Evidence-Based Assessment of Personality Disorders

Thomas A. Widiger and Douglas B. Samuel
University of Kentucky

The purpose of this article is to provide a foundation for the development of evidence-based guidelines for the assessment of personality disorders, focusing in particular on integrated assessment strategies. The general strategy recommended herein is to first administer a self-report inventory to alert oneself to the potential presence of particular maladaptive personality traits followed by a semistructured interview to verify their presence. This strategy is guided by the existing research that suggests particular strengths of self-report inventories and semistructured interviews relative to unstructured clinical interviews. However, the authors also consider research that suggests that further improvements to the existing instruments can be made. The authors emphasize, in particular, a consideration of age of onset, distortions in self-perception and presentation, gender bias, culture and ethnicity, and personality change.

The purpose of this article is to provide a foundation for the development of evidence-based guidelines for the assessment of the personality disorders, including those in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; 1994). We focus on integrated assessment strategies rather than specific personality disorder instruments. We also emphasize issues that are considered to be fundamental to a valid personality disorder assessment, being informed by what is generally known about personality disorders, their developmental course, and their comorbidity. We also address matters of clinical utility, because an assessment strategy is unlikely to be effective if it is unrealistic to implement or fails to address matters of importance to clinical intervention.

Strategies

For the purpose of obtaining an accurate assessment of a *DSM–IV* personality disorder, the general strategy recommended herein is a two-step procedure: (a) Administer a self-report inventory to alert oneself to the potential presence of particular maladaptive personality traits, and (b) administer a semistructured interview to verify and document their presence (Widiger, 2002). This integrated strategy is guided by the existing evidence that suggests particular strengths of self-report inventories and semistructured interviews relative to a reliance on unstructured clinical interviews. However, no method is infallible, and there is compelling research to suggest ways in which existing self-report inventories and semistructured interviews can be improved. We discuss the rationale and empirical support for using semistructured interviews and self-report inventories and present potential limitations that could be addressed in future research.

Semistructured Interviews

The preferred method for assessing personality disorders in general clinical practice appears to be an unstructured clinical interview (Watkins, Campbell, Nieberding, & Hallmark, 1995; Westen, 1997), whereas the preferred method in research is the semistructured interview (Rogers, 2001; Segal & Coolidge, 2003; Zimmerman, 2003). Semistructured interviews have a number of advantages over unstructured interviews (Rogers, 2003). Semistructured interviews ensure and document that a systematic and comprehensive assessment of each personality disorder diagnostic criterion has, in fact, occurred. The administration of a semistructured interview is particularly advantageous in clinical situations in which the credibility or

Thomas A. Widiger and Douglas B. Samuel, Department of Psychology, University of Kentucky.

Correspondence concerning this article should be addressed to Thomas A. Widiger, Department of Psychology, University of Kentucky, Lexington, KY 40506-0044. E-mail: widiger@uky.edu

validity of the assessment might be questioned (e.g., forensic or disability evaluations), because the administration of the interview will document that the assessment was reasonably comprehensive, replicable, and objective.

Studies have indicated that diagnoses based on unstructured clinical interviews often fail to consider all of the necessary diagnostic criteria (Blashfield & Herkov, 1996; Zimmerman & Mattia, 1999). As Cantor and Genero (1986) demonstrated, "There is a tendency, once having categorized, to exaggerate the similarity among nonidentical stimuli by overlooking within-group variability, discounting disconfirming evidence, and focusing on stereotypic examples of the category" (p. 235). Clinicians tend to diagnose personality disorders hierarchically, failing to assess for additional symptoms once a particular disorder has been identified (Adler, Drake, & Teague, 1990; Blashfield & Flanagan, 1998; Herkov & Blashfield, 1995). The personality disorder that is provided preferential attention may even be based on idiosyncratic interests (Gunderson, 1992; Mellsop, Varghese, Joshua, & Hicks, 1982).

Studies also indicate that personality disorder assessments in the absence of structured clinical interviews can be quite unreliable (Mellsop et al., 1982; Spitzer, Forman, & Nee, 1979), and that semistructured interviews increase the likelihood that a reliable and replicable assessment will occur (Farmer, 2000; Rogers, 2001, 2003; Segal & Coolidge, 2003; Wood, Garb, Lilienfeld, & Nezworski, 2002). Semistructured interviews provide specific, carefully selected questions for the assessment of each diagnostic criterion, the application of which increases the likelihood that assessments will be consistent across interviewers. In addition, the manuals that accompany a semistructured interview will often provide a considerable amount of helpful information for understanding the rationale of each diagnostic criterion, for interpreting vague or inconsistent symptomatology, and for resolving diagnostic ambiguities (e.g., Loranger, 1999; Widiger, Mangine, Corbitt, Ellis, & Thomas, 1995).

Reliability in diagnosis is due as much to the instruments used for their assessment as it is to the availability of more specific and explicit criterion sets (Rogers, 2003; Segal & Coolidge, 2003). Researchers would be hard pressed to get their findings published if they failed to document that their diagnoses were based on a systematic, replicable, and objective method, yet no such requirements are currently provided for clinical diagnoses, with the exception of mental retardation and learning disorders. However, a recommendation of the APA and National Institute of Mental Health *DSM–V* Nomenclature Work Group is to consider incorporating more structured assessments into the *DSM–V* criterion sets to require that they be assessed systematically. "At present, results of psychological testing are not included in *DSM–IV* diagnostic criteria, with the exception of IQ testing and academic skills. . .[and] this exception points the way for research that could lead to incorporation of psychological test results as diagnostic criteria for other disorders" (Rounsaville et al., 2002, p. 24). Few clinicians would attempt to diagnose mental retardation in the absence of a structured test, yet this is the norm for most other diagnoses. A proposal for *DSM–V* is to make comparable requirements for the diagnosis of the anxiety, mood, personality, and other mental disorders (Rounsaville et al., 2002; Widiger & Clark, 2000).

Advocates of unstructured clinical interviews raise a compelling criticism that semistructured interviews can degenerate into a mindless symptom counting that fails to consider adequately the context of a maladaptive behavior in a broader life history or to appreciate the manner in which the person relates to the interviewer over an extended period of time (Perry, 1992; Westen, 1997). Excellent construct validity in the assessment of personality disorders by practicing clinicians using unstructured interviews has been obtained with the Shedler-Westen Assessment Procedure-200 (SWAP-200), "a method for studying personality and personality pathology that strives to capture the richness and complexity of psychoanalytic constructs and formulations without forsaking the benefits of empirical rigor" (Shedler, 2002, p. 429). The SWAP-200 is a clinician ranking form of 200 items, drawn from the psychoanalytic and personality disorder literature (Shedler, 2002; Westen & Shedler, 1999a), although one does not have to have a psychodynamic orientation to use the SWAP-200 in a reliable and valid manner. SWAP-200 items are not ranked on the basis of a systematic administration of a series of questions; instead, the SWAP-200 relies on "the empathically attuned

and dynamically sophisticated clinician given free rein to practice his or her craft" (Shedler, 2002, p. 433). Initial research with the SWAP-200 has reported good to excellent convergent and discriminant validity (e.g., Westen & Shedler, 1999b; Westen, Shedler, Durrett, Glass, & Martens, 2003).

