Criminology*, 49(5), 530–546.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.* 113 S. Ct. 2786 (1993).
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: The Free Press.
- Dawes, R. M. (1999). Two methods for studying incremental validity of a Rorschach variable. *Psychological Assessment*, 11, 297–302.
- Erstad, D. (1996). *An investigation of older adults' less frequent human movement and color responses on the Rorschach*. Unpublished doctoral dissertation, Marquette University.
- Exner, J. E. (1993). *The Rorschach: A Comprehensive System: Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.
- Exner, J. E. (1995). *A Rorschach workbook for the Comprehensive System* (5th ed.). Ashville, NC: Rorschach Workshops.
- Exner, J. E. (2000). *A primer for Rorschach interpretation*. Asheville, NC: Rorschach Workshops.
- Exner, J. E. (2003). *The Rorschach: A Comprehensive System: Vol. 1. Basic foundations* (4th ed.). New York: Wiley.
- Federal rules of evidence*. (1992). Boston: Little, Brown
- Gacono, C. B. (2002). Introduction to a special series: Forensic psychodiagnostic testing. *Journal of Forensic Psychology Practice*, 2(3), 1–10.
- Gacono, C., & Bodholdt, R. (2001). The role of the Psychopathy Checklist–Revised (PCL–R) in violence risk and threat assessment. *Journal of Threat Assessment*, 1(4), 65–79.
- Gacono, C., DeCato, C., Brabender, V., & Goertzel, T. (1997). Vitamin C or pure C: The Rorschach of Linus Pauling. In J. R. Meloy, M. W. Acklin, C. B. Gacono, J. R. Murray, & C. A. Peterson (Eds.), *Contemporary Rorschach interpretation* (pp. 421–451). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gacono, C. B., & Evans, B. (2004). Entertaining reading but not science: Book review of What's wrong with the Rorschach. *International Journal of Offender Therapy and Comparative Criminology*, 48(2), 253–257.
- Gacono, C. B., & Gacono, L. A. (2006, April). Some caveats for evaluating the research on psychopathy. *The Correctional Psychologist*, 38(2), 7–9.

- Gacono, C. B., Gacono, L. A., Meloy, J. R., & Baity, M. (2005). The Rorschach Extended Aggression Scores. *Rorschachiana*, 27, 164–190.
- Gacono, C. B., Loving, J., & Bodholdt, R. (2001). The Rorschach and psychopathy: Toward a more accurate understanding of the research findings. *Journal of Personality Assessment*, 77(1), 16–38.
- Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gacono, C. B., & Meloy, J. R. (2002). Assessing antisocial and psychopathic personalities. In J. Butcher (Ed.), *Foundations of clinical personality assessment* (2nd ed.). Oxford, England: Oxford University Press.
- Gacono, C., Meloy, J., & Berg, J. (1992). Object relations, defensive operations, and affective states in narcissistic, borderline, and antisocial personality. *Journal of Personality Assessment*, 59, 32–49.
- Gacono, C. B., Meloy, J. R., & Bridges, M. (2000). A Rorschach study of psychopaths, sexual homicide perpetrators, and nonviolent pedophiles: Where Angels fear to tread. *Journal of Clinical Psychology*, 55(6), 1–21.
- Ganellen, R. J. (1994). Attempting to conceal psychological disturbance: MMPI defensive sets and the Rorschach. *Journal of Personality Assessment*, 6, 423–437.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Towards a resolution of the Rorschach controversy. *Psychological Assessment*, 13, 433–448.
- Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsberg, G. P., Merlino, M. I., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges in judging expert evidence in a *post-Daubert* world. *Law and Human Behavior*, 25(5), 433–458.
- Grisso, T. (1988). *Competency to stand trial evaluations: A manual for practice*. Sarasota, FL: Professional Resource Exchange.
- 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, and Law*, 5(1), 224–242.
- Halon, R. (2001). The Millon Clinical Multiaxial Inventory–III: The normal quartet in child custody cases. *American Journal of Forensic Psychology*, 19, 57–75.
- Harder, D. W. (1984). Character style of the defensively high self-esteem man. *Journal of Clinical Psychology*, 40(1), 26–35.
- Haywood, T. W., Kravitz, H. M., Grossman, L. S., & Wasylw, O. E. (1994). Profiling psychological distortion in alleged child molesters. *Psychological Reports*, 75, 915–927.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16, 257–272.
- 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(3), 278–296.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11, 266–277.
- Krause, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, 7(2), 276–310.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lampel, A. (1999) Use of the Millon Clinical Multiaxial Inventory–III in evaluating child custody litigants. *American Journal of Forensic Psychology*, 17, 19–31.
- Loving, J., & Russell, W. (2000). Selected Rorschach variables of psychopathic juvenile offenders. *Journal of Personality Assessment*, 75, 126–142.
- Marlowe, D. B. (1995) A hybrid decision framework for evaluating psychometric evidence. *Behavioral Sciences and the Law*, 13, 207–228.
- Martin, H. (2003). Scientific critique or confirmation bias? *The National Psychologist*, 121(5), 19.

- Masling, J. M. (1997). On the nature and utility of projective tests and objective tests. *Journal of Personality Assessment*, 69, 257–270.
- McCann, J. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment*, 70(1), 125–144.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690–702.
- McDowell, C., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interrater reliability: Conceptual and empirical foundations. *Journal of Personality Assessment*, 66(2), 308–320.
- Medhoff, D. (1999). MMPI–2 validity scales in child custody evaluations: Clinical versus statistical significance. *Behavioral Sciences and the Law*, 17, 409–411.
- Meloy, J. R. (1991, Fall/Winter). Rorschach testimony. *Journal of Psychiatry and the Law*, 221–235.
- Meloy, R., Hansen, T., & Weiner, I. (1997). Authority of the Rorschach: Legal citations during the past 50 years. *Journal of Personality Assessment*, 69, 53–62.
- Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment*, 9(4), 480–489.
- Meyer, G. J. (1997b). Thinking clearly about reliability: More critical correlations regarding the Rorschach Comprehensive system. *Psychological Assessment*, 9, 495–498.
- Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment*, 75(1), 46–81.
- 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., Baxter, D., Exner, J. E., Jr., Fowler, C. J., Hilsenroth, M. J., Piers, C. C., & Resnick, J. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment*, 78, 219–274.
- Netter, B., & Viglione, D. J. (1994). An empirical study of malingering schizophrenia on the Rorschach. *Journal of Personality Assessment*, 62, 45–57.
- Nieberding, R., Gacono, C., Pirie, M., Bannatyne, L., Viglione, D., Cooper, B., Bodholdt, R., & Frackowiak, M. (2003). MMPI–2 based classification of forensic psychiatric outpatients: An exploratory cluster analytic study. *Journal of Clinical Psychology*, 59(9), 907–920.
- 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(3), 367–373.
- Piotrowski, C. (1996). Use of the Rorschach in forensic practice. *Perceptual and Motor Skills*, 82, 254.
- Posthuma, A., & Harper, J. (1998). Comparison of MMPI–2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice*, 2, 437–443.
- Preston, J., & Liebert, D. (1990). Defending the Rorschach in forensic cases. *American Journal of Forensic Psychology*, 8(1), 59–66.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1986). *Structured Interview of Reported Symptoms manual*. Lutz, FL: PAR.
- Rogers, R., & Cavanaugh, J. (1983). Usefulness of the Rorschach: A survey of forensic psychiatrists. *Journal of Psychiatry and the Law*, 11, 55–67.
- Schwartz, N. S., Mebane, D. L., & Malony, H. N. (1990). Effects of alternate modes of administration on Rorschach performance of deaf adults. *Journal of Personality Assessment*, 54(3–4), 671–683.
- Shaffer, T., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS–R, and MMPI–2. *Journal of Personality Assessment*, 73, 305–316.
- Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist*, 48, 1117–1131.
- Shrout, P. E., & Fliess, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Smith, A. M., Gacono, C. B., & Kaufman, L. (1997). A Rorschach comparison of psychopathic and non-psychopathic conduct disordered adolescents. *Journal of Clinical Psychology*, 53(4), 239–300.

