Annu. Rev. Psychol. 2002. 53:519–43

CLINICAL ASSESSMENT*

James M. Wood[1], Howard N. Garb[2], Scott O. Lilienfeld[3], and M. Teresa Nezworski[4]

[1]*Department of Psychology, University of Texas at El Paso, El Paso, Texas 79968; e-mail: jawood@utep.edu*

[2]*Behavioral Health, V.A. Health Care System, Highland Drive, Pittsburgh, Pennsylvania 15206-1297 and Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213; e-mail: Garb.Howard_N@pittsburgh.va.gov*

[3]*Department of Psychology, Emory University, Room 206, Atlanta, Georgia 30322; e-mail: slilien@emory.edu*

[4]*Callier Center for Communication Disorders, School of Human Development, University of Texas at Dallas, 1966 Inwood Road, Dallas, Texas 75235-7298; e-mail: nezworsk@utdallas.edu*

Key Words managed care, multicultural assessment, forensic assessment, clinical judgment, computerized assessment

■ **Abstract** Are clinical psychologists' assessment practices cost-effective? Are they scientifically sound? Are they fair and unbiased? Financial pressures from managed care interests, recent developments in the law, and multicultural issues are forcing the profession to confront these hard questions regarding accountability. Our review discusses the important changes that have begun to alter the field of personality assessment and describes recent research on clinical judgment and its implications for the future. We conclude that clinical psychology can adapt to future conditions by developing assessment strategies that are economical, scientifically sound, and culturally sensitive.

CONTENTS

INTRODUCTION 520
THE CHANGING WORLD OF CLINICAL ASSESSMENT 520
Managed Care 520
Multicultural Clinical Assessment 521
Forensic Assessment 523
The Controversy Over the Comprehensive System for the Rorschach 525
Self-Report Tests 527
CLINICAL JUDGMENT AND DECISION-MAKING 530
Diagnosis 530

*The US government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

Case Formulation 531
Treatment Decisions 532
Prediction of Violence and Detection of Deception 533
Clinical Judgment, Computers, and Mechanical Prediction 533
FUTURE DIRECTIONS 534

INTRODUCTION

Economic, cultural, legal, and scientific forces are causing profound changes in clinical assessment. The financial pressures of managed care have already altered the assessment practices of many clinicians, and recent developments in the law have encouraged the critical examination of assessment techniques that are used in forensic settings. The rapidly changing ethnic composition of the United States presents new challenges to old practices.

In the first half of this review we discuss the important changes that have begun to alter the field of personality assessment. In the second we discuss progress in research on clinical judgment and its implications for the future. A theme that runs through our review is accountability. Clinical psychologists are being held increasingly accountable for their assessment practices. Forces from both inside and outside the profession are posing hard questions: Are psychologists' assessment practices cost-effective? Are they scientifically sound? Are they fair and unbiased? The answers are already reshaping both the science and the practice of psychological assessment.

THE CHANGING WORLD OF CLINICAL ASSESSMENT

Managed Care

The greatest challenge confronting practicing psychologists over the past decade has been economic. In the era of managed care, clinical practitioners have faced intense and sometimes unreasonable pressure to reduce or defend their services. Psychologists in private practice report a variety of concerns, including loss of income, excessive paperwork requirements, and ethical dilemmas (Phelps et al. 1998).

A survey of 137 psychologists by Piotrowski et al. (1998) gauged the effects of managed care on assessment practices: 55% of respondents reported that in response to managed care they were spending less time giving tests, were using fewer tests, or had discontinued testing altogether (see also Archer & Newsom 2000). Piotrowski (1999) concluded that most psychologists have coped with the pressures of managed care in two ways: (*a*) Some continue to use the same tests as in the past, but less frequently overall, whereas (*b*) others have selectively abandoned tests that are especially time consuming (e.g., the Wechsler Intelligence scales, Rorschach) while continuing to use briefer instruments.

Piotrowski (1999, pp. 792–93) somberly predicted that "economic reality will guide practice in assessment" and "the comprehensive test battery . . . will become a

moribund clinical activity." Groth-Marnat (1999), who shared Piotrowski's vision of the future, suggested ways that researchers and clinicians can respond proactively to the financial considerations of managed care. For example, Groth-Marnat suggested that psychologists develop and promote assessment approaches that (*a*) focus on diagnostic issues and client characteristics that are most clearly linked to treatment choice and outcomes, (*b*) reduce the risk of negative outcomes and litigation (e.g., assessment of danger to self or others), (*c*) identify conditions that, when correctly assessed, are likely to result in cost savings, (*d*) are time efficient, and (*e*) integrate treatment planning, progress monitoring, and outcome evaluation. Groth-Marnat urged researchers to demonstrate that their instruments meet such financial criteria as cost-effectiveness, cost-benefit, and cost-containment.

If Piotrowski (1999) and Groth-Marnat (1999) are correct, clinical assessment practices must change if they are to survive the challenge of managed care. Psychologists are already taking a closer look at the cost-effectiveness and utility of traditional assessment techniques, some of which require an excessive amount of professional time and expense for very little tangible benefit (e.g., the Rorschach) (Groth-Marnat 1999, Hunsley & Bailey 1999). Such trends dovetail with the increasing emphasis placed on demonstrating the incremental validity of psychological assessment devices (Butcher et al. 1995, Kuncel et al. 2001).

In addition, there are signs that clinical assessment techniques are being developed or reconceptualized to meet the needs of managed care, as Groth-Marnat (1999) recommended. Several notable examples appear in a recent book by Maruish (1999) that focuses on measurement of treatment planning and outcome. One of the tests described in Maruish's book provides a model of how psychological assessment can adapt to changing trends. The Outcome Questionnaire (OQ-45) (Lambert et al. 1998, 1999), a brief and cost-effective measure, was designed to track treatment progress and outcome for patients with a wide variety of diagnoses. The validity research on the OQ-45 completed thus far is very promising (Lambert et al. 1996, Mueller et al. 1998, Vermeersch et al. 2000). The success of the OQ-45 shows that psychologists can proactively develop new assessment techniques that are both scientifically sound and compatible with the financial constraints of managed care.

Multicultural Clinical Assessment

The ethnic composition of the United States is rapidly changing. The US Bureau of the Census (2000, p. 17) reported that in 1999, 11% of the US population was Hispanic. The projected figure for 2050 is 24%. Other minority groups, especially those from Asia, are also growing. Within 50 years, approximately half of Americans are expected to be people of color (Hall 1997, US Bureau of the Census 2000).

Attention to multicultural clinical assessment has grown in recent years, as evidenced by the publication of several handbooks on the topic (Cuellar & Paniagua 2000, Dana 2000, Fletcher-Janzen et al. 2000, Suzuki et al. 2000). Nevertheless, the

quality of research continues to lag far behind what is needed. Studies of American minorities and non-Americans are scarce for many popular assessment techniques, including the Millon Clinical Multiaxial Inventory, third edition (MCMI-III), Personality Assessment Inventory (PAI), Rorschach, Thematic Apperception Test (TAT), projective drawings, and the Hare Psychopathy Checklist-Revised (e.g., see Velasquez 1995). Thus, it is unclear whether these instruments can appropriately be used with American minorities and non-Americans. For example, several recent studies (Boscan 1999, Ephraim 2000, Vinet 2000) indicate that Rorschach scores for relatively normal community samples of Mexicans, Central Americans, and South Americans often differ strikingly from the norms of the Comprehensive System for the Rorschach (Exner 1993, 2001b). In light of these findings, there is substantial reason to doubt whether the norms should be used with Hispanic adults and children in the United States (Dana 1998, Lilienfeld et al. 2000, Wood & Lilienfeld 1999). Studies on "slope bias" (Cleary et al. 1975) are sorely needed to examine possible differences in the validity of Rorschach scores across diverse cultural and linguistic groups.

Although evaluation of multicultural assessment is too often neglected, some recent developments may serve as models for future investigations. First, multicultural research on the MMPI (Minnesota Multiphasic Personality Inventory) and MMPI-2 has been much more common, and often more sophisticated, than research on other assessment techniques (see reviews by Butcher et al. 1998, Handel & Ben-Porath 2000, Zalewski & Greene 1996). This body of research has begun to yield findings of substantial practical importance. For instance, there is now convincing evidence that mean MMPI and MMPI-2 scores of US blacks and Hispanics are usually very similar to the US normative values (Hall et al. 1999, Handel & Ben-Porath 2000, Zalewski & Greene 1996). Similarly, a number of studies now indicate that methodologically careful translations of the MMPI and MMPI-2 tend to yield scale means and standard deviations in international samples that are similar to US normative values (Butcher et al. 1998). Furthermore, in a methodological advance, investigators have used item response theory (Embretson & Reise 2000, Santor & Ramsay 1998) to examine possible race bias on the MMPI-2. By obtaining latent trait estimates for the underlying constructs assessed by MMPI-2 factor scales, Waller et al. (2000) showed that "Whites and Blacks can be meaningfully compared on these scales with little fear that obtained group differences are due to measurement bias" (p. 142) (for other applications of item response theory to the MMPI-2, see Childs et al. 2000, Waller 1998). Similarly, three recent studies (Arbisi et al. 1998, McNulty et al. 1997, Timbrook & Graham 1994) have examined the correlation of MMPI-2 scores with external criteria such as case records and therapists' ratings of patients. Reviewing these studies, Greene (2000, p. 482) concluded that in all three "the most striking finding was the high degree of similarity between the blacks and whites on these external criteria."