However, the positive results obtained with the SWAP-200 should perhaps be tempered by potential methodological limitations of the initial research. For example, clinicians who have provided the personality disorder criterion rankings have generally been the same persons who have provided the SWAP-200 rankings. This is comparable to having the interviewers provide the criterion diagnoses in a study addressing the convergent validity of a semistructured interview. On the other hand, Westen and Muderrisoglu (2003) did report strong convergence between clinician-based SWAP-200 and *DSM–IV* diagnoses provided by independent interviews. An additional methodological concern is that the clinicians in each study were provided with guidelines for the distribution of their rankings. For example, in Westen and Shedler (1999b), the clinicians were required to identify half of the personality disorder symptoms as being absent and only eight SWAP-200 items could be given the highest rankings, no matter the actual opinions of the clinicians or the symptoms that were, in fact, present (similar constraints were placed on the other rankings). Discriminant validity of a semistructured interview would also be improved dramatically if the interviewers were instructed to code half of the diagnostic criteria as absent and to provide only a few of the diagnostic criteria with the highest ratings. However, it is also possible that the improvement in the discriminant validity provided by the SWAP-200 is not due largely to the encouragement of the clinicians to adhere to a particular distribution of scores. This hypothesis could be tested in future studies in which no particular distribution of scores is encouraged.

Self-Report Inventories

Zimmerman and Mattia (1999) reported that clinicians adjusted their personality disorder diagnoses when informed of the results provided by a semistructured interview, "inconsistent with the notion that personality disorder diagnoses based on semistructured interviews are

not viewed as valid by clinicians" (p. 1570). Nevertheless, clinicians will be understandably reluctant to administer an entire semistructured interview because of the amount of time that would be required. We, therefore, recommend that one first administer a self-report inventory to identify which personality disorders should be emphasized during the interview and which disorders could be ignored with minimal risk (Lenzenweger, Loranger, Korfine, & Neff, 1997; Widiger, 2002). In most instances, this will reduce substantially the number of personality disorders that would need to be assessed with a semistructured interview. In some instances, patients could obtain clinically significant elevations on a substantial number of self-report scales, but these would also be cases in which an extensive and thorough interview should be conducted anyway. Brief screening questionnaires can also be used, although there might be little advantage in using a screening instrument in preference to an inventory that was constructed to provide a comprehensive and valid assessment. Self-report inventories also have the additional advantage of including validity scales that can alert the clinician to response sets, biases, and distortions that might compromise the validity of the clinical assessment (e.g., Millon Clinical Multiaxial Inventory–III [MCMI–III]; Millon, Millon, & Davis, 1997; and the Minnesota Multiphasic Personality Inventory—2 [MMPI–2]; Colligan, Morey, & Offord, 1994). There is a semistructured interview devoted to the assessment of malingering and deception (Rogers, 1997), but the personality disorder semistructured interviews do not themselves currently include validity scales, relying instead on the expertise of the clinician to detect responses sets and biases. A further benefit of self-report inventories is that they can alert the clinician to personality disorders that might have otherwise been missed as a result of false expectations or assumptions (e.g., failing to notice histrionic personality traits in male patients).

A potential limitation of our recommended strategy is the occurrence of false-negative assessments by the self-report inventory. The existing research suggests that self-report inventories, on the contrary, err heavily in the direction of false-positive rather than false-negative errors (L. A. Clark & Harrison, 2001; Farmer, 2000; Kaye & Shea, 2000; Widiger & Coker,

2002). Nevertheless, false-negative errors could occur, one potential source of which would be denial or inadequate insight on the part of the respondent that is not addressed effectively by an inventory's indirect items or validity scales. Projective tests could conceivably be more successful than self-report inventories in circumventing respondent denial (Wiggins, 2003). Compelling concerns, though, have been raised regarding the validity and utility of projective tests (Wood et al., 2002), and there is currently only limited research on the use of projective tests for the assessment of most of the *DSM–IV* personality disorders. Nevertheless, even the toughest critics of projective tests do acknowledge that some Rorschach indices possess moderate validity, including, for example, the Rorschach Oral Dependency score (Bornstein, 1999; Lilienfeld, Wood, & Garb, 2000; Wood et al., 2002). Comparable empirical support for the assessment of other maladaptive personality traits is needed if projective tests are to be included within an empirically valid assessment of the *DSM–IV* personality disorders (e.g., Blais, Hilsenroth, Castlebury, Fowler, & Baity, 2001; Hilsenroth, Fowler, Padawer, & Handler, 1997).

Psychometric Properties

Quite a few personality disorder self-report inventories and semistructured interviews have been developed. A complete summary of them is beyond the scope of this article, but a number of extensive reviews of each of them have been published (L. A. Clark & Harrison, 2001; Farmer, 2000; Kaye & Shea, 2000; Rogers, 2001; Widiger, 2002; Widiger & Coker, 2002). We briefly consider herein general concerns regarding their reliability, convergent validity, discriminant validity, and normative data.

Reliability

Many studies indicate that good to excellent interrater reliability is obtained in the assessment of personality disorders when semistructured interviews are used (Rogers, 2001, 2003; Segal & Coolidge, 2003). Nevertheless, it is also worth noting that the reliability data that are reported in most studies have been confined to the agreement in the coding of respondents' answers to interview questions. This may not be the more important or fundamental concern with respect to the reliability of a personality disorder assessment (L. A. Clark & Harrison, 2001). As the structure of an interview increases, the reliability of response coding can be no more demanding than obtaining agreement as to whether respondents said "yes" or "no" in response to a straightforward question. Of greater importance to the replicability of clinical and research assessments are studies addressing whether semistructured interviews are administered reliably (Segal & Coolidge, 2003). For example, are the questions being administered by different interviewers in a consistent manner? Are some interviewers providing substantially more follow-up queries than other interviewers? Do patients respond to the same open-ended questions in a consistent manner over time? Sophisticated reliability studies are being conducted (e.g., Lenzenweger, 1999; Trull, 2001; Zanarini, Frankenburg, & Vujanovic, 2002), but further research is needed on the agreement between independent administrations of the same interview to the same patient.

Convergent Validity

There are five semistructured interviews for the assessment of the 10 *DSM–IV* (American Psychiatric Association, 1994) personality disorders: (a) Diagnostic Interview for Personality Disorders (DIPD; Zanarini, Frankenburg, Chauncey, & Gunderson, 1987); (b) International Personality Disorder Examination (IPDE; Loranger, 1999); (c) Personality Disorder Interview–IV (PDI–IV; Widiger et al., 1995); (d) Structured Clinical Interview for *DSM–IV* Axis II Personality Disorders (SCID–II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997); and (e) Structured Interview for *DSM–IV* Personality Disorders (Pfohl, Blum, & Zimmerman, 1997). Only three studies have provided data concerning their convergent validity (O'Boyle & Self, 1990; Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991; Pilkonis, Heape, Proietti, Clark, McDavid, & Pitts, 1995), only two of these studies administered the interview schedules to the same patients (O'Boyle & Self, 1990; Skodol et al., 1991), and all three studies were confined to just two of the five semistructured interviews.