United States v. Frye 293F. 1013 (D. C. Cir. 1923).

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. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment*, 13, 452–471.

Viglione, D. J., & Perry, W. (1991). A general model for psychological assessment and psychopathology applied to depression. *British Journal of Projective Psychology*, 36, 1–16.

Viglione, D. J., & Rivera, B. (2003). Assessing personality and psychopathology with projective tests. In J. R. Graham & J. A. Naglieri & I. B. Weiner (Eds.), *Comprehensive handbook of psychology: Assessment psychology* (Vol. 10, pp. 531–553). New York: Wiley.

Weber, C., Meloy, J., & Gacono, C. (1992). A Rorschach study of attachment and anxiety in inpatient conduct-disordered and dysthymic adolescents. *Journal of Personality Assessment*, 58, 435–438.

Weiner, I. (1995). Variable selection in Rorschach research. In J. Exner, Jr. (Ed.), *Issues and methods in Rorschach research* (pp. 73–98). Mahwah, NJ: Lawrence Erlbaum Associates.

Weiner, I. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment*, 8(2), 206–213.

Weiner, I. (2001a). Considerations in collecting Rorschach reference data. *Journal of Personality Assessment*, 77(1), 122–127.

Weiner, I. (2001b). Assessment advocacy—report of the SPA coordinator. *SPA Exchange*, 12, 7.

Weiner, I., Exner, J., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment*, 67(2), 422–424.

Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment*, 6, 327–339.

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., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach?* San Francisco: Wiley.

Young, M. H., Justice, J. V., Erdberg, P. S., & Gacono, C. B. (2000). The incarcerated psychopath in psychiatric treatment: Management or treatment? In C. B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 313–331). Mahwah, NJ: Lawrence Erlbaum Associates.

AN OVERVIEW OF RORSCHACH PSYCHOMETRICS FOR FORENSIC PRACTICE

Donald J. Viglione

Alliant International University

Gregory J. Meyer

University of Toledo

This chapter addresses current evidence concerning the Rorschach Inkblot Test relevant to forensic practice. We present a selective overview of research findings and some new data to help explicate the scientific and empirical foundations of the test. The focus is primarily on psychometric issues of reliability, validity, normative reference values, and utility. Even when limiting ourselves to these topics, we are selective because it is not possible to address them comprehensively within a single chapter. We focus on topics of most interest in the forensic arena and that have attracted the most research and controversy lately.¹ There is no attempt to select research that supports or does not support the test, but rather a bias for selecting recent versus older and well-known and established evidence.

This review emphasizes Rorschach variables from the Comprehensive System (CS; Exner, 2003), but non-CS variables are included where relevant. In response to pressing concerns of most forensic psychologists when using the Rorschach, we address the recent criticisms of the Rorschach by synthesizing research findings. In doing so, we identify legitimate and spurious criticisms and describe and illuminate related limitations of the Rorschach. This entails our using the existing research literature and theory about the Rorschach to recommend certain alterations to interpretive practices and to identify important research needs.

CRITICISMS OF THE RORSCHACH FROM A HISTORICAL PERSPECTIVE

Before addressing psychometric issues, we present a brief historical perspective. Exner (1974) published the first edition of the Comprehensive System (CS), which was eventually recognized as being largely successful in meeting historical psychometric chal-

¹For coverage of issues not included in the chapter, see Meyer and Archer (2001) and Viglione and Hilsenroth (2001).

lenges of reliability and validity. In the 1980s into the 1990s, the CS became the dominant system in teaching and practice (Hilsenroth & Handler, 1995; Mihura & Weinle, 2002) and it has become to be used extensively on an international basis (e.g., in Argentina, Belgium, Brazil, Denmark, Finland, France, Holland, Japan, Israel, Italy, Peru, Portugal, Sweden, and Spain). Exner's works are contained in three volumes with eight editions and in five editions of his workbook.

Since 1995, the Rorschach has once again been subjected to a series of repetitive critical reviews from a group of coauthors (e.g., Garb, 1999; Grove, Barden, Garb, & Lilienfeld, 2002; Hunsley & Bailey, 1999, 2001; Lilienfeld, Wood, & Garb, 2000; Nezworski & Wood, 1995; Wood & Lilienfeld, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001a; Wood, Nezworski, Garb, & Lilienfeld, 2001b; Wood, Nezworski, & Stejskal, 1996), although controversy has existed since its origin (e.g., Hirt, 1962; Murstein, 1965; Rabin, 1981; Viglione & Rivera, 2003). Some of these criticisms are written to challenge the Rorschach in court (e.g., Dawes, 1999; Grove & Barden, 1999; Grove et al, 2002; Lilienfeld et al., 2000; Wood et al., 1996). Criticisms and controversies have waxed and waned in the literature. A regular tension has emerged between practitioners using the Rorschach, many of whom find the Rorschach to be indispensable in their applied work, and some academic researchers who consider the Rorschach and its evidentiary foundation to be fundamentally flawed.

Atkinson, Quarrington, Alp, and Cyr (1986), after presenting results from one of the earliest meta-analytic reviews on Rorschach validity, questioned why its validity is continuously challenged despite the evidence. They asserted bluntly, "The oft-cited explanation is that deprecation of the Rorschach is a sociocultural, rather than scientific, phenomenon" (p. 244). Others have asked whether the debate about the utility of the Rorschach is more philosophical and political, rather than academic and scientific (Viglione & Rivera, 2003).

To a degree, these recent challenges of the Rorschach and the CS prompted the current book on forensic issues. Although the controversy is part political and philosophical debate and part scientific and rational debate, one goal is to focus on the latter. Nevertheless, because it is probably impossible to step outside of the former, we note that we consider ourselves political centrists when it comes to the Rorschach. That is, we believe the evidence supports its use in clinical practice, but we also believe that, like all tests, it has its limitations. Continued research is needed to specify the applications and limitation for many interpretive postulates. Like all tests, it needs to be used cautiously and conscientiously.