The normative studies of neuropsychological tests by Ardila and his colleagues in South America (e.g., Ardila et al. 1994, Ostrosky-Solis et al. 1999) have been a second important development in evaluation of multicultural assessment. Although

this review does not cover most developments in neuropsychological testing, the work of Ardila is worth noting because it has systematically identified moderating variables, such as educational level, that should be taken into account in assessments of Spanish-speaking patients (see reviews by Ponton & Ardila 1999, Puente & Ardila 2000).

A third important development has been the publication of several acculturation scales that are suitable for US minorities (Cuellar et al. 1995, Marin & Gamba 1996, Stephenson 2000). Recently developed instruments all conceptualize acculturation as two separate dimensions (i.e., orientation to mainstream US culture and orientation to ethnic culture of origin) rather than as a single bipolar dimension (i.e., acculturated to mainstream US culture versus not acculturated). Acculturation is important in clinical work because, among other reasons, it sometimes moderates the validity of test scores (Cuellar 2000). Perhaps psychologists will eventually be able to routinely assess a client's level of acculturation, linguistic preference, age, and educational attainment and then choose the tests and norms that are most appropriate (Puente & Ardila 2000). Useful advice for researchers in the field of multicultural assessment can be found in Allen & Walsh (2000), Arnold & Matus (2000), Butcher et al. (1998), Handel & Ben-Porath (2000), Okazaki & Sue (1995), and Velasquez (1995, Velasquez et al. 2000).

Forensic Assessment

During the past decade, an increasing number of psychologists have begun to practice in the field of forensic assessment. Custody, competency, and pre-sentencing evaluations probably account for the bulk of forensic assessments. Assessments also play a role in parole decisions, personal injury suits, civil commitments, workers' compensation hearings, Social Security disability evaluations, and even criminal appeals.

At the same time that the field of forensic assessment has been growing, developments in the law have imposed new requirements for accountability. Over the past decade the US Supreme Court has handed down several decisions (Daubert v. Merrell Dow Pharmaceuticals Inc. 1993, General Electric Co. v. Joiner 1997, Kumho Tire Co. Ltd. v. Carmichael 1999) that delineate the legal standards governing the admissibility of scientific and expert evidence in federal courts. The court has described six factors, generally called the "Daubert criteria," that trial judges should consider when deciding whether to admit scientific or expert evidence into court (see discussion by Grove & Barden 1999). Courts in approximately half of the states have also adopted these criteria, which address the following questions: (*a*) Is the theory or technique that forms the basis of the evidence testable? (*b*) Has it in fact been tested? (*c*) Is it generally accepted by the relevant community of scientists? (*d*) Has it been subjected to peer review? (*e*) Does it have a known error rate? (*f*) Are there established standards for its application?

Recent articles have evaluated whether popular assessment techniques meet the Daubert criteria. Serious questions have been raised regarding the legal

admissibility of the Rorschach Comprehensive System (Grove & Barden 1999, but see McCann 1998a), projective drawings (Lally 2001), and the MCMI-III (Rogers et al. 1999, 2000b; but see Dyer & McCann 2000, McCann & Dyer 1996). Over the next decade psychologists and lawyers are certain to debate the scientific adequacy of many popular psychological tests.

In the meantime, the expansion of forensic psychology as a practice specialty has stimulated the development of new assessment instruments that address the specific needs and requirements of the legal system. The most extensively researched of these instruments is the Hare Psychopathy Checklist-Revised (PCL-R) (Hare 1991), including the closely related Psychopathy Checklist: Youth Version (PCL:YV) (Forth et al. 1990). The ratings of the PCL-R and PCL:YV are based on structured interviews, observation, and file reviews and have repeatedly been shown to predict violence in prison and psychiatric populations (Hemphill et al. 1998, Salekin et al. 1996). Briefer measures of psychopathy, such as the Psychopathy Checklist: Screening Version (Hart et al. 1994) and the Psychopathic Personality Inventory (Lilienfeld & Andrews 1996), also seem to predict violence in forensic populations, although the supporting research thus far is modest (Douglas et al. 1999b, Edens et al. 1999, Rogers et al. 2000a). A major advantage of these and other brief measures is their potential applicability outside of prison settings (Levenson et al. 1995, Lilienfeld 1998).

Another interesting development has been the revival of interest in mechanical prediction methods (also known as actuarial or statistical methods) (Meehl 1954). Although supported by an extensive research literature extending over half a century (Grove et al. 2000), mechanical prediction is seldom used for clinical decision-making. However, psychologists and other professionals who conduct forensic risk assessments now routinely employ mechanical prediction instruments, which usually consist of linear combinations of demographic and life-history variables (e.g., number of previous crimes, age of onset of crime). Instruments have been developed and validated to predict violence among criminals and psychiatric patients (Douglas et al. 1999a, Hanson 1998, Monahan et al. 2001, Quinsey et al. 1998), recidivism among juvenile offenders (Wiebush et al. 1995), and repeated child abuse and neglect (Baird & Wagner 2000).

Because litigants in civil or criminal cases may sometimes be motivated to exaggerate symptoms of psychopathology, psychologists in forensic settings have a pressing need for measures of malingering. Interest in this area has boomed in recent years (McCann 1998b, Reynolds 1998, Rogers 1997). Malingering subscales have been developed or validated for several conventional personality tests, including the MMPI-2 (Arbisi & Ben-Porath 1995, 1998; Bagby et al. 2000; Nicholson et al. 1997; Rogers et al. 1994), as well as some cognitive and neuropsychological tests, such as Raven's Standard Progressive Matrices (McKinzey et al. 1999) and the Luria-Nebraska Neuropsychological Battery (McKinzey et al. 1997). In addition, promising work has been reported on stand-alone measures of malingering, including the Structured Interview for Reported Symptoms (SIRS) (Rogers 1995, but see Pollock 1996) and the Test of Memory Malingering (TOMM) (Rees et al.

1998, Tombaugh 1997). Effective malingering measures have yet to be developed for the Rorschach and other projective measures, which can be highly susceptible to faking bad (Exner 1991, Schretlen 1997).

Although work on malingering measures is impressive, considerably more research is needed. Replications by independent researchers are often lacking. In addition, research suggests that informed malingerers can sometimes evade detection (Lamb et al. 1994, Storm & Graham 2000, Walters & Clopton 2000; but see Bagby et al. 2000, Iverson et al. 1995).

In closing, it should be noted that there is one area of forensic practice that remains especially problematic: custody evaluations. As O'Donohue & Bradley (1999) pointed out, psychologists who assess parents in custody evaluations commonly rely on techniques that are subjective, unvalidated, or bear no demonstrated relationship to parental fitness. These authors concluded that many custody recommendations are based on little more than "educated guesswork" (p. 321) or the evaluator's own values and prejudices. We hope that future research will provide a better scientific basis for the practice of custody evaluations. In the meantime, we urge custody evaluators to take concrete steps (e.g., arranging for a colleague to supervise or review cases) to ensure that their work is ethical, unbiased, and based on sound evidence.

The Controversy Over the Comprehensive System for the Rorschach

The 1996 chapter on clinical assessment in the *Annual Review of Psychology* (Butcher & Rouse 1996, p. 91) praised Exner's Comprehensive System (CS) for the Rorschach: "Much of the strength of the Rorschach method in contemporary assessment comes from the broad use of the Exner Comprehensive System (Exner 1991, 1993, 1995; Exner & Weiner 1994), which provides a more reliable and objective basis for interpretation than was available prior to its introduction."

At the time that the chapter by Butcher & Rouse (1996) appeared, there were no indications that the CS was about to become the subject of a fierce controversy. Coincidentally, the chapter was published the same year as the first major published critique of the CS (Wood et al. 1996). In ensuing years, heated debates followed in six other journals, including *Psychological Assessment* (Hunsley & Bailey 1999, Viglione 1999), *Assessment* (Acklin 1999, Garb 1999, Weiner 1999, Wood & Lilienfeld 1999), *Journal of Clinical Psychology* (Weiner 2000; Wood et al. 2000a,b), *Clinical Psychology: Science and Practice* (Aronow 2001; Exner 2001a; Hunsley & Di Giulio 2001; Meyer 2001; Widiger 2001; Wood et al. 2001b,c), *Journal of Personality Assessment* (Bornstein 2001, Gacono et al. 2001, Ganellen 2001, Wood et al. 2001a), and *Journal of Forensic Psychology Practice* (Wood et al. 2001d, Hamel et al. 2001). In addition, a full-issue article in *Psychological Science in the Public Interest* critically reviewed the scientific evidence regarding the Rorschach and other widely used projective techniques (Lilienfeld et al. 2000).

We discuss a subset of issues that are central to evaluations of the scientific and clinical merit of the CS. First, it has become apparent that the CS norms for many important variables (Exner 1993, 2001b) tend to make many normal adults and children appear psychologically disturbed (Hamel et al. 2000; Shaffer et al. 1999; Wood et al. 2001b,c; but see Exner 2001a, Meyer 2001). Even Meyer (Meyer & Richardson 2001), a staunch Rorschach proponent, has presented compelling evidence that the CS norms for form quality were inadvertently based on the wrong scoring rules and have been seriously in error since 1983.