The most comprehensive study to date was by Skodol et al. (1991). They administered the IPDE and SCID–II to 100 inpatients of a per-

sonality disorders treatment unit. Both interviews were administered blind to one another on the same day (one in the morning, the other in the afternoon). Order of administration was staggered. Kappa for individual diagnoses ranged from a low of .14 (schizoid) to a high of .66 (dependent), with a median kappa of .53 (borderline). The authors considered the agreement for some of the categorical diagnoses to be discouraging. "It is fair to say that, for a number of disorders (i.e., paranoid, schizoid, schizotypal, narcissistic, and passive–aggressive) the two [interviews] studied do not operationalize the diagnoses similarly and thus yield disparate results (Skodol et al., 1991, p. 22). However, the median agreement obtained for the personality disorders was consistent with agreement rates obtained for the diagnosis of many Axis I mental disorders when their assessments are conducted blind to one another (Loranger, 1992). In addition, agreement with respect to a more quantitative assessment of the extent to which each personality disorder was present was considerably better, with correlations ranging from a low of .58 (schizoid) to a high of .87 (antisocial). Skodol et al. (1991) concluded that "the greater agreement shown by comparing dimensions of disorder than by comparing strict categorical diagnoses suggests that patients are providing interviewers with reliable information about areas of difficulty in personality functioning and interviewers are able to judge when at least some of these reports indicate clinically significant psychopathology" (p. 22).

A number of studies have been published on the convergent validity of semistructured interviews with self-report inventories as well as the convergent validity among self-report inventories. Widiger and Coker (2002) tabulated the findings from 41 of these studies. It is apparent from this research that convergent validity increases as the structure of the assessment increases. The weakest convergent validity has been obtained with unstructured clinical interviews; the highest has been obtained with self-report inventories; the convergence of semistructured interviews with self-report inventories falls somewhere in between.

One implication of the convergent validity research is that self-report inventories might provide a more valid assessment of personality disorders than semistructured interviews. Semistructured interviews are generally preferred over self-report inventories in clinical research (Rogers, 2001; Segal & Coolidge, 2003), and they are often used as the criterion measure with which the validity of a self-report inventory is tested (Zimmerman, 2003). Rarely are self-report inventories used as criterion measures for the validity of a semistructured interview (although there are exceptions; e.g., Trull et al., 1998). The preference of researchers for semistructured interviews, however, could be analogous to the preference of clinicians for unstructured interviews. One reason why semistructured interviews are preferred over self-report inventories is that they provide the researcher the opportunity to have a direct, personal impact on the assessment; follow-up queries can be provided and inadequacies in self-insight and awareness can be addressed (Kaye & Shea, 2000; Rogers, 2001; Segal & Coolidge, 2003). However, the opportunity of the interviewer to personally impact the assessment might also contribute to less reliable and ultimately less valid assessments (comparable to the lower reliability obtained with unstructured interviews relative to semistructured interviews). The findings obtained with self-report inventories are more likely to replicate across research sites than the findings obtained with semistructured interviews (L. A. Clark, Vittengl, Kraff, & Jarrett, 2003) simply because there is little to no room for interrater disagreement in the administration and scoring of a self-report inventory (considered by some, though, to be a limitation rather than a strength). There are data to suggest that semistructured interviews might be more successful than self-report inventories in differentiating maladaptive personality traits from other mental disorders (an issue discussed further later), but, given the considerable expense of administering semistructured interviews, additional research on the relative validity of these two methods of assessment is perhaps warranted.

Discriminant Validity

Only a few studies have provided discriminant validity data (Widiger & Coker, 2002). The absence of much attention to discriminant validity reflects in part a recognition that the diagnostic constructs assessed by these measures do not themselves have compelling discriminant validity (L. A. Clark & Harrison, 2001;

Farmer, 2000). For example, the report by Skodol et al. (1991) was confined to the convergent validity of the IPDE and SCID–II and did not discuss or provide any data on discriminant validity. Instead, Oldham et al. (1992) subsequently used the same data to report an excessive co-occurrence among the personality disorder diagnostic categories and concluded that most of this was due to overlap among the disorders' criterion sets rather than flaws or inadequacies of the IPDE or SCID–II. Because the *DSM–IV* personality disorders overlap extensively (Bornstein, 1998; Farmer, 2000), a valid assessment of an individual personality disorder should perhaps obtain weak discriminant validity with respect to its near neighbor diagnostic constructs. For example, perhaps a valid assessment of borderline personality disorder should not result in the absence of overlap with the dependent, histrionic, and narcissistic personality disorders. The scales of some personality disorder self-report inventories (e.g., MCMI–III and MMPI–2) overlap substantially in order to compel the obtainment of a particular degree and direction of co-occurrence that would be consistent with theoretical expectations.

Morey et al. (2002) administered the NEO Personality Inventory—Revised (NEO-PI–R; Costa & McCrae, 1992), a self-report measure of the five-factor model (FFM) of general personality functioning, to 86 patients diagnosed with schizotypal, 175 with borderline, 157 with avoidant, and 153 with obsessive–compulsive personality disorder in the multisite Collaborative Longitudinal Personality Disorders Study (CLPS). A discriminant function analysis indicated that the four personality disorders were differentiated significantly in terms of the 30 facets of the FFM, but it was also apparent from a visual inspection of the FFM profiles of each personality disorder that "all four of the disorders displayed a similar configuration of FFM traits" (Morey et al., 2002, p. 229). However, there was considerable diagnostic co-occurrence among the personality disorders in their sample. Morey et al. repeated the analyses using a subsample of 24 schizotypal, 72 borderline, 103 avoidant, and 105 obsessive–compulsive personality-disordered patients who did not meet criteria for one of the three other respective personality disorders under study. "The elimination of patients with comorbid study di-

agnoses did appear to sharpen the distinction between the personality disorder groups, whereas only 18 facets revealed substantive differences (i.e., effect sizes larger than .50) among the cell-assigned personality disorder diagnoses, 31 facets achieved this threshold using the non-comorbid groups" (pp. 224–225). Differentiation would probably increase further if the additional diagnostic co-occurrence with the six other personality disorders was also considered. In other words, it is unclear whether Morey et al.'s results indicated weak discriminant validity for the FFM or for the personality disorder constructs. Lynam and Widiger (2001) indicated that much of the diagnostic co-occurrence among the *DSM–IV* personality disorders can, in fact, be explained well by their overlapping FFM personality trait profiles.