RELIABILITY: DO WE MEASURE CONSISTENTLY?

Reliability can be globally defined as the extent to which a construct is assessed consistently. Once we are measuring something consistently, it is necessary to establish that what is being measured is actually what we want to measure (validity) and that the measured information is helpful in some applied manner (utility). We focus on reliability first.

There are four main types of reliability—internal consistency, stability, alternate forms, and interrater. *Internal consistency reliability* refers to the consistency or homogeneity of content over items, that is, whether the items of a scale or test measure the same

construct. In the Rorschach, the notion of an item would have two meanings. First, responses or cards could be considered items. This form of internal consistency reliability entails an assumption that each card or response provides an equal opportunity to measure the same construct (Exner, Armbruster, & Viglione, 1978). However, it is readily recognized that each card does not allow an equal opportunity for all scores (e.g., cards vary greatly in their pull for color or texture determinants), so that internal consistency reliability is infrequently evaluated or reported and it is considered largely inapplicable to the test.

In terms of internal consistency reliability, an item also translates to the individual subcomponents or criteria of composite indices (e.g., the subcomponents of the *DEPI* or Ego Impairment Index, *EII*). As an example of this version of internal consistency research, Hilsenroth, Fowler, and Padawer (1998) and Stokes, Pogge, Grosso, and Zaccario (2001) examined the internal consistency of the six criteria forming the Schizophrenia Index (*SCZI*), whereas Dao and Prevatt (2006) examined the five criteria of its successor, the Perceptual Thinking Index (*PTI*). Although they found evidence for a reasonable degree of homogeneity ($KR-20 = .79, .70$, and $.75$, respectively), these analyses are difficult to interpret because the six *SCZI* and five *PTI* criteria draw on just two types of scores, form quality and the cognitive special scores. As such, there should be a certain degree of artificial correlation among the criteria, although the precise magnitude would be hard to determine.

More substantively, the *SCZI* or *PTI* and all the other CS Constellation Indices were created as composites that draw on the full range of information available in a protocol to maximize validity; they were not developed as scales designed to measure a single homogeneous construct. As Streiner (2003) has pointed out, internal consistency reliability is important for scales assessing a homogenous construct but immaterial for a composite index. Indeed, efficiency in measurement is achieved through low rather than high inter-correlations among subcomponents or items. Accordingly, weak internal consistency reliability can accompany strong validity and utility.

Another type of reliability that has been largely considered inapplicable to the Rorschach is *alternate forms reliability*, which assesses the consistency of scores across parallel versions of an instrument. Although Holtzman specifically developed his set of inkblots to have two parallel forms and Behn-Eschenberg made an early effort at developing a set of inkblots to parallel Rorschach's inkblots (see, e.g., Exner, 2003, p. 12), at present a good parallel set of the 10 standard Rorschach inkblots does not exist.

Stability reliability, also known as temporal consistency or test-retest reliability, is essentially the consistency of scores over time. It has been applied to the Rorschach and the results generally have been acceptable to good (Grønnerød, 2003; Meyer & Archer, 2001; Viglione & Hilsenroth, 2001). Comprehensive System scores thought to measure traitlike aspects of personality have produced relatively high retest coefficients, even over extended time periods. Also, scores thought to reflect statelike emotional process have produced relatively low retest coefficients even over short time intervals.

However, the most recent large-scale and well-designed study of CS stability found lower than anticipated consistency over a 3-month retest period (Sultan, Andronikof, Réveillère, & Lemmel, 2006). For instance, stability coefficients for *R* and *Lambda*,

which index the overall richness or complexity of a protocol, were .75 and .72, respectively. Because these scores are related to the frequency of other scores in the protocol, when they are unstable most other scores will be unstable as well. Indeed, in this study the median level of stability reliability across a core set of 47 scores was .53 and the median across 87 ratios, percentages, and derivations in the lower portion of the Structural Summary was .55. Number of responses (*R*) and *Lambda*, as markers of task engagement, moderated stability. Stability reliability was greater among those individuals whose *R* and *Lambda* did not change much over time, as compared to the stability among those individuals whose *R* and *Lambda* differed at the two testings.

Conducted in France, the Sultan et al. study was a carefully executed investigation with a sound methodology and adequate controls. It also used the most sophisticated statistical analyses to date to examine potential moderators of stability, and several were identified that would increase stability if they were controlled (e.g., engagement with the Rorschach task, situational distress/emotional status). Variation over time due to situational distress or emotional status is not related to the true stability reliability of the test, so that test–retest statistics underestimate the Rorschach’s true reliability. Nevertheless, even taking this situational variation into consideration, the stability for the majority of the Rorschach CS variables in this study was limited.

More investigation of Rorschach stability reliability is needed (Meyer & Archer, 2001; Viglione & Hilsenroth, 2001), and Sultan et al.’s (2006) findings should be replicated. However, given the care that went into designing and executing this study, forensic examiners should be aware of the challenges to the CS that might emerge in the courtroom from these data. The Sultan data indicate that nonpatient volunteers for a study can provide notably different protocols when tested by one reasonably trained examiner and again 3 months later by a different reasonably trained examiner. This finding will remain even if it is subsequently discovered that certain methodological factors account for the lower than expected stability or if the majority of future studies find superior stability.

Putting these results in context might be illuminating. Forensic examiners should recognize that the global stability of Rorschach scores might, under some circumstances, be more similar to the stability of memory tests than the stability of intelligence tests. For instance, although the manual for the third edition of the Wechsler Memory scale (WMS; Psychological Corporation, 1997) does not report data for all subscales, the 1-month stability for 13 of its subscores is .71 ($N = 297$). Over a 7 ½-month retest interval, the average stability coefficient for 5 of its subscores was .66 (Dikmen, Heaton, Grant, & Temkin, 1999) and over a 9-month interval the average stability for 10 scores was .68 (Martin et al., 2002). Although these coefficients are higher than those observed in Sultan et al., more similar stability values are found for tests like the California Verbal Learning Test (CVLT) and the Hopkins Verbal Learning Test (HVLT). Over a retest interval of 1 to 2 months, the average stability of HVLT scores was about .50 (Barr, 2003; Benedict, Schretlen, Groninger, & Brandt, 1998). Average stability for CVLT scores also has been about .50 over a 1-year retest interval (Paolo, Tröster, & Ryan, 1997). Finally, as another example, the average stability of scores on the Extended Complex Figure Test was .46 over the course of a 1-week interval ($N = 55$; Woodrome & Fastenau, 2005). It should be

pointed out that memory ability is thought to be a stable trait similar to many personality and information-processing variables accessed by the Rorschach and, as such, should possess stability reliability.

Forensic examiners addressing work-related issues might also note that the Sultan et al. (2006) findings are similar to the stability of job performance measures. In a recent meta-analysis, Sturman, Cheramie, and Cashen (2005) found that over a 6-month retest interval, the temporal consistency of objective job performance measures was .45. For both objective and subjective measures of job performance, consistency was .56.