Second, critics of the CS and many proponents now agree that most Rorschach scores bear little or no relation to psychiatric diagnoses (Bornstein 2001; Weiner 1999; Wood et al. 2000a,b). Although a few Rorschach scores are moderately valid for detecting conditions marked by thought disorder, such as schizophrenia and bipolar disorder, convincing evidence of incremental validity is often lacking. As Exner's co-author Weiner (1999, pp. 336–37) stated, "The Rorschach Inkblot Method is not a diagnostic test, it was not designed as a diagnostic test, it is not intended to be a diagnostic test, and it does not in fact work very well as a diagnostic test, especially if what is meant by diagnosis is a DSM category."

Third, critics and proponents agree that CS scores are generally unrelated to self-report measures that were once thought to measure the same or similar constructs (e.g., Archer & Krishnamurthy 1993a,b; Greenwald 1990, 1991, 1999; Meyer 1992, 1993, 1996; Nezworski & Wood 1995). CS proponents have recently argued that the negligible relationship between self-report measures and Rorschach scores implies that projective techniques assess different aspects of personality (i.e., implicit characteristics) than do self-report techniques (i.e., explicit characteristics) (Bornstein 2001). If such an argument were correct, the Rorschach should provide substantial incremental validity beyond self-report measures for psychologically relevant external criteria. However, evidence of the Rorschach's incremental validity is limited to a few variables, and the gain in predictive power is often small (Lilienfeld et al. 2000).

Fourth, despite claims that "every variable in the Comprehensive System has demonstrated substantial interrater reliability" (Ritzler 1995, p. 230), there is now considerable evidence that the scoring reliability of many variables is mediocre or poor. Nunnally (1978, pp. 245–46) recommended that test scores used in clinical assessments should have a minimum reliability of 0.90. In contrast, a recent study by Acklin et al. (2000) found that approximately 50% of CS variables had interrater reliabilities below 0.85 and some had reliabilities below 0.30 (see also Gronnerod 1999, Nakata 1999, Shaffer et al. 1999).

Fifth, both critics and proponents agree that at least a few Rorschach scores are valid for certain purposes. Several meta-analyses of published Rorschach studies (e.g., Garb et al. 1998, Hiller et al. 1999, Parker et al. 1988) have yielded a mean weighted validity coefficient of 0.30 ± 0.05. Although this figure may be inflated owing to publication bias and methodological flaws, the findings indicate that "some Rorschach indexes can possess moderate validity" (Hunsley & Bailey 1999, p. 269). Despite the findings of such global meta-analyses, only a handful

of individual Rorschach scores possess well-demonstrated and adequate validity (Lilienfeld et al. 2000).

In summary, widely held assumptions regarding the superior psychometric properties of the CS have been abandoned or drastically modified in recent years. Archer (1999, p. 309) concluded that "the assumption that the Rorschach Comprehensive System rests solidly and uniformly on an empirical foundation has been forced to undergo a significant re-examination." Ironically, the Board of Professional Affairs (1998, p. 392) of the American Psychological Association recently commended the CS as "perhaps the most powerful psychometric instrument ever envisioned." In response, Wood & Lilienfeld (1999, p. 348) suggested that the Board's commendation is at least as much an overstatement as the old claim that the Rorschach is an X-ray of the mind (Klopfer 1940). It is unclear what will become of practicing psychologists' long, bittersweet romance with the Rorschach. Perhaps the Rorschach will be reconceptualized as an aid to self-exploration in psychotherapy rather than as an assessment device (Aronow 2001, Widiger 2001).

Self-Report Tests

Important developments concerning self-report tests (e.g., their use in multicultural assessment and detection of malingering) were described above. This section focuses on three issues: (*a*) the impact of response options on self-reports, (*b*) the advantages of the MMPI-2 over the Rorschach in clinical practice, and (*c*) the treatment utility of self-report and projective techniques. Each of these topics points to both the strength and continued viability of self-report tests, as well as to potential challenges confronting users of these tests in clinical practice.

IMPACT OF RESPONSE OPTIONS ON SELF-REPORT Until recently, most personality assessment researchers paid relatively little attention to the selection of response options (e.g., labeling of anchor points on questionnaire items) in the design of self-report measures (see Clark & Watson 1995 for a useful discussion of the pros and cons of differing questionnaire response options). Schwarz and his colleagues have questioned this indifference, arguing persuasively that questionnaire items can be viewed as implicit forms of communication between test developers and test takers (Schwarz 1999). Specifically, they maintain that when participants respond to self-report items, they attempt to discern the pragmatic meaning or intent of the item in addition to its literal meaning. As a consequence, seemingly trivial differences in response options across self-report items can sometimes lead to substantial differences in participants' responses.

For example, in one study (see Schwarz 1999), 39% of patients with psychosomatic disorders reported physical symptom frequencies of more than twice a month when the item anchors ranged from "never" to "more than twice a month." Yet 62% of patients with psychosomatic disorders reported symptom frequencies of more than twice a month when the item anchors ranged from "twice a month or less" to "several times a day." The impact of these different response options

was greatest for ambiguous symptoms (e.g., "responsiveness to changes in the weather"). Schwarz and his collaborators contend that these different response options led the two sets of participants to ascribe different interpretations to the same questions. Specifically, participants who were provided with a response scale that implied a higher frequency of symptoms (i.e., "twice a month or less" to "several times a day") interpreted the questions as inquiring about less severe symptoms than did the other participants.

Still other research by Schwarz and his colleagues demonstrates that minor differences in questionnaire format, including the order of items, can sometimes influence substantially not only the mean levels of item endorsement but also the intercorrelations among items (e.g., Schwarz et al. 1991). Although the findings of Schwarz and his collaborators warrant replication by independent investigators, these findings suggest that researchers may no longer be able to treat differing questionnaire response scales as essentially interchangeable. Instead, researchers should remain cognizant of the differing meanings that participants may impute to self-report items and the potential impact of response options on these interpretations.

ADVANTAGES OF THE MMPI-2 OVER THE RORSCHACH At a time when the scientific status of projective techniques is being vigorously challenged (Lilienfeld et al. 2000), it is worth asking whether self-report instruments could withstand the same level of scrutiny that projective techniques have been exposed to (Archer 1999, Widiger 2001). In this section we compare the MMPI-2 (clinical psychology's most widely used self-report test) and the CS for the Rorschach (the most popular projective technique) and identify five advantages of the MMPI-2 over the CS in clinical practice and managed care settings.

The first advantage of the MMPI-2 is obvious: It can be scored easily and with nearly perfect reliability by a computer. In contrast, scoring of the CS is laborious and complicated, so that Rorschach proponents consider a kappa of 0.61 for interscorer reliability to be "substantial and acceptable" (Acklin et al. 2000, p. 34).

Second, the MMPI-2 is considerably less expensive than the CS, a consideration that is of considerable importance in the era of managed care. If administered and scored by a clerical worker, the MMPI-2 can be interpreted by a psychologist in approximately 30 minutes (Ball et al. 1994). In comparison, the CS must be administered, scored, and interpreted by a psychologist, taking approximately 2.5 hours. In addition, the training required to score and interpret the CS is extremely time-consuming and expensive (Groth-Marnat 1999).

Third, the norms of the MMPI-2 are on much firmer scientific footing than those for the CS. The MMPI-2 norms are based on a stratified probability sample of 2600 American adults that was collected in the late 1980s. In contrast, the CS norms (Exner 2001b) are based on a nonprobability sample of 600 adults that was collected in the late 1970s and early 1980s (Hunsley & Di Giulio 2001). Furthermore, as already discussed, considerable evidence indicates that the CS norms are unrepresentative of normal American adults (Shaffer et al. 1999, Wood et al. 2001b).

Fourth, validity research is stronger for the MMPI-2 than for the CS. Positive findings have been reported for CS scores, but typically they have not been replicated by independent researchers (Lilienfeld et al. 2000). Furthermore, research indicates that the clinical judgments of psychologists become more accurate when they use the MMPI but not when they use the Rorschach (Garb 1998, also see Whitehead 1985).

Fifth, the MMPI-2 can be used more confidently with minority groups than can the CS. As discussed earlier, the cross-cultural validity of the MMPI-2 has been examined in many groups and with increasingly sophisticated methods. In comparison, research on ethnic and cross-cultural differences for the CS is extremely limited (Velasquez 1995). In fact, Dana (1993, p. 160) concluded, "The Rorschach and the Exner Comprehensive versions are not recommended for routine cross-cultural applications."

TREATMENT UTILITY One critical question that has received surprisingly little research attention concerns the treatment utility of self-report tests. The issue of treatment utility applies with equal force to projective techniques. Despite the widespread use of both classes of techniques in clinical practice (Watkins et al. 1995), there is virtually no evidence that they enhance treatment outcome (Hunsley & Bailey 1999, Lilienfeld et al. 2000). One seeming exception to this absence of evidence is the work of Finn and his colleagues (Finn 1996, Finn & Tonsager 1992), who reported that providing clients (college students awaiting psychotherapy) with feedback based on their MMPI-2 scores decreased their psychological distress.

Although Finn's findings (1996, Finn & Tonsager 1992) are a promising first step toward demonstrating treatment utility, they are open to multiple interpretations. Halperin & Snyder (1979), for example, found that snake-phobic clients who received bogus "Barnum" feedback (i.e., highly vague and generalized interpretations of test scores) after completing two psychological tests exhibited enhanced treatment outcome compared with clients who received no test feedback. Moreover, classic research by Sundberg (1955) demonstrates that individuals typically cannot discriminate genuine feedback based on their MMPIs from bogus MMPI feedback at greater than chance levels. Consequently, the work of Finn and his colleagues demonstrates only that assessment feedback to clients can be therapeutic, although it demonstrates neither the treatment utility of the MMPI-2 per se nor even the necessity for accurate MMPI-2 feedback.