Normative Data

None of the test manuals for any of the *DSM–IV* personality disorder semistructured interviews provide normative data. In contrast, a substantial amount of normative data have been obtained and reported for some of the self-report inventories (e.g., Colligan et al., 1994; Millon et al., 1997). The test manuals for semistructured interviews are, in fact, quite weak in their coverage of reliability and validity data (Kaye & Shea, 2000; Rogers, 2001). Diagnoses obtained through the administration of a semistructured interview are used as the criterion by which the validity of other instruments is evaluated (Kaye & Shea, 2000; Zimmerman, 2003), but semistructured interviews might be relying too heavily on simple face validity for their own derivation (Farmer, 2000; Segal & Coolidge, 2003). In defense of the validity of semistructured interviews, the most compelling published research concerning etiology, course, pathology, and treatment has been based on assessments provided by semistructured interviews (Rogers, 2001). The results of this extensive research provide considerable support for the construct validity of the interviews used in these studies. Nevertheless, a clinician's or researcher's interpretation of the results obtained by a particular semistructured interview would be improved substantially if reliability, validity, and normative data were provided within their respective test manuals.

Further Issues for an Evidenced-Based Assessment

No method of assessment is infallible, and further improvements to the existing self-report inventory and semistructured interviews can be made. We emphasize in particular here such issues as age of onset, distortions in self-perception and presentation, gender bias, culture and ethnicity, and treatment evaluation.

Age of Onset

Personality traits and personality disorders, by definition, have an age of onset that "can be traced back at least to adolescence or early adulthood" (American Psychiatric Association, 1994, p. 689), consistent with most theoretical models of personality functioning. An onset no later than young adulthood is also helpful in avoiding the confusion of a mood, anxiety, substance dependence, or other Axis I disorder with a personality disorder (Triebwasser & Shea, 1996). Personality disorders are often comorbid with anxiety, mood, eating, substance, and other Axis I disorders (Dolan-Sewell, Krueger, & Shea, 2001), and one of the more well-established and consistently replicated findings is the considerable effect of Axis I psychopathology on the assessment of personality (Farmer, 2000; Widiger & Coker, 2002). Diagnostic assessments are needed most at the beginning of treatment, yet it is at this time that it can be most difficult to differentiate between personality and Axis I disorders. Most persons with a personality disorder probably seek treatment when they are in crisis or at least experiencing acute levels of distress, anxiety, or depression, and persons who are significantly anxious, depressed, angry, or distraught will often fail to provide an accurate description of their usual way of thinking, feeling, behaving, and relating to others. Requiring that the assessment of a personality disorder document presence back to late adolescence or young adulthood is one means to ensure that the personality disorder was, in fact, present prior to the onset of a current Axis I disorder (Triebwasser & Shea, 1996).

Personality disorder assessment instruments, however, vary substantially in how rigorously they address the *DSM–IV* age of onset requirement. Self-report inventories are particularly weak in this regard. For example, the MCMI–III (Millon et al., 1997) instructs respondents to answer the questions in reference to their current problems. There is no instruction to describe one's characteristic manner of thinking, feeling, behaving, or relating to others prior to the onset of a recent mental disorder. Semistructured interviews make more of an effort, but they also vary substantially in the manner and extent to which they do so. The PDI–IV (Widiger et al., 1995) encourages but does not require the interviewer to document that each diagnostic criterion was evident in young adulthood. The IPDE (Loranger, 1999) is more explicit in its requirements but is also more liberal, as it requires that only one diagnostic criterion for a respective personality disorder be present since the age of 25; all of the others can be evident only within the past few years. The SCI-D–II (First et al., 1997) generally requires that each diagnostic criterion be evident over a 5-year period; the DIPD (Zanarini et al., 1987) focuses its assessment on the prior 2 years.

The DIPD is being used in the CLPS (Gunderson et al., 2000), which is contributing a substantial amount of useful information on the course and treatment of the borderline, avoidant, schizotypal, and obsessive–compulsive personality disorders. One of the more intriguing findings of this research program has been the extent to which persons fail to maintain personality disorder symptomatology. For example, 23 of 160 persons (14%) diagnosed with borderline personality disorder (BPD) at the study's baseline assessment met criteria for two or fewer of the nine diagnostic criteria 6 months later (Gunderson et al., 2003). Eighteen sustained this reduction from 6 months to 1 year. Gunderson et al. (2003) concluded that only 1 of these 18 persons had been inaccurately diagnosed at baseline; the rest were considered to be valid instances of sudden and dramatic remission. However, it is difficult to imagine so many persons who met the diagnostic criteria for BPD since late childhood and who continued to manifest these symptoms throughout their adult life experienced, apparently for the first time, dramatic changes in personality functioning soon after the onset of the study. For example, the purportedly valid diagnoses include 1 person whose original symptoms were determined to be secondary to the use of a stimulant for weight reduction: "The most dramatic improvement following a treatment intervention occurred

when a subject discontinued a psychostimulant she had used the year prior to baseline for purposes of weight loss. . .Discontinuation was followed by a dramatic reduction of her depression, panic, abandonment fears, and self-destructiveness" (Gunderson et al., 2003, p. 116). For other cases, "the changes involved gaining relief from severely stressful situations they were in at or before the baseline assessment" (p. 115), including the resolution of a traumatic divorce or custody battle. The CLPS project is providing findings that are increasing our understanding of the course of personality disorders, but it may also be helpful in alerting clinicians and researchers to additional complications and difficulties in obtaining valid assessments.

Inaccuracy in Respondents' Perception, Insight, and Presentation

One of the principal arguments for using unstructured clinical interviews and projective tests to assess personality disorders is that they might be less susceptible to the distortions in perception, insight, and presentation that are evident in persons with personality disorders (Bornstein, 1999; Hilsenroth, Handler, & Blais, 1996; Westen, 1997). The self-descriptions of persons who are characterized in part by a grandiose self-image, a lack of honesty, or a wary suspiciousness should not be taken at face value (Kaye & Shea, 2000). Semistructured interviews and self-report inventories have at times been characterized as naively direct inquiries as to the presence of each of the *DSM–IV* diagnostic criteria (Shedler, 2002; Westen, 1997). There are indeed items within most of the inventories and interviews that have at least the appearance of naively trusting a person's insight or forthrightness.

However, it is not the case that semistructured interviews are simply a series of direct questions as to the presence of each *DSM–IV* diagnostic criterion. Semistructured interviews are "semi"-structured because they include many open-ended questions and indirect inquiries as well as observations of the respondents' manner of responding and relating to the interviewer (Segal & Coolidge, 2003). Interviewers do not simply record respondents' answers to direct questions. They follow up respondents' answers with further queries to ensure that a diagnostic criterion is, in fact, present (or absent). The diagnostic rating is not simply that patient's opinion regarding the presence of each diagnostic criterion; it is, instead, the interviewer's opinion based on the substantial amount of information that was generated by the semistructured interview.