In a summary of the research data available at the time, Viglione and Hilsenroth (2001) reported that CS stability was adequate or better in all respects, especially in the context of comparing Rorschach findings to other personality tests. Revisiting the data about other tests leads to the conclusion that the level of stability reported by Sultan is similar to that reported for the MMPI in a meta-analysis over a 1-year period (Mauger, 1972; Stone, 1965; Sines, Silver, & Lucero, 1961; all as cited in Dahlstrom, Welsh, & Dahlstrom, 1975; Milott, Lira, & Miller, 1977; Ryan, Dunn, & Paolo, 1995). The Sultan stability reliability coefficients are also similar to those reported in a comprehensive meta-analysis of self-report, observer, and performance tests of personality (Roberts & Del Vecchio, 2000), but less than that reported in a more limited and less definitive meta-analysis of eight self-report tests over a 1-year period (Schuerger, Zarrella, & Holtz, 1989). At this point, forensic examiners should be alert to the possibility, based on this one study, that CS scores can be more changeable and responsive to statelike influences than previously thought. In forensic cases, when making dispositional attributions, examiners might consider repeating a Rorschach and other personality assessment measures to more definitively differentiate state and trait influences.

The type of reliability that has received the most attention recently—and one that may be most relevant to forensic practice—is *interrater reliability*, or the consistency of judgments across raters. For the Rorschach, this type of reliability concerns coding (scoring) reliability as well as the reliability of interpretation across test users. We address research in coding reliability because it has received most of the recent research attention. For issues involving interpretive reliability, we refer to Meyer, Mihura, and Smith (2005).

Exner (2003) has primarily presented percentage agreement (%A) between coders as a means of addressing interrater reliability and coding accuracy. Percentage agreement is the proportion of responses in which two raters agree on a code, that is, code a given response parameter the same way. He had required that any code have a %A of 80% to be included in the CS. Weiner (1991) also required that studies submitted to the *Journal of Personality Assessment* meet this %A benchmark for 20 records. For example, for *Human Movement (M)*, if two raters independently code 50 responses and agree 45 times on the presence or absence of *M*, then %A = 90%. However, *M* only occurs in about one fifth of responses from adults, so that two raters are expected to agree, by chance, about 70% of the time. This high incidence of chance agreement occurs largely because raters with knowledge of base rates could agree that *M* is absent even if they randomly scored *M*. Accordingly, %A does not consider base rates and chance agreement, and it overestimates reliability for single scores, so that it has been subjected to criticism (Wood et al., 1996).

Although true in some respects, criticism of %A has been greatly overextended to all types of coding and response combinations. It is not nearly as problematic for response segments that have multiple choices for codes. The term *response segment* refers to a coding category, for example, determinants or content. To achieve agreement for determinants, one would have to agree on all determinants in a given response (e.g., *FT.CF* and *FT.CF* represents an agreement, whereas *FT.CF* and *FT.FC* do not.) Obviously, chance agreement for response segments (e.g., determinants or content) is much lower than it is for individual codes. For determinants and content chance %A is about 20%; for all special scores chance %A is about 40%; for location, *DQ*, and *FQ* chance %A is about 30%–50% (Meyer, 1997a, 1997c). Thus, it is mathematically impossible to discount 80% agreement for response segments among 20 records as being due to chance.

Nevertheless, there are preferred statistics that do take base rate into consideration, namely, kappa for response level data and the intraclass correlation (ICC) for protocol level data. Kappa is appropriate for nominal or categorical variables, as represented by individual Rorschach scores or codes. Accordingly, if one wanted to evaluate how reliably two raters or two teams of raters scored *Texture* (*T*) on a response by response basis, one could use kappa. This statistic could, for example, estimate reliability for the presence or absence of any form of *T*. Alternatively, it could detect whether or not raters reliably distinguished between *FT*, *TF*, *T* and no *Texture*.

Whereas kappa is applied to response-level variables, the ICC is applied to dimensional variables at the protocol level. In other words, if one wanted to evaluate the reliability of the sum of all *T* responses, *X* – %, or the Suicide Constellation across records, ICC is ideal. Score levels and interpretation of ICC are equivalent to kappa, and it is an excellent statistic for Rorschach summary scores (i.e., those types of scores that are found on the CS Structural Summary). Given that the preponderance of interpretive inferences emerges from the Structural Summary, the ICC is more related to the foundation of interpretation and how the test is used in practice. Kappa, however, may be more useful in training raters and evaluating the ease to which a new score can be coded.

Janson (Janson & Olsson, 2001, 2004) has introduced a new statistic called iota. As a more general statistic, it can be used instead of kappa or ICC. It is a multivariable extension of kappa and can be applied to response level variables (e.g., individual codes), response segments (e.g., determinants or contents in a given response), or even all the codes of a response or protocol in its entirety. Like the ICC, it also can be applied to dimensional or protocol level variables. Accordingly, it has considerable flexibility and is recommended for research and training. For training or forensic applications, one could measure the reliability or agreement of two raters for a single record across all scores.

Given that kappa, ICC, and iota are more demanding types of reliability statistics, the benchmarks for interpreting their magnitude differ from those associated with Pearson *r* and %A. Kappa, ICC, and iota at .75 or above is considered excellent, .60 and above good, and .40 and above fair (Cicchetti, 1994; Shrout & Fleiss, 1979).

There are four meta-analyses addressing Rorschach interrater reliability. Two related studies address CS reliability (Meyer, 1997a, 1997c; Meyer et al., 2002) and the others address two non-CS scales, the Rorschach Prognostic Rating scale and the Rorschach Oral Dependency scale (see Meyer, 2004). Meyer (2004) compared these interrater reli-

ability data to all the other meta-analyses of interrater reliability available at the time. Comparisons with these other types of judgments allow forensic psychologists—or indeed an attorney, judge, or jury—to derive a “gut feel” sense of how the reliability of the Rorschach fares.

These interrater reliability comparisons are presented in Table 2–1.² Reliabilities are presented separately for scale-level judgments and item-level judgments. With each type of judgment, the average reliability coefficient is listed along with the number of pairs of ratings summarized. For the Rorschach, scale data corresponds to protocol level summary scores, whereas item data corresponds to coding determinations made on individual responses. A consistent pattern is that scale reliabilities exceed item reliabilities because random errors tend to cancel each other out when items are aggregated to form scales. The overall reliability of the Rorschach CS and Rorschach Oral Dependent scale are excellent with summary score coefficients about .90 and response-level judgments in the range between .80 and .85. The Rorschach Prognostic Rating scale reliability is not as high, with $r = .84$ for summary scores, but still more than adequate.

Thus, one must conclude that the Rorschach interrater reliability is good to excellent and compares favorably to a wide range of determinations made in psychology and medicine. Attorneys, judges, or juries may be very interested to know that the Rorschach raters agree much more than do superiors’ evaluations of job performance, surgeons/nurses’ diagnoses of breast abnormalities, and physicians’ estimations of the quality of medical care from record review, all of which are subject to considerable disagreement and inconsistency across raters. Rorschach CS and Oral Dependent scale coding determinations have the same degree of agreement or reliability as do simple, physical measurements in medicine. For example, Rorschach coding is as reliable as estimating the size of the spinal canal and spinal cord from MRI, CT, or x-ray scans, or counts of decayed, filled, or missing teeth in early childhood. These comparisons are consistent with the conclusion that Rorschach coding for the trained examiner is typically a relatively straightforward process, one in which consistency and agreement are attainable across raters.