A more informative test of treatment utility would involve the use of *manipulated assessment* designs (Hayes et al. 1987). In such designs therapists (who in essence serve as participants) are randomly assigned to receive either information from a given assessment device (e.g., an MMPI-2) or no such assessment information. The extent to which the provision of this information contributes to improved treatment outcome constitutes a direct test of the assessment device's treatment utility (see Harkness & Lilienfeld 1997). To our knowledge, manipulated assessment designs have yet to find their way into the personality assessment literature. Nevertheless, such designs should become a priority among researchers in

this area, as the pressures of managed care force practitioners to demonstrate that the psychological tests they administer are therapeutically useful.

CLINICAL JUDGMENT AND DECISION-MAKING

Literature reviews on clinical assessment typically focus on the development and validation of new tests. However, in this section we focus on an equally important component of the assessment process: clinical judgment and decision-making. Steady progress has been made in this area in recent years. Many of the results point out ways that clinical judgments can become more reliable and valid.

Diagnosis

Important insights can be drawn from research on the diagnostic process. First, results from recent studies indicate that agreement between diagnoses made by mental health professionals in the course of their clinical work and research diagnoses based on structured interviews ranges from poor [kappa = 0.24 (Shear et al. 2000), kappa = 0.25 (Strakowski et al. 1997)] to fair [(kappa = 0.45, 0.51, 0.52 (Basco et al. 2000)]. Similar results have been reported in earlier studies (see Garb 1998, pp. 53–54).

When clinicians' diagnoses are compared with diagnoses based on structured interviews, it becomes clear that clinicians underdiagnose a range of mental disorders. This is true for the diagnosis of mental disorders in the mentally retarded, major depressive disorder in terminally ill patients, personality disorders in clients receiving mental health treatment, substance abuse disorders in psychiatric patients, and obsessive-compulsive disorder (and perhaps other anxiety disorders) in substance abuse patients (Garb 1998, pp. 74–77; Hansen et al. 2000; Zimmerman & Mattia 1999a). Underdiagnosis also seems to occur for posttraumatic stress disorder in routine clinical settings (Zimmerman & Mattia 1999b).

Second, and somewhat paradoxically, it also appears that under some circumstances clinicians tend to "overpathologize" patients, perceiving them as more psychopathological than they really are (e.g., Kullgren et al. 1996). In some cases, the tendency to overpathologize clients has been due to the inadequacies of popular psychological tests. For example, as already noted, research indicates that the norms of the CS of the Rorschach are flawed, so that normal individuals tend to appear more psychopathological than they really are (Shaffer et al. 1999, Wood et al. 2001b). Some evidence suggests that CS scores erroneously indicate the presence of depression or a personality disorder in about 75% of normal individuals (see discussions by Exner 1991, pp. 432–33; Wood et al. 2001c).

Third, important results have been obtained regarding the effect of bias on diagnoses. Agreement between clinical diagnoses and research diagnoses has not been significantly different for males and females (Basco et al. 2000, Shear et al. 2000). However, it has sometimes differed for ethnic groups: Agreement between clinical and research diagnoses has been better for whites than non-whites (54%

vs. 35%) (Strakowski et al. 1997), worse for whites than for minority patients (kappa = 0.47 and 0.49 for whites, 0.57 and 0.59 for minority patients) (Basco et al. 2000), and nonsignificantly different for both groups (Shear et al. 2000). It is not clear why results on race bias varied across studies.

Structured interviews really do make a difference, as revealed by the disparity between clinical and structured interview diagnoses. When structured interviews are used, it is more likely that clinicians will adhere to diagnostic criteria and interrater reliability will be at least fair. Also, construct validity is at least fair to good for many structured interviews: For example, structured interviews have routinely been used in studies on psychopathology that have obtained important results (e.g., Keller et al. 2000, McCullough et al. 2000). For these reasons, psychologists should probably increase their use of structured interviews in making diagnoses. Furthermore, clinical graduate programs should place greater emphasis on training students to use such interviews (for further training recommendations, see Grove 2000).

Case Formulation

Perhaps the most difficult judgment task facing mental health professionals involves case formulation. Research reveals that it is surprisingly difficult for clinicians to explain why a client behaves a particular way (Garb 1998, pp. 85–101). Given this body of research, one would hope that psychologists would be cautious when making causal judgments, but this is not always the case. For example, discussing the theory of one psychoanalyst, apparently with approval, Brown et al. (1998) paraphrased her by stressing that "memory for infant trauma is encoded accurately and indelibly" (p. 205) and that implicit memory even for "birth trauma" can have "a profound influence on later development, even when no narrative memory" is available (p. 206). These assertions are not based on convincing empirical evidence, and psychologists who rely on them may be misled into forming false causal conclusions (McNally 1999). A similar situation can be observed for the controversial diagnosis of dissociative identity disorder (formerly known as multiple personality disorder). Here, too, some mental health professionals make questionable causal judgments, tending to attribute their clients' problems to severe childhood trauma and dissociative identity disorder (Spanos 1996), even though " . . . a large proportion—perhaps a majority—of Dissociative Identity Disorder patients . . . exhibit few or no unambiguous signs of this condition prior to therapy" (Lilienfeld et al. 1999, p. 511).

Some things can be done to improve case formulation. Widiger & Clark (2000) recommended that "a means of characterizing a developmental, life span history of a patient's symptomatology should perhaps be provided in DSM-V by recording, for example, age of onset, lifetime history of disorders, and their longitudinal course" (p. 956). Though more descriptive than explanatory in nature, this procedure could help clinicians make more valid causal judgments. Also, collecting this information could lead to a transformation in how psychopathology is viewed. Widiger & Clark gave the following example: "If one comes to understand how an

anxiety disorder develops into a depressive disorder with which it shares a common genetic vulnerability, it could be impossible to persist with the notion that they are separate and distinct disorders" (p. 956).

Another promising approach involves using functional analytical clinical case models (Haynes et al. 1997). This approach calls on the clinician to make "low-level" causal inferences (e.g., the clinician may conclude that marital stress led to a client's presleep worry, and that this worry in combination with pregnancy led to a sleep disturbance). An attractive feature of this approach is that causal relations are described pictorially—one can gain an understanding of a client quickly by looking at the "vector-graphic representation of variables and functional relationships" (Haynes et al. 1997, p. 334). In addition, because making causal judgments is in many ways the most difficult task facing mental health professionals (Garb 1998), functional analytical clinical case modeling is promising because it requires judgments that are tied relatively closely to events and observed behavior. One would expect that the interrater agreement among different clinicians would be good given the low level of causal inference typically required, but this needs to be investigated empirically.

Treatment Decisions

Many psychologists, and certainly many mental health professionals, are unfamiliar with the scientific literature on therapeutic interventions. This is one reason why treatment decisions are sometimes inappropriate. The problem of inappropriate interventions is in part an assessment issue, not simply a treatment issue. In a sense, the problem is that some clinicians make inappropriate decisions when assessing clients and formulating treatment plans.

Interestingly, the American Psychological Association (APA) may encourage poor decisions by offering continuing education credits for a range of treatment techniques that have not been empirically supported. APA continuing education credits used to be offered for workshops on thought field therapy, and are still offered for workshops on calligraphy therapy, Jungian sandplay therapy, and neurotherapy (a form of electroencephalographic biofeedback that has sometimes been advertised as a treatment for depression, learning disabilities, attention-deficit hyperactivity disorder, epilepsy, and coma) (Kline et al. 2001, Lilienfeld 1999). APA continuing education credits are even offered for techniques that appear to be harmful. For example, credits are approved for training in crisis debriefing for victims of traumatic events, even though several studies have found this intervention to have negative effects (e.g., Gist & Lubin 1999, Mayou et al. 2000).

One of the most important recent results on decision-making concerns race bias. Race bias has been observed for the prescription of antipsychotic medicine (Segal et al. 1996; also see Garb 1998, pp. 126–29). This finding was replicated in a study on adherence to treatment recommendations conducted by the Schizophrenia Patient Outcome Research Team (Lehman et al. 1998). In this study, 27.4% of minority patients and 15.9% of white patients were placed on excessive dosages

of antipsychotic medicine. This finding may be related to data indicating that the risk of violence is overestimated for black psychiatric inpatients and black prison inmates (Garb 1998, pp. 113–14; Hoptman et al. 1999). In a study conducted in Israel (J Rabinowitz, T Shlezinger, M Davidson, manuscript submitted for publication), dosage of psychotropic medicine was found to be related to "the extent to which the patient is believed to constitute a threat to the physician" (p. 2). Thus, minority patients may be more likely than white patients to be perceived as being dangerous, and for this reason a substantial number are put on excessive doses of medicine. Other research has also found a relationship between perception of dangerousness and dosage of medicine (Baldessarini et al. 1995).

Prediction of Violence and Detection of Deception

Researchers have reported encouraging findings about clinicians' ability to predict violence. At one time, some psychologists believed that mental health professionals could not predict violence. For example, in an article in *Science*, Faust & Ziskin (1988, p. 32) concluded that "studies on the prediction of violence are consistent: clinicians are wrong at least twice as often as they are correct." We now know that mental health professionals can make valid short-term and long-term predictions of violence (Garb 1998, pp. 107–9; Mossman 1994). For example, in one study (Hoptman et al. 1999), psychiatrists at a forensic psychiatric hospital were asked to predict assaultive behavior. They made predictions during a 3-month period for a sample of 183 recently admitted male patients. Sixty of the patients became assaultive. The clinicians' overall hit rate was 71%. Fifty-four percent of the predictions of "assaultive behavior" and 79% of the predictions of "no assaultive behavior" were correct.