Nevertheless, some interviews may obtain inadequate or inaccurate information as a result of the lack of cooperation or insight of a respondent. The Psychopathy Checklist–Revised (PCL–R; Hare, 1991) relies heavily on an inmate's criminal record to rate many of its diagnostic criteria for psychopathy, because it is believed that psychopathic persons cannot be trusted to provide accurate self-descriptions. For example, the assessment of a lack of empathy is often based on the nature of the crimes that a person has committed rather than the person's responses to verbal queries (Hare, 1991). The reliance of the PCL–R on the availability of a criminal record, however, complicates its usage within other settings that lack detailed institutional records (Kaye & Shea, 2000) and makes it difficult to know the basis on which the PCL-R diagnostic criteria are, in fact, being assessed (Salekin, Rogers, & Sewell, 1996).

"To help overcome this difficulty, supplementary information from other informants may be helpful" (American Psychiatric Association, 1994, p. 686). These informants (e.g., spouses, friends, or colleagues) can be acutely aware of and perhaps more able or willing to describe the target person's arrogance, dishonesty, suspiciousness, hostility, or dependency (Bernstein et al., 1997). Agreement between self-descriptions and peer descriptions of personality traits has generally been good to excellent when sampling within nonclinical populations (McCrae, Stone, Fagan, & Costa, 1998), but agreement between self-descriptions and informant descriptions of personality disorders within clinical samples has at times been only poor to adequate, with considerable variation in convergence across the personality disorders (Klonsky, Oltmanns, & Turkheimer, 2002). The existence of this disagreement raises many fundamental questions, not the least of which is how best to resolve the disagreement and which source of information is providing the more valid information. Some studies have suggested that the self (e.g., patient) is providing the most

valid descriptions (e.g., Dreesen, Hildebrand, & Arntz, 1998), whereas others have suggested that the informant can be the more valid source (Klein, 2003). Most studies suggest that both sources of information can provide uniquely valid information (Klein, 2003; Klonsky et al., 2002; Kraemer et al., 2003; Ready, Watson, & Clark, 2002). There does not yet appear to be a consistent explanation for the disagreement. "With the exception of trait observability, [research] has failed to identify the moderators of self-informant agreement" (Ready & Clark, 2002, p. 46). The ambiguity of this research does at least argue for the importance of obtaining multiple sources of input for an assessment and for conducting additional research into the bases for the disagreement.

Gender

Many of the personality disorders have a differential sex prevalence rate, and some appear to involve maladaptive variants of gender-related personality traits. The suggestion that these differential sex prevalence rates reflect gender biases has been among the more difficult and heated diagnostic issues. Gender bias concerns have been raised with respect to the conceptualization of personality disorders, the wording of diagnostic criteria, the application of diagnostic criteria by clinicians, thresholds for diagnosis, clinical presentation, research sampling, the self-awareness and openness of respondents, and the items included within self-report inventories (Bornstein, 1996; Cale & Lilienfeld, 2002; Morey, Warner, & Boggs, 2002; Sprock, Crosby, & Nielsen, 2001; Widiger, 1998; Zlotnick, Rothschild, & Zimmerman, 2002). Most, if not all, of these concerns (and those that follow in the Ethnicity and Culture section) apply to other domains of psychopathology (as well as to the assessment of personality in general), but our discussion is confined to the assessment of personality disorders.

Studies have indicated that the failure of clinicians to assess diagnostic criterion sets in a thorough or systematic manner has contributed to excessive diagnoses of histrionic personality disorder in women (Garb, 1997). There have not yet been any studies to indicate whether semistructured interviews are prone to gender-biased assessments, although gender biases that are inherent within the diagnostic criterion sets would be evident in the findings obtained by a semistructured interview (Bornstein, 1996; Sprock et al., 2001; Vitale & Newman, 2001).

Studies have suggested that some self-report inventories are providing gender-biased assessments (Widiger, 1998). Some self-report inventories include gender-related items that are keyed in the direction of adaptive rather than maladaptive functioning. An item need not assess for dysfunction to contribute to a valid assessment of personality disorders. For example, items assessing for gregariousness can identify histrionic persons, items assessing for confidence can identify narcissistic persons, and items assessing conscientiousness can identify obsessive–compulsive persons (Millon et al., 1997). Items keyed in the direction of adaptive rather than maladaptive functioning can also be helpful in countering the tendency of some respondents to deny or minimize personality disorder symptomatology. However, these items will not be useful in differentiating abnormal from normal personality functioning and are likely to contribute to the overdiagnosis of personality disorders in normal or minimally dysfunctional populations, such as student counseling centers, child custody disputes, and personnel selection (Boyle & Le Dean, 2000). When these items are related to the sex or gender of respondents, as many are in the case of the histrionic, dependent, narcissistic, and obsessive–compulsive personality disorder scales of the MCMI-III (Millon et al. (1997) and MMPI-2 (Colligan et al., 1994), they may contribute to gender-biased assessments (Lampel, 1999; Lindsay, Sankis, & Widiger, 2000).

Culture and Ethnicity

One might expect considerable variation in the diagnosis and assessment of personality disorders across different cultural and ethnic groups (Alarcon, 1996). *DSM–IV* narcissistic personality disorder is not even included within the World Health Organization's (1992) international classification of mental disorders. However, there has been relatively little research on the impact of culture and ethnicity on the diagnosis or assessment of personality disorders (Alarcon, 1996; Cooke, 1996).

Items within self-report inventories are generally written from the perspective of a member of the predominant ethnic, cultural group, and

such items may not have the same meaning or implications when provided to members of a minority ethnic group (Okazaki, Kallivayalil, & Sue, 2002). Hindering the effort of psychologists to identify the cultural contexts in which assessment techniques should be interpreted differently, or the adjustments in test interpretation that should be made across different ethnic groups, is the absence of much research on the mechanisms for cultural or ethnic group differences. Much of the existing research has been confined to the reporting of group differences, without an assessment of the purported mechanism by which the differences could be explained or understood (Okazaki et al., 2002). For example, studies have reported significantly higher scores obtained by African Americans (compared with Caucasian Americans) on the paranoid personality disorder scales of the MC-MI–III (e.g., Frueh, Smith, & Libet, 1996), but there has not yet been any published research that has attempted to explain or account for these group differences. One possible social–cultural explanation for the different elevations is the presence of racial discrimination and prejudice (Whaley, 1997).

R. Clark, Anderson, Clark, and Williams (1999) documented well the importance of considering racism as a stressor on the psychological functioning of African Americans. Membership within a minority ethnic group that has historically been severely mistreated and exploited, and does still experience prejudicial, discriminatory, and antagonistic behaviors, would understandably contribute to feelings of mistrust, skepticism, and suspicion that would not be shared by members of the majority ethnic group. African Americans who have experienced a history of racial discrimination might respond differently than Caucasian Americans to such paranoid personality disorder items as "I am sure I get a raw deal from life" or "The people I work with are not sympathetic with my problems" (Colligan et al., 1994; Millon et al., 1997). Similar hypotheses might be generated for the interpretation of personality disorder test items by members of other ethnic or cultural groups.