Clearly, the answer to the question, “Do we code reliably?” is yes, as well-trained and motivated raters code reliably. However, there are limitations. Several studies reported that standard errors of reliabilities of low base rate variables are large so that their reliability estimates are erratic (Acklin, 1999; Acklin, McDowell, & Verschell, 2000; Meyer, 1997a, 1997c; Meyer et al., 2002; Viglione & Taylor, 2001). Low base rate variables, for example, sex, reflections, color projection, or refined variables³ can be loosely defined as occurring on the average once or less often per record. This is a

²Meyer (2004) compared types of statistics, contrasting r with kappa or the ICC. Across 16 topics that provided both types of statistics, the mean kappa/ICC was .70 and the mean r was .74. Because these differences are not large, the findings for those 16 topics were combined in our version of the table. Our table also differs slightly from Meyer’s (2004) in that it presents two coefficients for job selection interviews (one for joint interviews and one for separately conducted interviews), rather than just a single undifferentiated coefficient.

³Weiner (2001) described refined variables as coding combinations that encompass multiple categories, so that $M-$, $WS+$, or M^a with *Pure H* are refined variables. In contrast, M , W , and H are unrefined variables. He stated that refined variables are more likely to demonstrate validity in research. There is not a great deal of research with refined variables, presumably because large samples are needed.

TABLE 2-1

Meta-Analyses of Interrater Reliability in the Psychological and Medical Literature

		<i>n(k-1) = independent pairs of judgments</i>		<i>Reliability r/κ/ICC</i>	
<i>Target reliability construct</i>		<i>Scale</i>	<i>Item</i>	<i>Scale</i>	<i>Item</i>
1. Measured bladder volume by real-time ultrasound			360		.92 ^b
2. Measured size of spinal canal and spinal cord on MRI, CT, or X-ray		200	86	.90 ^a	.88 ^a
3. Count of decayed, filled, or missing teeth (or surfaces) in young children		113	237	.97 ^a	.79 ^c
4. Rorschach Oral Dependency Scale scoring		974	6,430	.91 ^b	.84 ^c
5. Scoring the Rorschach Comprehensive System:	Summary scores	784		.91 ^b	
	Response segments		11,518		.86 ^c
	Scores per response		11,572		.83 ^c
6. Neuropsychologists' test-based judgments of cognitive impairment			901		.80 ^c
7. Hamilton Depression Rating Scale scoring from joint interviews ^d		3,847	495	.86 ^b	.71 ^b
8. Level of drug sedation by ICU physicians or nurses		1,116	165	.86 ^b	.71 ^c
9. Functional independence measure scoring (joint and separate interviews)		1,365	1,345	.91 ^c	.62 ^c
10. TAT Personal Problem-Solving Scale scoring		385		.85 ^b	
11. Rorschach Prognostic Rating Scale scoring		472		.84 ^a	
12. TAT Social Cognition and Object Relations Scale scoring		934		.82 ^b	
13. TAT Defense Mechanism Manual scoring		743		.80 ^b	
14. Hamilton Anxiety Rating Scale scoring from joint interviews ^d		752	214	.80 ^b	.72 ^c
15. Borderline personality disorder (joint and separate interviews)	Diagnosis	402		.82 ^c	
	Specific symptoms		198		.64 ^c
16. Signs and symptoms of temporomandibular disorder (separate exams)		192	562	.86 ^c	.56 ^c
17. Hamilton Depression Rating Scale scoring from separate interviews		1,012	597	.82 ^b	.52 ^b
18. Therapist or observer ratings of therapeutic alliance in treatment		(<i>S</i> = 31)		.78 ^a	
19. Job selection ratings by joint interviews		9,364		.77 ^a	
20. Hamilton Anxiety Rating Scale scoring from separate interviews		268	208	.76 ^b	.58 ^c
21. Axis I psychiatric diagnosis by SCID in joint interviews		216		.75 ^c	
22. Type A behavior pattern by structured interview		(<i>S</i> = 3)		.74 ^a	
23. Axis II psychiatric diagnosis by semistructured joint interviews		740		.73 ^c	
24. Personality or temperament of mammals (variable observations)		151	637	.71 ^a	.49 ^a

		<i>n(k-1) = independent pairs of judgments</i>		<i>Reliability r/κ/ICC</i>	
Target reliability construct		Scale	Item	Scale	Item
25. Visual analysis of plotted behavior change in single-case research			1,277		.57 ^b
26. Editors' ratings of the quality of manuscript reviews or reviewers			3,721		.54 ^b
27. Presence of clubbing in fingers or toes ^c			630		.52 ^c
28. Stroke classification by neurologists			1,362		.51 ^c
29. Child or adolescent problems:	Teacher ratings	2,100		.64 ^a	
	Parent ratings	4,666		.59 ^a	
	Externalizing	7,710		.60 ^a	
	Internalizing	5,178		.54 ^a	
	Direct observers	231		.57 ^a	
	Clinicians	729		.54 ^a	
30. Job performance ratings by supervisors		1,603	10,119	.57 ^a	.48 ^a
31. Axis I psychiatric diagnosis by SCID in separate interviews		693		.56 ^c	
32. Job selection ratings by separate interviews		3,185		.53 ^a	
33. Axis II Psychiatric diagnosis by semistructured separate interviews		358		.52 ^c	
34. Self and partner ratings of conflict:	Men's aggression	616		.55 ^a	
	Women's aggression	616		.51 ^a	
35. Determination of systolic heart murmur by cardiologists			500		.45 ^c
36. Abnormalities on clinical breast examination by surgeons or nurses			1,720		.42 ^c
37. Mean quality scores from two grant panels:	Dimensional ratings		2,467		.43 ^b
	Yes/No decision		398		.39 ^c
38. Job performance ratings by peers		1,215	6,049	.43 ^a	.37 ^a
39. Number of factors in a correlation matrix by scree plots ^f			2,300		.35 ^c
40. Medical quality of care as determined by physician peers			9,841		.31 ^c
41. Job performance ratings by subordinates		533	4,500	.29 ^a	.31 ^a
42. Definitions of invasive fungal infection in the research literature			21,653		.25 ^c
43. Research quality by peer-reviewers:	Dimensional ratings		31,068	.25 ^b	
	Yes/No decision		4,807		.21 ^c

Note. Adapted from Meyer (2004), which provides a complete description of the meta-analytic data sources contributing to this table. ICC = intraclass correlation, ICU = intensive care unit, S = number of studies contributing data, SCID = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), and TAT = Thematic Apperception Test.