Positive results were also obtained in a study of deception detection (Ekman et al. 1999). Clinical psychologists and other participants watched silent videotapes of people who were lying or telling the truth about their opinions. Subjects who lied and those who told the truth exhibited differences in facial movements. Clinical psychologists with a special interest in detecting deception were more accurate than other clinical psychologists. This finding is particularly interesting because judgment research indicates that it can be surprisingly difficult to draw inferences from nonverbal behavior (Garb 1998, p. 18). The results of this study, along with those of a prior study in which Secret Service agents performed better than chance in detecting deception (Ekman & O'Sullivan 1991), suggest that experience, training, and/or social intelligence can improve performance in this area. Future studies may clarify the importance of individual differences among judges and examine whether the laboratory findings can be duplicated in more realistic situations.

Clinical Judgment, Computers, and Mechanical Prediction

Clinicians can use algorithms programmed into computers to interpret test results and make judgments and decisions (e.g., diagnoses, descriptions of traits and symptoms, behavioral predictions, and treatment decisions). A recent meta-analysis by

Grove and his colleagues (2000) provides the most thorough and sophisticated review of the research on such "mechanical" algorithms. Grove et al. included studies from both psychology and medicine in which mechanical algorithms and human judges were used to "predict human behavior, make psychological or medical diagnoses or prognoses, or assess states and traits (including abnormal behavior and normal personality)" (p. 20). The analysis included 136 studies, making it the largest review ever conducted on this topic.

The results supported the use of mechanical algorithms to make judgments. This held true across categories: "It holds in general medicine, in mental health, in personality, and in education and training settings" (Grove et al. 2000, p. 25). Mechanical algorithms substantially outperformed predictions made by human judges in 33–47% of the studies. In contrast, judges substantially outperformed algorithms in 6–16% of the studies. In the remaining studies, clinicians and algorithms were about equally accurate. On average, the mechanical algorithms were about 10% more accurate than clinicians. The algorithms were usually superior, regardless of whether clinicians were "inexperienced or seasoned judges" (p. 25).

The Grove et al. (2000) meta-analysis is a landmark study. Its findings strongly suggest that computers can supplement and in some cases improve upon the decisions that clinicians make in their work. However, two important limitations of computerized decision-making should be noted. First, well-validated mechanical decision rules are currently unavailable for most clinical tasks. For example, substantial progress has been made in developing computerized algorithms to predict violence, child abuse and neglect, and recidivism among juvenile offenders (see Forensic Assessment above). However, there are still no well-validated algorithms for making diagnoses, predicting behavior, describing personality traits and psychopathology, or making treatment decisions (Garb 2000). In the future, research is needed to develop and validate computer programs for such tasks.

Second, despite the positive findings of Grove et al. (2000), the highly popular computer programs that clinical psychologists currently use to interpret test results (e.g., the MMPI-2 and the Rorschach) have not generally performed well in validity studies, perhaps because these programs are generally based on the "canned" interpretations of experts, rather than on empirically developed actuarial decision rules. When Butcher et al. (2000) reviewed the literature, they were able to find only four validity studies during the 1990s that had examined these programs. Two of the studies reported negative findings, a third reported mixed results, and the fourth reported only mildly positive results. As Snyder (2000, p. 55) concluded, "studies regarding CBTIs' [computer-based test interpretations'] validity are scarce; for most CBTI systems, they are lacking entirely."

FUTURE DIRECTIONS

As this review has shown, the field of clinical assessment is in the midst of significant change. New pressures are forcing psychologists to demonstrate that their assessment techniques are cost-effective, scientifically sound, and culturally fair.

In concluding, we offer four predictions.

1. The economic constraints on health and psychological services that characterize the era of managed care will not disappear any time soon. Psychologists in the future will be held accountable to show that their assessment techniques yield tangible benefits, are cost-effective, and provide incrementally valid information beyond what can be obtained from less expensive sources. Some traditional assessment techniques will have to be abandoned and new ones developed.
2. As the US population becomes more diverse, valid multicultural assessment will become increasingly important. For many years multicultural issues have been regarded as peripheral to research and practice in clinical assessment. In the future they will become a major focus of attention.
3. Owing to changes in the law, the assessment techniques used by forensic psychologists will face increasingly stringent scrutiny. Current debates concerning the legal admissibility of the MCMI-III and the Rorschach CS are only the first signs of a trend toward greater accountability. Prudent forensic psychologists will pay attention to such debates and base their opinions as much as possible on well-validated assessment approaches. Those who fail to do so will run an ever-higher risk of being embarrassed in court or having their recommendations discounted.
4. Developments in science are likely to introduce radically new and unexpected elements into the assessment process, beyond those we have identified here. For example, Plomin & Crabbe (2000) predicted that advances in behavioral genetics will change clinical psychology: "Here is what the future might look like for clinical psychologists. DNA will be routinely collected. The most powerful potential for DNA is to predict risk so that genes can be used to aid in diagnosis and plan treatment programs" (p. 823). Similarly, neuroimaging techniques for visualizing brain structure and function may become increasingly sophisticated and affordable. It is interesting to imagine a time when psychologists may replace their Rorschach cards with a DNA kit and a pocket scanner.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

Acklin MW. 1999. Behavioral science foundations of the Rorschach test: research and clinical applications. *Assessment* 6:319–26

Acklin MW, McDowell CJ, Verschell MS, Chan D. 2000. Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *J. Pers. Assess.* 74: 15–47

Allen J, Walsh JA. 2000. A construct-based approach to equivalence: methodologies for cross-cultural/multicultural personality assessment research. See Dana 2000, pp. 63–85

Arbisi PA, Ben-Porath YS. 1995. An MMPI-2 infrequent response scale for use with psychopathological populations: the Infrequency-Psychopathology scale, *F*(*p*). *Psychol. Assess.* 7:424–31

Arbisi PA, Ben-Porath YS. 1998. The ability of Minnesota Multiphasic Personality Inventory–2 validity scales to detect fake-bad responses in psychiatric inpatients. *Psychol. Assess.* 10:221–28

Arbisi PA, Ben-Porath YS, McNulty JL. 1998. *The impact of ethnicity on the MMPI-2 in inpatient psychiatric settings*. Presented at Annu. Meet. Am. Psychol. Assoc., San Francisco

Archer RP. 1999. Introduction to a special section: perspectives on the Rorschach. *Assessment* 6:307–11

Archer RP, Krishnamurthy R. 1993a. Combining the Rorschach and the MMPI in the assessment of adolescents. *J. Pers. Assess.* 60:132–40

Archer RP, Krishnamurthy R. 1993b. A review of MMPI and Rorschach interrelationships in adult samples. *J. Pers. Assess.* 61:277–93

Archer RP, Newsom CR. 2000. Psychological test usage with adolescent clients: survey update. *Assessment* 7:227–35

Ardila A, Rosselli M, Puente AE. 1994. *Neuropsychological Evaluation of the Spanish Speaker*. New York: Plenum

Arnold BR, Matus YE. 2000. Test translation and cultural equivalence methodologies for use with diverse populations. See Cuellar & Panigua 2000, pp. 121–36

Aronow E. 2001. CS norms, psychometrics, and possibilities for the Rorschach technique. *Clin. Psychol. Sci. Pract.* 8:383–85

Bagby RM, Nicholson RA, Buis T, Bacchiochi JR. 2000. Can the MMPI-2 validity scales detect depression feigned by experts? *Assessment* 7:55–62

Baird C, Wagner D. 2000. The relative validity of actuarial- and consensus-based risk assessment systems. *Child. Youth Serv. Rev.* 22:839–71

Baldessarini RJ, Kando JC, Centorrino F. 1995. Hospital use of antipsychotic agents in 1989 and 1993: stable dosing with decreased length of stay. *Am. J. Psychiatry* 152:1038–44

Ball JD, Archer RP, Imhof EA. 1994. Time requirements of psychological testing: a survey of practitioners. *J. Pers. Assess.* 63:239–49

Basco MR, Bostic JQ, Davies D, Rush AJ, Witte B, et al. 2000. Methods to improve diagnostic accuracy in a community mental health setting. *Am. J. Psychiatry* 157:1599–605

Board Professional Affairs. 1998. Awards for distinguished professional contributions: John Exner. *Am. Psychol.* 53:391–92

Bornstein RF. 2001. The clinical utility of the Rorschach Inkblot Method: reframing the debate. *J. Pers. Assess.* 77:39–47

Boscan DC. 1999. *The Rorschach test: a Mexican sample using the Comprehensive System.* PhD thesis. Fielding Inst., Santa Barbara, CA

Brown D, Scheflin AW, Hammond DC. 1998. *Memory, Trauma Treatment, and the Law*. New York: Norton

Butcher JN, Graham JR, Ben-Porath YS. 1995. Methodological problems and issues in MMPI, MMPI-2, and MMPI-A research. *Psychol. Assess.* 7:320–29

Butcher JN, Lim J, Nezami E. 1998. Objective study of abnormal personality in cross-cultural settings: the Minnesota Multiphasic Personality Inventory (MMPI–2). *J. Cross-Cult. Psychol.* 29:189–211

Butcher JN, Perry JN, Atlis MM. 2000. Validity and utility of computer-based test interpretation. *Psychol. Assess.* 12:6–18