Clinical Utility

The procedure we recommend for the assessment of personality disorders in clinical practice is to first administer a self-report inventory, followed by a semistructured interview that focuses on the disorders that received elevated scores. This procedure is in part an effort to be responsive to the time constraints of clinical practice, wherein the administration of an entire semistructured interview is not feasible. A more substantive matter of clinical utility is the extent to which the assessments contribute to treatment planning and outcome (First et al., in press; Zimmerman, 2003). The assessment of personality disorders should be of considerable importance to treatment planning given the extensive documentation that the presence of maladaptive personality traits can significantly disrupt, impair, or otherwise alter routine treatments of other mental disorders (Dolan-Sewell et al., 2001). This was, in fact, one of the major reasons why personality disorders were placed on a separate axis of the diagnostic manual (Frances, 1980). There are also compelling studies to indicate that maladaptive personality traits provide a vulnerability to future Axis I disorders that should perhaps be a focus of treatment (e.g., dependent personality traits as a vulnerability to the development of future mood disorders; Widiger & Bornstein, 2001). Much of this research has been conducted with semistructured interviews and self-report inventories.

The unique axis placement, however, has also contributed to a misperception that personality disorders are themselves untreatable (Widiger, 2003). Personality disorders were placed on a separate axis because of their particular importance to treatment decisions, not to suggest that they were themselves untreatable. In fact, systematic reviews of the treatment outcome research suggest that, on the contrary, cost-effective and personally meaningful changes to personality functioning do occur (American Psychiatric Association, 2001; Perry, Banon, & Ianni, 1999; Sanislow & Mc-Glashan, 1998), and much of this research has again been conducted with semistructured interviews and self-report inventories.

However, some of the existing personality disorder semistructured interviews might not be optimally suited for the assessment of changes in personality functioning (Shea, 1997). The intention of the existing instruments is to determine whether maladaptive personality traits have been evident since childhood, not to de-

termine whether there has been a recent or meaningful change to personality functioning. These two goals might not be entirely compatible. It is also unclear how long a personality trait should appear to be absent before it is considered to be treated successfully. An innovation of the DIPD is the inclusion of a monthly "follow-along" assessment of personality disorder symptoms as well as an exploration of various durations in time that might be useful for determining whether and when a personality disorder has gone into remission (Gunderson et al., 2000, 2003).

The research on changes in personality functioning secondary to treatment also suggests that assessments of outcome should focus on components of maladaptive personality functioning (e.g., impulsivity, self-destructiveness, callousness, feelings of vulnerability) rather than being more globally concerned with the presence versus absence (or remission) of a particular personality disorder (Perry et al., 1999; Sanislow & McGlashan, 1998). Personality disorders involve constellations of maladaptive personality traits, not all of which will be equally responsive to a clinical intervention. Semistructured interviews that are devoted to a particular disorder generally provide scales for these components (e.g., the Revised Diagnostic Interview for Borderlines; Zanarini et al., 2002), as do instruments that provide assessments of dimensional models of personality disorder. These latter instruments, such as the Dimensional Assessment of Personality Pathology-Basic Questionnaire (Livesley & Jackson, in press), Schedule for Nonadaptive and Adaptive Personality (L. A. Clark, 1993), the NEO-PI–R (Costa & McCrae, 1992), or Structured Interview for the Five-Factor Model (Trull et al., 1998), assess dimensions of maladaptive personality functioning (e.g., anxiousness, self-harm, intimacy problems, workaholism, introversion) that are hypothesized to underlie the existing diagnostic categories.

Conclusions

No method for the assessment of personality disorders is infallible. Using multiple methods whose errors are uncorrelated is preferable to relying on any single instrument. The general strategy recommended herein is to first administer a self-report inventory to avoid unnecessary interviewing and to alert the clinician to maladaptive personality functioning that might not have otherwise been anticipated, followed by a semistructured interview to assess systematically the respective diagnostic criteria of the disorders that were elevated on the self-report inventory. The administration of instruments to an informed source (e.g., relative, friend, or close colleague) is also likely to provide additional valid information. However, the validity of this strategy can be improved by addressing further issues, including age of onset, distortions in self-perception and presentation, gender bias, culture and ethnicity, and personality change.

References

Adler, D. A., Drake, R. E., & Teague, G. B. (1990). Clinicians' practices in personality assessment: Does gender influence the use of *DSM–III* Axis II? *Comprehensive Psychiatry, 31,* 125–133.

Alarcon, R. D. (1996). Personality disorders and culture in *DSM–IV*: A critique. *Journal of Personality Disorders, 10,* 260–270.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2001). *Practice guidelines for the treatment of patients with borderline personality disorder.* Washington, DC: Author.

Bernstein, D. P., Kasapis, C., Bergman, A., Weld, E., Mitropoulou, V., Horvath, T., et al. (1997). Assessing Axis II disorders by informant interview. *Journal of Personality Disorders, 11,* 158–167.

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.

Blashfield, R. K., & Flanagan, E. (1998). A prototypic nonprototype of a personality disorder. *Journal of Nervous and Mental Disease, 186,* 244–246.

Blashfield, R. K., & Herkov, M. J. (1996). Investigating clinician adherence to diagnosis by criteria: A replication of Morey and Ochoa (1989). *Journal of Personality Disorders, 10,* 219–228.

Bornstein, R. F. (1996). Sex differences in dependent personality disorder prevalence rates. *Clinical Psychology: Science and Practice, 3,* 1–12.

Bornstein, R. F. (1998). Reconceptualizing personality disorder diagnosis in the *DSM-V*: The discriminant validity challenge. *Clinical Psychology: Science and Practice, 5,* 333–343.

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

Boyle, G. J., & Le Dean, L. (2000). Discriminant validity of the Illness Behavior Questionnaire and Millon Clinical Multiaxial Inventory-III in a heterogeneous sample of psychiatric outpatients. *Journal of Clinical Psychology, 56,* 779–791.

Cale, E. M., & Lilienfeld, S. O. (2002). Sex differences in psychopathy and antisocial personality disorder. A review and integration. *Clinical Psychology Review, 22,* 1179–1207.

Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon & G. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 233–256). New York: Guilford Press.

Clark, L. A. (1993). *Manual for the Schedule for Nonadaptive and Adaptive Personality.* Minneapolis: University of Minnesota Press.

Clark, L. A., & Harrison, J. A. (2001). Assessment instruments. In W. J. Livesley (Ed.), *Handbook of personality disorders. Theory, research, and treatment* (pp. 277–306). New York: Guilford Press.

Clark, L. A., Vittengl, J., Kraft, D., & Jarrett, R. B. (2003). Separate personality traits from states to predict depression. *Journal of Personality Disorders, 17,* 152–172.

Clark, R., Anderson, N. B., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans. *American Psychologist, 54,* 805–816.

Colligan, R. C., Morey, L. C., & Offord, K. P. (1994). MMPI/MMPI-2 personality disorder scales. Contemporary norms for adults and adolescents. *Journal of Clinical Psychology, 50,* 168–200.