^aPearson's *r*. ^bCombination of *r* and κ or agreement ICC. ^c κ or agreement ICC. ^dCategory includes videotaped interviews and instances when the patient's report fully determined both sets of ratings (e.g., identical questions in written and oral format). ^eOne study produced outlier results ($\kappa = .90$) relative to the others (κ range from .36–.45) so the results should be considered tentative.

^fFinding should be treated cautiously because agreement varied widely across studies, with values below .10 in several samples but above .70 in several others.

statistical issue and one would need large samples to accurately estimate reliability for low base rate variables.

In addition, there are some codes for which reliabilities are lower so that they are presumably more of a challenge to code accurately. Table 2–2 identifies these CS codes associated with lower reliabilities in multiple research reports. Forensic examiners should pay special care to code these variables accurately, consistent with CS principles. Some examiners have protocols in high-stakes cases blindly rescored by a colleague. Viglione wrote *Rorschach Coding Solutions* (2002) to address these and other coding challenges. Along with the workbook (Exner et al., 2001) and volume I text (Exner, 2003), it is a good resource to consult to eliminate rater drift from CS standards. Indeed, interrater reliability is not a fixed property of the score or instrument. In forensic practice, this means that what counts is the reliability of the person who coded the protocol, not the general reliability found in the literature. As such, it would behoove forensic examiners to document that they have achieved good interrater reliability with another expert rater.

In the forensic arena, the single most problematic implication of the data on variables with lower reliabilities might be the possibility of over coding *ALOG*, *DR*, and *FQ*—so as to overestimate pathology, thought disorder, and the likelihood of a psychotic or schizo-

TABLE 2–2
CS Codes Decisions with Lower Reliabilities in Some Studies

Developmental Quality

DQv and *DQv/+*

Form Dominance

FC vs. *CF* vs. *C*

Form Shading vs. *Shading Form* vs. *Shading*

Shading Subtypes

Y vs. *T* vs. *C'* vs. *V*

Form Quality

Occasionally *FQ* subcategories, especially *FQu*

Failure to code or neglect of *FQ+*

Contents

Art, *Ay*, *Sc*, *Bt* vs. *Na* vs. *Ls*, *Id*

Special Scores

DV vs. *INC*

ALOG vs. no special score, coding too many *ALOGs*

CONTAM vs. *INC*

PER or *DR* vs. task comment, coding too many *DRs*

Level 1 vs. *Level 2*

phrenic diagnosis. In forensic assessment, such an error might translate to underestimates of, for example, sanity, capacity, culpability, or parenting ability. Some comfort can be drawn by the fact that the research indicates that the summary scores for cognitive special scores, *WSum6* and *Sum6*, generally demonstrate better reliability than do the individual scores (e.g., *DV2* or *ALOG* individually). This superior reliability is important because interpretation is primarily based on these summary scores rather than on individual cognitive special score codes.

Research reports from around the world (Erdberg, 2005; Viglione, 1999; Viglione & Hilsenroth, 2001) also reveal that the CS is transportable to other languages and cultures and that coding reliability is very similar to the results from the meta-analyses. For the most part, those codes that achieve lower or more variable reliabilities in U.S. samples are the same as codes that are more variable in the international samples (Exner et al., 1999).

Another issue or complication is that most reliability research studies generally use raters who work or train in the same setting. If local guidelines develop to contend with scoring ambiguity, agreement among those who work or train together may be greater than agreement across different sites or workgroups. Thus, existing reliability research may then give an overly optimistic view of reliability across sites or across forensic examiners working independently.

In a preliminary presentation, Meyer, Viglione, Erdberg, Exner, and Shaffer (2004) examined this across site interreliability issue by having 40 randomly selected protocols from Exner's new CS nonpatient reference group sample and 40 protocols from Shaffer, Erdberg, and Haroian's (1999) from a California (CA) sample recoded by a third group of trained raters. This third group, advanced graduate students supervised in Viglione's lab, were blind to the original coding, the origin of the samples, and the nature and purpose of the study. The coding assigned by the original sites was compared to the coding assigned by this single additional site and yielded an across site median ICC of .72, an acceptable level of reliability in the good range.

These across site results can be contrasted with within site data sets, that is, samples coded by raters working in the same setting. We have three such relevant within site research reports available to us: (a) the meta-analysis data in Table 2-1, (b) a large international sample (Erdberg, 2005), and (c) a smaller sample from Viglione's lab. All report greater reliabilities than our across site median ICC of .72. As noted earlier, the Table 2-1 meta-analysis yields a reliability estimate for summary scores of .91. Erdberg (2005) compiled 467 protocols from 17 internationally collected nonpatient reference samples. The initial median within site ICC from the international sample was .82, a reliability estimate in the excellent range. Although the pool of protocols was collected from many different countries, all the scoring for each protocol took place locally by examiners who trained together. Thus, these data provide a reasonable sample of within site scoring reliability across the world and attest to the cultural adaptability of the test and its administration procedures. The third within site reliability estimate is pertinent because it is from the same lab that provided the across site coding. Viglione and Taylor (2001) reported a median within site reliability of .92 for 84 protocols.

Although the across site reliability estimates are preliminary, these findings suggest that there are complexities in the coding process that are not fully clarified in the standard

CS training materials (Exner 2003; Exner et al., 2001). As a result, training sites (e.g., specific graduate programs) may develop guidelines for coding that help resolve these residual complexities but they may not generalize well to other training sites. Forensic examiners may find it helpful to consult an advanced coding text (Viglione, 2003) or to practice coding with colleagues trained in a different setting.

NORMATIVE DATA: HOW ADEQUATE ARE CS NORMS?

Rorschach normative reference group data have been criticized for pathologizing examinees. Wood et al. (2001b) compared CS reference values on 14 selected variables to the values reported in 8 to 19 comparison samples from the literature. They reported small to very large differences (Cohen's d from .18 to 1.67)⁴ for the 13 variables where mean differences could be computed.⁵ All differences were in the more pathological or problematic direction for the comparison samples. There were nine variables for which these differences were at least medium size: (a) lower values for $X+\%$, Afr , FC , P , $WSumC$, and $Pure H$; and (b) higher values for reflections, $X-\%$ and Y . Variability of these scores (i.e., the SD) was greater than in the original CS sample—a worrisome finding because it might suggest that current confidence intervals and normative interpretive ranges are too narrow.

The samples in the Wood et al. report were portrayed as nonpatient or normative reference samples but had serious problems and were not fully representative of nonpatients (Meyer, 2001). From a total pool of 32 studies, 22 samples (69%) did not have a procedure to exclude patients or low functioning or disturbed individuals; 16 (50%) samples were college students or the elderly; one had a mean R of 15, whereas another had a mean R of 39, suggesting atypical administration; respondents in one sample were held motionless with electrodes on their head; and just two samples had data for all 14 scores. Obviously, these samples are not representative of nonpatients and are not a good source for comparisons. Nevertheless, it is hard to dismiss these findings totally, as others (Viglione & Hilsenroth, 2001) have examined similar data and found that the distributions for form quality and R appeared to diverge to some degree from CS expectations.