Butcher JN, Rouse SV. 1996. Personality: individual differences and clinical assessment. *Annu. Rev. Psychol.* 47:87–111

Childs RA, Dahlstrom WG, Kemp SM, Panter AT. 2000. Item response theory in personality assessment: a demonstration using the MMPI-2 Depression scale. *Assessment* 7:37–54

Clark LA, Watson D. 1995. Constructing validity: basic issues in objective scale development. *Psychol. Assess.* 7:309–19

Cleary TA, Humphreys LG, Kendrick SA,

Wesman AG. 1975. Educational uses of tests with disadvantaged students. *Am. Psychol.* 30:15–41

Cuellar I. 2000. Acculturation as a moderator of personality and psychological assessment. See Dana 2000, pp. 113–29

Cuellar I, Arnold B, Maldonado R. 1995. Acculturation rating scale for Mexican-Americans-II: a revision of the original ARSMA scale. *Hisp. J. Behav. Sci.* 17:275–304

Cuellar I, Paniagua FA, eds. 2000. *Handbook of Multicultural Mental Health*: *Assessment and Treatment of Diverse Populations*. San Diego, CA: Academic

Dana RH. 1993. *Multicultural Assessment Perspectives for Professional Psychology*. Boston: Allyn Bacon

Dana RH. 1998. Cultural identity assessment of culturally diverse groups: 1997. *J. Pers. Assess.* 70:1–16

Dana RH, ed. 2000. *Handbook of Cross-Cultural and Multicultural Personality Assessment*. Mahwah, NJ: Erlbaum

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. 509 U.S., 113 S. Ct. 2786

Douglas KS, Cox DN, Webster CD. 1999a. Violence risk assessment: science and practice. *Legal Criminol. Psychol.* 4:149–84

Douglas KS, Ogloff JRP, Nicholls TL, Grant I. 1999b. Assessing risk for violence among psychiatric patients: the HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version. *J. Consult. Clin. Psychol.* 67:917–30

Dyer FJ, McCann JT. 2000. The Millon clinical inventories, research critical of their forensic application, and Daubert criteria. *Law Hum. Behav.* 24:487–97

Edens JF, Poythress NG, Lilienfeld SO. 1999. Identifying inmates at risk for disciplinary infractions: a comparison of two measures of psychopathy. *Behav. Sci. Law* 17:435–43

Ekman P, O'Sullivan M. 1991. Who can catch a liar? *Am. Psychol.* 46:913–20

Ekman P, O'Sullivan M, Frank MG. 1999. A few can catch a liar. *Psychol. Sci.* 10:263–66

Embretson SE, Reise SP. 2000. *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum

Ephraim D. 2000. Culturally relevant research and practice with the Rorschach Comprehensive System. See Dana 2000, pp. 303–27

Exner JE. 1991. *The Rorschach: A Comprehensive System*, Vol. 2. *Interpretation*. New York: Wiley. 2nd ed.

Exner JE. 1993. *The Rorschach: A Comprehensive System*, Vol. 1. *Basic Foundations*. New York: Wiley. 3rd ed.

Exner JE. 1995. *A Rorschach Workbook for the Comprehensive System*. Asheville, NC: Rorschach Workshops. 4th ed.

Exner JE. 2001a. A comment on: The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. *Clin. Psychol. Sci. Pract.* 8: 368–88

Exner JE. 2001b. *A Rorschach Workbook for the Comprehensive System*. Asheville, NC: Rorschach Workshops. 5th ed.

Exner JE, Weiner IB. 1994. *The Rorschach: A Comprehensive System*, Vol. 3. *Assessment of Children and Adolescents*. New York: Wiley. 2nd ed.

Faust D, Ziskin J. 1988. The expert witness in psychology and psychiatry. *Science* 241:31–35

Finn SE. 1996. *Manual for Using the MMPI–2 as a Therapeutic Intervention*. Minneapolis: Univ. Minn. Press

Finn SE, Tonsager ME. 1992. Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychol. Assess.* 4:278–87

Fletcher-Janzen E, Strickland TL, Reynolds CR. 2000. *Handbook of Cross-Cultural Neuropsychology*. New York: Kluwer Acad./Plenum

Forth AE, Hart SD, Hare RD. 1990. Assessment of psychopathy in male young offenders. *Psychol. Assess.* 2:342–44

Gacono CB, Loving JL, Bodholdt RH. 2001. The Rorschach and psychopathy: toward a more accurate understanding of the research findings. *J. Pers. Assess.* 77:16–38

Ganellen RJ. 2001. Weighing evidence

concerning the Rorschach's psychometric properties: a response to Wood et al. 1999. *J. Pers. Assess.* 77:1–15

Garb HN. 1998. *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington, DC: Am. Psychol. Assoc.

Garb HN. 1999. Call for a moratorium on the use of the Rorschach Inkblot in clinical and forensic settings. *Assessment* 6:313–15

Garb HN. 2000. Computers will become increasingly important for psychological assessment: not that there's anything wrong with that! *Psychol. Assess.* 12:31–39

Garb HN, Florio CM, Grove WM. 1998. The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: results from meta-analyses. *Psychol. Sci.* 9:402–4

General Electric Co. v. Joiner. 1997. 118 S. Ct. 512

Gist R, Lubin B. 1999. *Response to Disaster: Psychosocial, Community, and Ecological Approaches*. Philadelphia: Brunner/Mazel

Greene RL. 2000. *The MMPI-2: An Interpretive Manual*. Boston: Allyn Bacon. 2nd ed.

Greenwald DF. 1990. An external construct validity study of Rorschach personality variables. *J. Pers. Assess.* 55:768–80

Greenwald DF. 1991. Personality dimensions reflected by the Rorschach and the 16PF. *J. Clin. Psychol.* 47:708–15

Greenwald DF. 1999. Relationships between the Rorschach and the NEO-Five Factor Inventory. *Psychol. Rep.* 85:519–27

Gronnerod C. 1999. Rorschach interrater agreement estimates: an empirical evaluation. *Scand. J. Psychol.* 40:115–20

Groth-Marnat G. 1999. Financial efficacy of clinical assessment: rational guidelines and issues for future research. *J. Clin. Psychol.* 55:813–24

Grove WM. (Chair). 2000. *APA Division 12* (*Clinical*) *Presidential Task Force "Assessment for the Year 2000." Report of the Task Force*. Washington, DC: Am. Psychol. Assoc., Div. 12 (Clin. Psychol.)

Grove WM, Barden RC. 1999. Protecting the integrity of the legal system: the admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychol. Public Policy Law* 5:224–42

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. 2000. Clinical versus mechanical prediction: a meta-analysis. *Psychol. Assess.* 12:19–30

Hall CCI. 1997. Cultural malpractice: the growing obsolescence of psychology with the changing U.S. population. *Am. Psychol.* 52:642–51

Hall GCN, Bansal A, Lopez IR. 1999. Ethnicity and psychopathology: a meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychol. Assess.* 11:642–51

Halperin K, Snyder CR. 1979. Effect of enhanced psychological test feedback on treatment outcome: therapuetic implications of the Barnum effect. *J. Consult. Clin. Psychol.* 47:140–46

Hamel M, Gallagher S, Soares C. 2001. The Rorschach: Here we go again. *J. Forensic Psychol. Pract.* 1:(3):79–87

Hamel M, Shaffer TW, Erdberg P. 2000. A study of nonpatient preadolescent Rorschach protocols. *J. Pers. Assess.* 75:280–94

Handel RW, Ben-Porath YS. 2000. Multicultural assessment with the MMPI-2: issues for research and practice. See Dana 2000, pp. 229–45

Hansen SS, Munk-Jorgensen P, Guldbaek B, Solgard T, Lauszus KS, et al. 2000. Psychoactive substance use diagnoses among psychiatric in-patients. *Acta Psychiatr. Scand.* 102:432–38

Hanson RK. 1998. What do we know about sex offender risk assessment? *Psychol. Public Policy Law* 4:50–72

Hare R. 1991. *Manual for the Revised Psychopathy Checklist*. Toronto: Multihealth Systems

Harkness AR, Lilienfeld SO. 1997. Individual differences science for treatment planning: personality traits. *Psychol. Assess.* 9:349–60

Hart SD, Hare RD, Forth AE. 1994. Psychopathy as a risk marker for violence: development and validation of a screening version of

the Revised Psychopathy Checklist. In *Violence and Mental Disorder: Developments in Risk Assessment*, ed. J Monahan, HJ Steadman, pp. 81–97. Chicago: Univ. Chicago Press

Hayes SC, Nelson RO, Jarrett RB. 1987. The treatment utility of assessment: a functional approach to evaluating assessment quality. *Am. Psychol.* 42:963–74

Haynes SN, Leisen MB, Blaine DD. 1997. Design of individualized behavioral treatment programs using functional analytic clinical case models. *Psychol. Assess.* 9:334–48

Hemphill JF, Hare RD, Wong S. 1998. Psychopathy and recidivism: a review. *Legal Criminol. Psychol.* 3:139–70

Hiller JB, Rosenthal R, Bornstein RF, Berry DTR, Brunell-Neuleib S. 1999. A comparative meta-analysis of Rorschach and MMPI validity. *Psychol. Assess.* 11:278–96

Hoptman MJ, Yates KF, Patalinjug MB, Wack RC, Convit A. 1999. Clinical prediction of assaultive behavior among male psychiatric patients at a maximum-security forensic facility. *Psychiatr. Serv.* 50:1461–66