Cooke, D. J. (1996). Psychopathic personality in different cultures: What do we know? What do we need to find out? *Journal of Personality Disorders, 10,* 23–40.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Dolan-Sewell, R. G., Krueger, R. F., & Shea, M. T. (2001). Co-occurrence with syndrome disorders. In W. J. Livesley (Ed.), *Handbook of personality disorders* (pp. 84–104). New York: Guilford Press.

Dreesen, L., Hildebrand, M., & Arntz, A. (1998). Patient-informant concordance on the Structured Clinical Interview for *DSM–III–R* personality disorders (SCID-II). *Journal of Personality Disorders, 12,* 149–161.

Farmer, R. F. (2000). Issues in the assessment and conceptualization of personality disorders. *Clinical Psychology Review, 20,* 823–851.

First, M., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the Structured Clinical Interview for DSM–IV Axis II personality disorders.* Washington, DC: American Psychiatric Press.

First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (in press). Using clinical utility as a criterion for revising psychiatric diagnoses. *American Journal of Psychiatry.*

Frances, A. J. (1980). The *DSM–III* personality disorders section: A commentary. *American Journal of Psychiatry, 137,* 1050–1054.

Frueh, B. C., Smith, D. W., & Libet, J. M. (1996). Racial differences on psychological measures in combat veterans seeking treatment for PTSD. *Journal of Personality Assessment, 66,* 41–53.

Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice, 4,* 99–120.

Gunderson, J. G. (1992). Diagnostic controversies. In A. Tasman & M. B. Riba (Eds.), *Review of psychiatry* (Vol. 11, pp. 9–24). Washington, DC: American Psychiatric Press.

Gunderson, J. G., Bender, D., Sanislow, C., Yen, S., Rettew, J. B., Dolan-Sewell, R., et al. (2003). Plausibility and possible determinants of sudden "remissions" in borderline patients. *Psychiatry, 66,* 111–119.

Gunderson, J. G., Shea, M. T., Skodol, A. E., McGlashan, T. H., Morey, L. C., Stout, R. L., et al. (2000). The Collaborative Longitudinal Personality Disorders: Study I. Development, aims, design, and sample characteristics. *Journal of Personality Disorders, 14,* 300–315.

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised manual.* North Tonawanda, NY: Multi-Health Systems.

Herkov, M. J., & Blashfield, R. K. (1995). Clinicians' diagnoses of personality disorder: Evidence of a hierarchical structure. *Journal of Personality Assessment, 65,* 313–321.

Hilsenroth, M. J., Fowler, J. C., Padawer, J. R., & Handler, L. (1997). Narcissism in the Rorschach revisited: Some reflections on empirical data. *Psychological Assessment, 9,* 113–121.

Hilsenroth, M. J., Handler, L., & Blais, M. A. (1996). Assessment of narcissistic personality disorder: A multi-method review. *Clinical Psychology Review, 16,* 655–683.

Kaye, A. L., & Shea, M. T. (2000). Personality disorders, personality traits, and defense mechanisms measures. In *Handbook of psychiatric measures* (pp. 713–749). Washington, DC: American Psychiatric Association.

Klein, D. N. (2003). Patients' versus informants' reports of personality disorders in predicting 7.5-year outcome in outpatients with depressive disorders. *Psychological Assessment, 15,* 216–222.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9,* 300–311.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566–1577.

Lampel, A. K. (1999). Use of the Millon Clinical Multiaxial Inventory-III in evaluating child custody litigants. *American Journal of Forensic Psychology, 17,* 19–31.

Lenzenweger, M. F. (1999). Stability and change in personality disorder features. *Archives of General Psychiatry, 56,* 1009–1015.

Lenzenweger, M. F., Loranger, A. W., Korfine, L., & Neff, C. (1997). Detecting personality disorders in a nonclinical population. Application of a 2-stage procedure for case identification. *Archives of General Psychiatry, 54,* 345–351.

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

Lindsay, K. A., Sankis, L. M., & Widiger, T. A. (2000). Gender bias in self-report personality disorder inventories. *Journal of Personality Disorders, 14,* 218–232.

Livesley, W. J., & Jackson, D. (in press). *Manual for the Dimensional Assessment of Personality Pathology-Basic Questionnaire.* Port Huron, MI: Sigma Press.

Loranger, A. W. (1992). Are current self-report and interview methods adequate for epidemiological studies of personality disorders? *Journal of Personality Disorders, 6,* 313–325.

Loranger, A. W. (1999). *International Personality Disorder Examination (IPDE).* Odessa, FL: Psychological Assessment Resources.

Lynam, D. R., & Widiger, T. A. (2001). Using the five factor model to represent the *DSM–IV* personality disorders: An expert consensus approach. *Journal of Abnormal Psychology, 110,* 401–412.

McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa, P. T. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality, 66,* 285–313.

Mellsop, G., Varghese, F. T. N., Joshua, S., & Hicks, A. (1982). The reliability of Axis II of *DSM–III.* *American Journal of Psychiatry, 139,* 1360–1361.

Millon, T., Millon, C., & Davis, R. (1997). *MCMI-III manual* (2nd ed.). Minneapolis, MN: National Computer Systems.

Morey, L. C., Gunderson, J. G., Quigley, B. D., Shea, M. T., Skodol, A. E., McGlashan, T. H., et al. (2002). The representation of borderline, avoidant, obsessive-compulsive, and schizotypal personality disorders by the five-factor model. *Journal of Personality Disorders, 16,* 215–234.

Morey, L. C., Warner, M. B., & Boggs, C. D. (2002). Gender bias in the personality disorders criteria: An investigation of five bias indicators. *Journal of Psychopathology and Behavioral Assessment, 24,* 55–65.

O'Boyle, M., & Self, D. (1990). A comparison of two interviews for *DSM–III–R* personality disorders. *Psychiatry Research, 32,* 85–92.

Okazaki, S., Kallivayalil, D., & Sue, S. (2002). Clinical personality assessment with Asian Americans. In J. T. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (2nd ed., pp. 135–153). London: Oxford University Press.

Oldham, J. M., Skodol, A. E., Kellman, H. D., Hyler, S. E., Rosnick, L., & Davies, M. (1992). Diagnosis of *DSM–III–R* personality disorders by two semi-structured interviews: Patterns of comorbidity. *American Journal of Psychiatry, 149,* 213–220.

Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry, 149,* 1645–1653.

Perry, J. C., Banon, E., & Ianni, F. (1999). Effectiveness of psychotherapy for personality disorders. *American Journal of Psychiatry, 156,* 1312–1321.

Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured interview for DSM–IV personality.* Washington, DC: American Psychiatric Press.

Pilkonis, P. A., Heape, C. L., Proietti, J. M., Clark, S. W., McDavid, J. D., & Pitts, T. E. (1995). The reliability and validity of two structured diagnostic interviews for personality disorders. *Archives of General Psychiatry, 52,* 1025–1033.