To investigate these normative issues with a better comparison sample, Meyer (2001) contrasted Exner et al.'s (1993) original CS adult normative reference sample to a composite of 2,125 protocols from nine adult samples presented in Erdberg and Shaffer's (1999) symposium on international CS reference data. These samples (which include the Shaffer et al., 1999, sample from the United States) provided data on all CS variables and encompassed great variability and thus generalizability across subject selection procedures, examiner training, examination context, language, culture, and national boundaries. Across 69 composite scores from the lower portion of the Structural Summary, distributions for 49 variables were similar in the original CS sample and international

⁴Cohen's d is an effect size measure for comparing two groups. It basically is the difference between the means of the groups in standard deviation units, i.e. the z -score for the differences. For example, a difference of 10 IQ points should result in a Cohen's d of 0.67.

⁵The 14th variable was EB style, a categorical variable for which means could not be computed.

data, a finding consistent with the conclusion that the original CS norms are generally adequate. These data, in addition to the similarities between U.S. and international findings for interrater reliability, again indicate considerable cultural and international adaptability of the Rorschach. One can adapt it to different cultures, languages, and regions, and the test behaves largely as it does in the United States.

Nevertheless, some differences between the CS sample and the composite of international samples persist, so that we need to adjust our normative expectations. International samples have higher scores for *Dd*, *S*, *FQu*, *FQ-*, *Hd*, (*Hd*), and *Sum6*, and lower scores for *WSumC*, *EA*, *FQo*, *P*, *COP*, *AG*, and *Afr*. In all cases, the CS norms come across as “healthier.” In other words comparison to the CS norms would lead to more pathological interpretations than would comparisons to the international norms. Accordingly, normative expectations for these and for variables that subsume them (e.g., *X-%* for *FQ-*) need to be adjusted. More specific recommendations are given here.

A reasonable question becomes, “Why do the original CS norms look healthier than other normative approximation samples?” The CS respondents were recruited largely through work, unions, or social organizations. Compensation was in the altruistic form of contributions to charity in name of the place of business or organizations, so that respondents were not paid themselves as volunteers. Thus, differences could be due to situational differences or examination context. The CS respondents may feel that their responses matter more than do volunteers in other studies, so that they may “tidy-up” their answers a bit more through filtering in the response process (Exner, 2003). One might speculate that making the examination matter to the respondent is a better approximation of the use of the test in the real world, and thus a better contrast sample. Alternatively, these recruiting practices involving employment and social involvement might lead to a selection bias in terms of attracting healthier and better adapted individuals to volunteer. Indeed, the literature indicates that the garden variety volunteers tend to possess problematic characteristics and are less well-adapted (Berman, Fallon, & Coccato, 1998; Rosenthal & Rosnow, 1975).

Other explanations of the observed health in the CS norms include differences in administration or coding. There are considerable differences between the initial CS form quality tables first published in 1974 (Exner, 1974) and the current version (Exner et al., 2001), with most of these differences resulting in more *FQ-* and fewer *FQo* responses (Meyer & Richardson, 2001; Viglione, 1989). In addition, criteria and examples for other coding distinctions have changed or been elaborated on over time in ways that alter the benchmarks for assigning a score (Meyer, 2001). Another explanation is simple aging of the norms and increasing mental health difficulties over time.

To address these normative issues, Exner started collecting a new adult normative reference group in 1999 (Exner, 2002; Exner & Erdberg, 2005). This new sample, which is now approaching 500 respondents, was collected largely in the same way as the original CS sample, but there are some differences. The new sample involves the workplace or organizations less formally, so that individuals may feel that they represent themselves instead of an organization. For example, charity donations are made in a respondent’s name rather than the organization’s name. In the original CS sample, a manager acted as the liaison between examiners and data collection sites and actually solicited respondents. In

the new sample, examiners recruit participants on their own. Respondents are now excluded due to “prolonged or significant history” of psychotropic medications or illegal drug use.

Exner and Erdberg (2005) provided data for 450 of the individuals in this sample. The more important differences in terms of mean differences and interpretive cutoffs between the two groups are summarized in Table 2–3. The selected frequencies differ by 5% in the two samples. As can be seen, form quality is less optimal in the new reference sample. There are fewer *Populars*, more special scores, and more of the serious *Level 2 Cognitive Special Scores*. There is less color overall and more color-dominated relative to form-dominated color responses. The *Afr* is lower, there is a notable increase in space responses, and there is a lower frequency of both cooperative and aggressive movement scores. In addition, it is more common for passive movement to exceed active and for the *Depression Index (DEPI)* to be elevated. Although the frequencies remain low, it is worth pointing out that the *SCON* did not exceed 7 in any of the old 600 records, but it does for 11 of the current 450 records. These changes incorporate many of the same variables discussed earlier as divergences between the old CS samples and the international composite pool of references samples collected by other researchers.

Another notable finding is that the standard deviation for *R* is 5.68, as compared to 4.40 in the original CS sample of 600. This change may be problematic because this increased variability of *R* should be associated with more variability for all other scores. Indeed, the great majority of *SDs* is larger in the new sample as compared to the original. This greater variability means that interpretive postulates need to have wider confidence intervals (i.e., the range of expected scores is broader).

Although the new CS reference sample reduces some of the differences with the composite of international reference samples, it does not eliminate them. For example, the new CS sample still has means for *Dd* and *X-%* that are lower and means for *X+ %* and *EA* that are higher than other reference samples.

The study that initiated the concerns about the original CS normative reference sample is mentioned in the previous reliability discussion and was published by Shaffer, Erdberg, and Haroian (1999). Its respondents had MMPI–2 *T* score means at approximately 50 and WAIS–R IQs of about 100, thus at normative values. Most Rorschach values were consistent with the original CS normative reference group, but values for the variables already identified as diverging from normative expectations also demonstrated such divergence in this sample. The Shaffer et al. California (CA) sample also differed from both the original and new CS samples in terms of overall complexity. The mean for *R* in the Shaffer et al. sample is only 20.8 versus 23.36 for the new CS sample, and the *Lambda* is 1.22 (median = .75) versus .58 (median = .47) in the new CS sample, with 41% of the Shaffer et al. sample having a *Lambda* greater than .99 versus 14% in the new CS sample. These findings indicate that the Shaffer et al. sample was not very productive and they produced relatively simplistic records in comparison to the CS and other samples included in the international group.