Hunsley J, Bailey JM. 1999. The clinical utility of the Rorschach: unfulfilled promises and an uncertain future. *Psychol. Assess.* 11:266–77

Hunsley J, Di Giulio G. 2001. Norms, norming, and clinical assessment. *Clin. Psychol. Sci. Pract.* 8:378–82

Iverson GL, Franzen MD, Hammond JA. 1995. Examination of inmates' ability to malinger on the MMPI-2. *Psychol. Assess.* 7:118–21

Keller J, Nitschke JB, Bhargava T, Deldin PJ, Gergen JA, Miller JA. 2000. Neuropsychological differentiation of depression and anxiety. *J. Abnorm. Psychol.* 109:5–10

Kline JP, Brann CN, Loney BR. 2001. A cacophony in the brainwave: a critical appraisal of neurotherapy for ADHD. *Sci. Rev. Ment. Health Pract.* In press

Klopfer B. 1940. Personality aspects revealed by the Rorschach method. *Rorschach Res. Exch.* 4:26–29

Kullgren G, Jacobsson L, Lynoe N, Kohn R, Levav I. 1996. Practices and attitudes among Swedish psychiatrists regarding the ethics of compulsory treatment. *Acta Psychiatr. Scand.* 93:389–96

Kumho Tire Co., Ltd. v. Carmichael. 1999. 119 S. Ct. 1167

Kuncel NR, Hezlett SA, Ones DS. 2001. A comprehensive meta-analysis of the predictive validity of the Graduate Record Examination: implications for graduate student selection and performance. *Psychol. Bull.* 127:162–81

Lally SJ. 2001. Should human figure drawings be admitted into court? *J. Pers. Assess.* 76:135–49

Lamb DG, Berry DTR, Wetter MW, Baer RA. 1994. Effects of two types of information on malingering of closed head injury on the MMPI-2: an analog investigation. *Psychol. Assess.* 6:8–13

Lambert MJ, Burlingame GM, Umphress V, Hansen NB, Vermeersch DA, et al. 1996. The reliability and validity of the Outcome Questionnaire. *Clin. Psychol. Psychother.* 3:249–58

Lambert MJ, Finch AE. 1999. The Outcome Questionnaire. See Maruish 1999, pp. 831–69

Lambert MJ, Okiishi JC, Finch AE, Johnson LD. 1998. Outcome assessment: from conceptualization to implementation. *Prof. Psychol. Res. Pract.* 29:63–70

Lehman AF, Steinwachs DM, the co-investigators of the PORT project. 1998. Patterns of usual care for schizophrenia: initial results from the Schizophrenia Patient Outcomes Research Team (PORT) client survey. *Schizophr. Bull.* 24:11–20

Levenson MR, Kiehl KA, Fitzpatrick CM. 1995. Assessing psychopathic attributes in a noninstitutionalized population. *J. Pers. Soc. Psychol.* 68:151–58

Lilienfeld SO. 1998. Methodological advances and developments in the assessment of psychopathy. *Behav. Res. Ther.* 36:99–125

Lilienfeld SO. 1999. Pseudoscience in contemporary clinical psychology: what it is and what we can do about it. *Clin. Psychol.* 51:3–9

Lilienfeld SO, Andrews BP. 1996. Development and preliminary validation of a self-report measure of psychopathic personality in noncriminal populations. *J. Pers. Assess.* 66:488–524

Lilienfeld SO, Lynn SJ, Kirsch I, Chaves JF, Sarbin TR, et al. 1999. Dissociative identity disorder and the sociocognitive model: recalling the lessons of the past. *Psychol. Bull.* 125:507–23

Lilienfeld SO, Wood JM, Garb HN. 2000. The scientific status of projective techniques. *Psychol. Sci. Public Interest* 1:27–66

Marin G, Gamba RJ. 1996. A new measurement of acculturation for Hispanics: the bidimensional acculturation scale for Hispanics (BAS). *Hisp. J. Behav. Sci.* 18:297–316

Maruish ME, ed. 1999. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*. Mahwah, NJ: Erlbaum. 2nd ed.

Mayou RA, Ehlers A, Hobbs M. 2000. Psychological debriefing for road traffic accident victims: three-year follow-up of a randomised controlled trial. *Br. J. Psychiatry* 176:589–93

McCann JT. 1998a. Defending the Rorschach in court: an analysis of admissibility using legal and professional standards. *J. Pers. Assess.* 70:125–44

McCann JT. 1998b. *Malingering and Deception in Adolescents: Assessing Credibility in Clinical and Forensic Settings*. Washington, DC: Am. Psychol. Assoc.

McCann JT, Dyer FJ. 1996. *Forensic Assessment with the Millon Inventories*. New York: Guilford

McCullough JP, Klein DN, Keller MB, Holzer CE, Davis SM, et al. 2000. Comparison of DSM-III-R chronic major depression and major depression superimposed on dysthymia (double depression): validity of the distinction. *J. Abnorm. Psychol.* 109:419–23

McKinzey RK, Podd MH, Krehbiel MA, Mensch AJ, Trombka CC. 1997. Detection of malingering on the Luria-Nebraska Neuropsychological Battery: an initial and cross-validation. *Arch. Clin. Neuropsychol.* 12:505–12

McKinzey RK, Podd MH, Krehbiel MA, Raven J. 1999. Detection of malingering on Raven's Standard Progressive Matrices: a cross-validation. *Br. J. Clin. Psychol.* 38:435–39

McNally RJ. 1999. Review of the book *Memory, Trauma Treatment, and the Law*. *Int. J. Clin. Exp. Hypnosis* 47:374–82

McNulty JL, Graham JR, Ben-Porath YS, Stein LAR. 1997. Comparative validity of MMPI-2 scores of African-American and Caucasian mental health center clients. *Psychol. Assess.* 9:464–70

Meehl PE. 1954. *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: Univ. Minn. Press

Meyer GJ. 1992. The Rorschach's factor structure: a contemporary investigation and historical review. *J. Pers. Assess.* 59:117–36

Meyer GJ. 1993. The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *J. Pers. Assess.* 60:153–80

Meyer GJ. 1996. The Rorschach and MMPI: toward a more scientifically differentiated understanding of cross-method assessment. *J. Pers. Assess.* 67:558–78

Meyer GJ. 2001. Evidence to correct misperceptions about Rorschach norms. *Clin. Psychol. Sci. Pract.* 8:389–96

Meyer GJ, Richardson C. 2001. *An examination of changes in Form Quality codes in the Rorschach Comprehensive System from* 1974 to 1995. Presented at Midwinter Meet. Soc. Pers. Assess., Philadelphia

Monahan J, Steadman HJ, Silver E, Appelbaum P, Robbins PC, et al. eds. 2001. *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*. New York: Oxford Univ. Press

Mossman D. 1994. Assessing predictions of violence: being accurate about accuracy. *J. Consult. Clin. Psychol.* 62:783–92

Mueller RM, Lambert MJ, Burlingame GM.

1998. Construct validty of the Outcome Questionnaire: a confirmatory factor analysis. *J. Pers. Assess.* 70:248–62

Nakata LM. 1999. *Interrater reliability and the Comprehensive System for the Rorschach: clinical and non-clinical protocols*. PhD thesis. Pacif. Grad. Sch. Psychol., Palo Alto, CA

Nezworski MT, Wood JM. 1995. Narcissism in the Comprehensive System for the Rorschach. *Clin. Psychol. Sci. Pract.* 2:179–99

Nicholson RA, Mouton GJ, Bagby RM, Buis T, Peterson SA, Buigas RA. 1997. Utility of MMPI-2 indicators of response distortion: receiver operating characteristic analysis. *Psychol. Assess.* 9:471–79

Nunnally JC. 1978. *Psychometric Theory*. New York: McGraw-Hill. 2nd ed.

O'Donohue W, Bradley AR. 1999. Conceptual and empirical issues in child custody evaluations. *Clin. Psychol. Sci. Pract.* 6:310–22

Okazaki S, Sue S. 1995. Methodological issues in assessment research with ethnic minorities. *Psychol. Assess.* 7:367–75

Ostrosky-Solis F, Ardila A, Rosselli M. 1999. NEUROPSI: a brief neuropsychological test battery in Spanish with norms by age and educational level. *J. Int. Neuropsychol. Soc.* 5:413–33

Parker KCH, Hanson RK, Hunsley J. 1988. MMPI, Rorschach, and WAIS: a meta-analytic comparison of reliability, stability, and validity. *Psychol. Bull.* 103:367–73

Phelps R, Eisman EJ, Kohout J. 1998. Psychological practice and managed care: results of the CAPP Practitioner Survey. *Prof. Psychol. Res. Pract.* 29:31–36

Piotrowski C. 1999. Assessment practices in the era of managed care: current status and future directions. *J. Clin. Psychol.* 55:787–98

Piotrowski C, Belter RW, Keller JW. 1998. The impact of "managed care" on the practice of psychological testing: preliminary findings. *J. Pers. Assess.* 70:441–47

Plomin R, Crabbe J. 2000. DNA. *Psychol. Bull.* 126:806–28

Pollock PH. 1996. A cautionary note on the determination of malingering in offenders. *Psychol. Crime Law.* 3:97–110

Ponton MO, Ardila A. 1999. The future of neuropsychology with Hispanic populations in the United States. *Arch. Clin. Neuropsychol.* 14:565–80

Puente AE, Ardila A. 2000. Neuropsychological assessment of Hispanics. See Fletcher-Janzen et al. 2000, pp. 87–104

Quinsey VL, Harris GT, Rice ME, Cormier CA. 1998. *Violent Offenders: Appraising and Managing Risk.* Washington, DC: Am. Psychol. Assoc.