Ready, R. E., & Clark, L. A. (2002). Correspondence of psychiatric patient and informant ratings of personality traits, temperament, and interpersonal problems. *Psychological Assessment, 14,* 39–49.

Ready, R. E., Watson, D., & Clark, L. A. (2002). Psychiatric patient- and informant-reported personality: Predicting concurrent and future behavior. *Assessment, 9,* 361–372.

Rogers, R. (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Rogers, R. (2001). *Diagnostic and structured interviewing. A handbook for psychologists.* New York: Guilford Press.

Rogers, R. (2003). Standardizing *DSM–IV* diagnoses: The clinical applications of structured interviews. *Journal of Personality Assessment, 81,* 220–225.

Rounsaville, B. J., Alarcon, R. D., Andrews, G., Jackson, J. S., Kendell, R. E., & Kendler, K. (2002). Basic nomenclature issues for *DSM–V.* In D. J. Kupfer, M. B. First, & D. E. Regier (Eds.), A

research agenda for *DSM–V* (pp. 1–29). Washington, DC: American Psychiatric Press.

Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3,* 203–215.

Sanislow, C. A., & McGlashan, T. H. (1998). Treatment outcome of personality disorders. *Canadian Journal of Psychiatry, 43,* 237–250.

Segal, D. L., & Coolidge, F. L. (2003). Structured interviewing and DSM classification. In M. Hersen & S. Turner (Eds.), *Adult psychopathology and diagnosis* (4th ed., pp. 72–103). New York: Wiley.

Shea, M. T. (1997). Core battery conference: Assessment of change in personality disorders. In H. H. Strupp, L. M. Horowitz, & M. J. Lambert (Eds.), *Measuring patient changes in mood, anxiety, and personality disorders. Toward a core battery* (pp. 389–400). Washington, DC: American Psychological Association.

Shedler, J. (2002). A new language for psychoanalytic diagnosis. *Journal of the American Psychoanalytic Association, 50,* 429–456.

Skodol, A. E., Oldham, J. M., Rosnick, L., Kellman, H. D., & Hyler, S. E. (1991). Diagnosis of *DSM–III–R* personality disorders: A comparison of two structured interviews. *International Journal of Methods in Psychiatric Research, 1,* 13–26.

Spitzer, R. L., Forman, J. B. W., & Nee, J. (1979). *DSM–III* field trials: I. Initial interrater diagnostic reliability. *American Journal of Psychiatry, 136,* 815–817.

Sprock, J., Crosby, J. P., & Nielsen, B. A. (2001). Effects of sex and sex roles on the perceived maladaptiveness of *DSM-IV* personality disorder symptoms. *Journal of Personality Disorders, 15,* 41–59.

Triebwasser, J., & Shea, M. T. (1996). Personality change resulting from another mental disorder. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, & W. W. Davis (Eds.), *DSM–IV sourcebook* (Vol. 2, pp. 861–868). Washington, DC: American Psychiatric Association.

Trull, T. J. (2001). Structural relations between borderline personality disorder features and putative etiological correlates. *Journal of Abnormal Psychology, 110,* 471–481.

Trull, T. J., Widiger, T. A., Useda, J. D., Holcomb, J., Doan, D.-T., Axelrod, S. R., et al. (1998). A structured interview for the assessment of the five-factor model of personality. *Psychological Assessment, 10,* 229–240.

Vitale, J. E., & Newman, J. P. (2001). Using the Psychopathy Checklist-Revised with female samples: Reliability, validity, and implications for clinical utility. *Clinical Psychology: Science and Practice, 8,* 117–132.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26,* 54–60.

Westen, D. (1997). Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *American Journal of Psychiatry, 154,* 895–903.

Westen, D., & Muderrisoglu, S. (2003). Assessing personality disorders using a systematic clinical interview: Evaluation of an alternative to structured interviews. *Journal of Personality Disorders, 4,* 351–369.

Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry, 156,* 258–272.

Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part II: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry, 156,* 273–285.

Westen, D., Shedler, J., Durrett, C., Glass, S., & Martens, A. (2003). Personality diagnoses in adolescence: *DSM–IV* Axis II diagnoses and an empirically derived alternative. *American Journal of Psychiatry, 160,* 952–966.

Whaley, A. L. (1997). Ethnicity/race, paranoia, and psychiatric diagnoses: Clinician bias versus sociocultural differences. *Journal of Psychopathology and Behavioral Assessment, 19,* 1–20.

Widiger, T. A. (1998). Sex biases in the diagnosis of personality disorders. *Journal of Personality Disorders, 12,* 95–118.

Widiger, T. A. (2002). Personality disorders. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment, treatment planning, and outcome for psychological disorders* (pp. 453–480). New York: Guilford Press.

Widiger, T. A. (2003). Personality disorder and Axis I psychopathology: The problematic boundary of Axis I and Axis II. *Journal of Personality Disorders, 17,* 90–108.

Widiger, T. A., & Bornstein, R. F. (2001). Histrionic, narcissistic, and dependent personality disorders. In H. E. Adams & P. Sutker (Eds.), *Comprehensive handbook of psychopathology* (3rd ed., pp. 507–529). New York: Plenum Press.

Widiger, T. A., & Clark, L. A. (2000). Toward *DSM–V* and the classification of psychopathology. *Psychological Bulletin, 126,* 946–963.

Widiger, T. A., & Coker, L. A. (2002). Assessing personality disorders. In J. N. Butcher (Ed.), *Clinical personality assessment. Practical approaches* (2nd ed., pp. 407–434). New York: Oxford University Press.

Widiger, T. A., Mangine, S., Corbitt, E. M., Ellis, C. G., & Thomas, G. V. (1995). *Personality Disorder Interview-IV. A semistructured interview for the assessment of personality disorders. Professional manual.* Odessa, FL: Psychological Assessment Resources.

Wiggins, J. S. (2003). *Paradigms of personality assessment.* New York: Guilford Press.

Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology, 53,* 519–543.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines.* Geneva, Switzerland: Author.

Zanarini, M. C., Frankenburg, F. R., Chauncey, D. L., & Gunderson, J. G. (1987). The Diagnostic Interview for Personality Disorders: Interrater and test-retest reliability. *Comprehensive Psychiatry, 28,* 467–480.

Zanarini, M. C., Frankenburg, F. R., & Vujanovic, A. A. (2002). Inter-rater and test-retest reliability of the Revised Interview for Borderlines. *Journal of Personality Disorders, 16,* 270–276.

Zimmerman, M. (2003). What should the standard of care for psychiatric diagnostic evaluations be? *Journal of Nervous and Mental Disease, 191,* 281–286.

Zimmerman, M., & Mattia, J. I. (1999). Differences between clinical and research practices in diagnosing borderline personality disorder. *American Journal of Psychiatry, 156,* 1570–1574.

Zlotnick, C., Rothschild, L., & Zimmerman, M. (2002). The role of gender in the clinical presentation of patients with borderline personality disorder. *Journal of Personality Disorders, 16,* 277–282.