Along with our interrater reliability investigations with these samples (Meyer et al., 2004), we have conducted some initial investigations into the differences between the CA normative reference sample and the new CS reference sample. In this research, we ex-

TABLE 2–3
Illustrative Changes in the New Target Reliability Construct Versus Original CS Normative Reference Samples

<i>Domain/Score</i>	<i>Original 600</i>	<i>New 450</i>
<i>Quality of Perception and Thinking</i>		
<i>X+%</i>	.77	.68
<i>Xu%</i>	.15	.20
<i>X–%</i>	.07	.11
<i>X+%</i> < .55	2%	12%
<i>X%</i> > .20	22%	45%
<i>X–%</i> > .20	3%	10%
<i>XA%</i> > .89	74%	45%
<i>WDA%</i> < .85	5%	16%
<i>P</i> > 7	31%	18%
<i>Sum6</i>	1.91	2.54
<i>WSum6</i>	4.48	7.12
<i>Lvl2 SS</i> > 0	6%	13%
<i>Color</i>		
<i>FC</i> > <i>CF</i> + <i>C</i> + 2	25%	15%
<i>FC</i> > <i>CF</i> + <i>C</i> + 1	41%	26%
<i>CF</i> + <i>C</i> > <i>FC</i> + 1	12%	26%
<i>CF</i> + <i>C</i> > <i>FC</i> + 2	4%	14%
<i>Extratensive</i>	38%	31%
<i>Miscellaneous</i>		
<i>S</i> > 2	14%	38%
<i>DQv</i> > 2	12%	2%
<i>T</i> > 1	11%	17%
<i>Ego</i> < .33	13%	20%
<i>Ego</i> > .44	23%	30%
<i>Afr</i> < .40	3%	9%
<i>Afr</i> < .50	11%	24%
<i>Zd</i> < 3.0	7%	14%
<i>Intell</i> > 5	2%	8%
<i>COP</i> = 0	17%	11%
<i>AG</i> = 0	37%	44%
<i>Hd</i>	.84	1.14
(<i>Hd</i>)	.21	.62
<i>DEPI</i> > 4	5%	14%
<i>p</i> > <i>a</i> + 1	2%	10%
<i>Mp</i> > <i>Ma</i>	14%	23%

amined whether coding conventions might contribute to the differences between the data sets. More specifically, we wondered if CS–CA differences would be reduced when records from both samples were recoded at a third site. If the Shaffer et al. records were coded according to somewhat different benchmarks than Exner’s protocols, the differences between the two samples would be reduced if records from both samples were coded by a third group.

To address this question and as described earlier, we obtained 80 protocols from both the CA and CS samples. These 80 protocols were then recoded by a new group of examiners who were trained together in one setting. We then computed two sets of difference scores, using Cohen’s d as the effect size index. The first difference score compared mean scores for the CS and CA samples using the original coding from the two sites. The second difference score compared the means for the CS and CA samples based on the new coding. Because the new coding was done by raters who trained together within one site, it eliminates the potential influence of site-specific differences in coding conventions. We anticipated that the initial differences would decrease with the revised coding; that is, the second set of differences from single site scores would be smaller than the first set of differences generated from separate sites.

Initially, with the original CS and CA scoring, across 129 structural summary variables the differences for 36 scores (28%) were moderate to large, with d values greater than .40 or less than $-.40$. Thus, the normative expectations differed for 36 of the 129 variables in our randomly selected protocols from both samples. However, with the new single site coding, there were only three means (2%) that remained different at this magnitude. Thus, almost all the seemingly important differences between the new CS sample and the CA sample disappeared when the protocols were rescored by a different group. In general, for most variables, our new coding split the difference between the CS sample and Shaffer et al. sample. By and large, the groups now were much more similar: Relative to the original scores, with the new coding, the CS sample looked less healthy than before and the Shaffer et al. sample looked healthier than before.

However, there were instances when the new scores were more similar to one of the reference samples than the other. For complexity variables (*Lambda*, *DQ+*, *Blends*, etc.) and for *Dd*, the values from the rescored protocols more closely resembled the CS reference sample than the CA sample. Furthermore, with the possible exception of *Dd*, the CS reference sample is more similar than the CA sample to the internationally collected reference samples for these particular complexity scores. In contrast, form quality values from the rescored protocols were more in line with the Shaffer et al. CA sample than the CS sample. Equally important, the CA reference sample is more similar than Exner’s CS sample to the form quality values observed in other U.S. and international reference samples.

The overall findings suggest that site-specific coding practices may contribute in important and previously unappreciated ways to some of the seeming differences across normative approximation samples. In addition, these initial data suggest a convergence between the CS and CA sample, with the international normative sample. These suggestions are hypotheses that need to be tested with additional samples and coding sites.

There is less research into the suitability of the CS normative reference samples for children. In a study similar to the Shaffer et al. (1999) study and from the same group of researchers, Hamel, Shaffer, and Erdberg (2000) reported on 100 6- to 12-year-old children. This research has also attracted a lot of attention. To establish this group as a normative reference sample, their parents identified them as average to psychologically healthy on a commonly used multidimensional rating scale. However, once again, the Rorschach data diverged from the CS normative reference groups in some respects. In many ways, the differences are similar to those found in the adult normative reference samples. Like the adult samples, Hamel et al. found more distorted form quality values, less color, more use of unusual blot locations, elevated rates of dysfunction on the constellation indices, and less complexity. However, unlike the adult CA versus CS sample differences, the reference values observed by Hamel et al. tended to be more extreme. For instance, the average *Dd* was 8.3, the average *X-%* was .41, 62% of the sample had an elevated *SCZI* (value of 4 or more), and the median *Lambda* value was 1.14 (mean = 1.91).⁶

Although Hamel et al. (2000) took a careful and conscientious approach to their study, several characteristics of the sample suggest it is idiosyncratic and challenge its trustworthiness as a contemporary CS reference sample for children. First, all administration and coding was done by a single examiner, so that generalizability may be limited. Second, for interrater reliability, %A was reported in an unusual way.⁷ This method would lead to the undetected possibility of coding inaccuracies for determinants, contents, and special scores. Also, in comparison to most research reports, %A was low for location and form quality. Third, the authors strongly emphasized the necessity for precision in documenting blot areas on the location sheet that appear to drift from CS standards:

Students should be clearly taught to very carefully and accurately encircle the precise portion of the blot utilized by the examinee ... to enable any other clinician to precisely replicate the coding for location. The precision of location cannot be overemphasized; not only does the location code clearly depend upon an accurate location sheet, but so do other segments of the coding. Form quality and Popular are heavily dependent upon location. A Form Quality of ordinary can easily be altered to unusual or minus on the basis of location alone. (Hamel et al., 2000, p. 291)

If carried through in administration, this emphasis on precision may distort the interaction between the examiner and respondent in the inquiry and also influence the documentation of response areas on the location sheet. Moreover, along with the slack in interrater reliability, it may be related to the extraordinary *Dd* elevation. Excessive *Dd* locations, in turn, could negatively affect form quality codes and *Popular* responses, as well as *SCZI* scores. Accordingly, we do not recommend using the Hamel et al. (2000) sample as a normative approximation sample.

Nevertheless, other samples suggest clinicians should be cautious about using the existing CS reference values for children. Besides Hamel et al. (2000), other child and ado-

⁶Because of the skew inherent with *Lambda*, we recommend that median *Lambda* values be reported and that *Pure F%* (*Pure F/R*) be used (Meyer, Viglione, & Exner, 2001).

⁷It should be pointed out that the Hamel reliability data was derived using an across site coding procedure where the comparison scoring was done by a person trained in the same lab that did the rescoring for the Meyer et al. (2004) across site reliability study.