Rees LM, Tombaugh TN, Gansler DA, Maczynski NP. 1998. Five validation experiments of the Test of Memory Malingering (TOMM). *Psychol. Assess.* 10:10–20

Reynolds CR, ed. 1998. *Detection of Malingering During Head Injury Litigation. Critical Issues in Neuropsychology*. New York: Plenum

Ritzler B. 1995. Putting your eggs in the content analysis basket: a response to Aronow, Reznikov, and Moreland. *J. Pers. Assess.* 64:229–34

Rogers R. 1995. *Diagnostic and Structured Interviewing: A Handbook for Psychologists*. Odessa, FL: Psychol. Assess. Res.

Rogers R, ed. 1997. *Clinical Assessment of Malingering and Deception*. New York: Guilford. 2nd ed.

Rogers R, Salekin RT, Hill C, Sewell KW, Murdock ME, Neumann CS. 2000a. The Psychopathy Checklist—Screening Version: an examination of criteria and subcriteria in three forensic samples. *Assessment* 7:1–15

Rogers R, Salekin RT, Sewell KW. 1999. Validation of the Millon Clinical Mutiaxial Inventory for Axis II disorders: Does it meet the Daubert standard? *Law Hum. Behav.* 23:425–43

Rogers R, Salekin RT, Sewell KW. 2000b. The MCMI-III and the Daubert standard: separating rhetoric from reality. *Law Hum. Behav.* 24:501–6

Rogers R, Sewell KW, Salekin RT. 1994. A meta-analysis of malingering on the MMPI-2. *Assessment* 1:227–37

Salekin RT, Rogers R, Sewell KW. 1996. A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: predictive validity of dangerousness. *Clin. Psychol. Sci. Pract.* 3:203–15

Santor DA, Ramsay JO. 1998. Progress in the technology of measurement: applications of item response models. *Psychol. Assess.* 10:345–59

Schretlen DJ. 1997. Dissimulation on the Rorschach and other projective measures. See Rogers 1997, pp. 208–22

Schwarz N. 1999. Self-reports: how the questions shape the answers. *Am. Psychol.* 54:93–105

Schwarz N, Strack F, Mai HP. 1991. Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. *Public Opin. Q.* 55:3–23

Segal SP, Bola JR, Watson MA. 1996. Race, quality of care, and antipsychotic prescribing practices in psychiatric emergency services. *Psychiatr. Serv.* 47:282–86

Shaffer TW, Erdberg P, Haroian J. 1999. Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *J. Pers. Assess.* 73: 305–16

Shear MK, Greeno C, Kang J, Ludewig D, Frank E, et al. 2000. Diagnosis of nonpsychotic patients in community clinics. *Am. J. Psychiatry* 157:581–87

Snyder DK. 2000. Computer-assisted judgment: defining strengths and liabilities. *Psychol. Assess.* 12:52–60

Spanos NP. 1996. *Multiple Identities and False Memories*. Washington, DC: Am. Psychol. Assoc.

Stephenson M. 2000. Development and validation of the Stephenson Multigroup Acculturation Scale (SMAS). *Psychol. Assess.* 12:77–88

Storm J, Graham JR. 2000. Detection of coached general malingering on the MMPI-2. *Psychol. Assess.* 12:158–65

Strakowski SM, Hawkins JM, Keck PE, McElroy SL, West SA, et al. 1997. The effects of race and information variance on disagreement between psychiatric emergency service and research diagnoses in first-episode psychosis. *J. Clin. Psychiatry* 58:457– 63

Sundberg ND. 1955. The acceptability of "fake" versus "bona fide" personality test interpretations. *J. Abnorm. Soc. Psychol.* 50: 145–47

Suzuki LA, Ponterotto JG, Meller PJ, eds. 2000. *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications*. San Francisco: Jossey-Bass. 2nd ed.

Timbrook RE, Graham JR. 1994. Ethnic differences on the MMPI-2? *Psychol. Assess.* 6:212–17

Tombaugh TN. 1997. The Test of Memory Malingering (TOMM): normative data from cognitively intact and cognitively impaired individuals. *Psychol. Assess.* 9:260–68

US Bureau Census 2000. *Statistical Abstracts of the U.S.* Washington, DC: US Bur. Census

Velasquez RJ. 1995. Personality assessment of Hispanic clients. In *Clinical Personality Assessment: Practical Approaches*, ed. JN Butcher, pp. 120–39. New York: Oxford Univ. Press

Velasquez RJ, Ayala GX, Pace T, Mendoza S, Choney SK, et al. 2000. Culturally competent use of the Minnesota Multiphasic Personality Inventory-2. See Cuellar & Paniagua 2000, pp. 389–417

Vermeersch DA, Lambert MJ, Burlingame GM. 2000. Outcome questionnaire: item sensitivity to change. *J. Pers. Assess.* 74:242–61

Viglione DJ. 1999. A review of recent research addressing the utility of the Rorschach. *Psychol. Assess.* 11:251–65

Vinet EV. 2000. The Rorschach Comprehensive System in Iberoamerica. See Dana 2000, pp. 345–65

Waller NG. 1998. Searching for structure in the MMPI. In *The New Rules of Measurement*, ed. SE Embretson, SL Hershberger, pp. 185–217. Mahwah, NJ: Erlbaum

Waller NG, Thompson JS, Wenk E. 2000. Using IRT to separate measurement bias from true group differences on homogeneous *and*

heterogeneous scales: an illustration with the MMPI. *Psychol. Methods* 5:125–46

Walters GL, Clopton JR. 2000. Effect of symptom information and validity scale information on the malingering of depression on the MMPI-2. *J. Pers. Assess.* 75:183–99

Watkins CE, Campbell VL, Nieberding R, Hallmark R. 1995. Contemporary practice of psychological assessment by clinical psychologists. *Prof. Psychol. Res. Pract.* 26:54–60

Weiner IB. 1999. What the Rorschach can do for you: incremental validity in clinical applications. *Assessment* 6:327–38

Weiner IB. 2000. Using the Rorschach properly in practice and research. *J. Clin. Psychol.* 56:435–38

Whitehead WC. 1985. *Clinical decision-making on the basis of Rorschach, MMPI, and automated MMPI report data*. PhD thesis. Univ. Tex. Health Sci. Center, Dallas

Widiger TA. 2001. The best and the worst of us? *Clin. Psychol. Sci. Pract.* 8:374–77

Widiger TA, Clark LA. 2000. Toward DSM-V and the classification of psychopathology. *Psychol. Bull.* 126:946–63

Wiebush RG, Baird C, Krisberg B, Onek D. 1995. Risk assessment and classification for serious, violent, and chronic juvenile offenders. In *Serious, Violent, and Chronic Juvenile Offenders: A Sourcebook*, ed. JC Howell, B Krisberg, JD Hawkins, JJ Wilson, pp. 171–212. Thousand Oaks, CA: Sage

Wood JM, Lilienfeld SO. 1999. The Rorschach Inkblot Test: a case of overstatement? *Assessment* 6:341–49

Wood JM, Lilienfeld SO, Garb HN, Nezworski MT. 2000a. Limitations of the Rorschach as a diagnostic tool: a reply to Garfield (2000), Lerner (2000), and Weiner (2000). *J. Clin. Psychol.* 56:441–48

Wood JM, Lilienfeld SO, Garb HN, Nezworski MT. 2000b. The Rorschach Test in clinical diagnosis: a critical review, with a backward look at Garfield (1947). *J. Clin. Psychol.* 56:395–430

Wood JM, Lilienfeld SO, Nezworski MT, Garb HN. 2001a. Coming to grips with negative evidence for the Comprehensive System for the Rorschach: a comment on Gacono, Loving, and Bodholdt (2001), Ganellen (2001), and Bornstein (2001). *J. Pers. Assess.* 77:48–70

Wood JM, Nezworski MT, Garb HN, Lilienfeld SO. 2001b. The misperception of psychopathology: problems with the norms of the Comprehensive System for the Rorschach. *Clin. Psychol. Sci. Pract.* 8:350–73

Wood JM, Nezworski MT, Garb HN, Lilienfeld SO. 2001c. Problems with the norms of the Comprehensive System for the Rorschach: methodological and conceptual considerations. *Clin. Psychol. Sci. Pract.* 8: 397–402

Wood JM, Nezworski MT, Stejskal WJ. 1996. The Comprehensive System for the Rorschach: a critical examination. *Psychol. Sci.* 7:3–10

Wood JM, Nezworski MT, Stejskal WJ, McKinzey RK. 2001d. Problems of the Comprehensive System for the Rorschach in forensic settings: recent developments. *J. Forensic Psychol. Pract.* 1:(3):89–103

Zalewski C, Greene RL. 1996. Multicultural usage of the MMPI-2. In *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications*, ed. LA Suzuki, PJ Meller, JG Ponterotto, pp. 77–114. San Francisco: Jossey-Bass

Zimmerman M, Mattia JI. 1999a. Differences between clinical and research practices in diagnosing borderline personality disorder. *Am. J. Psychiatry* 156:1570–74

Zimmerman M, Mattia JI. 1999b. Is posttraumatic stress disorder underdiagnosed in routine clinical settings? *J. Nerv. Ment. Dis.* 187:420–28