

Prototype Personality Diagnosis in Clinical Practice: A Viable Alternative for *DSM–5* and *ICD–11*

Drew Westen and Jared A. DeFife
Emory University

Bekh Bradley
Emory University and Atlanta Veterans Administration
Hospital, Atlanta, GA

Mark J. Hilsenroth
Adelphi University

Several studies suggest that a prototype-matching approach yields diagnoses of comparable validity to the more complex diagnostic algorithms outlined in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Furthermore, clinicians prefer prototype diagnosis of personality disorders to the current categorical diagnostic system or alternative dimensional methods. An important extension of this work was to investigate the degree to which clinicians are able to make prototype diagnoses reliably. The aim of this study was to assess the interrater reliability of a prototype-matching approach to personality diagnosis in clinical practice. Using prototypes derived empirically in prior research, outpatient clinicians diagnosed patients' personality after an initial evaluation period. External evaluators independently diagnosed the same patients after watching videotapes of the same clinical hours. Interrater reliability for prototype diagnosis was high, with a median $r = .72$. Cross-correlations between disorders were low, with a median $r = .01$. Clinicians and clinically trained independent observers can assess complex personality constellations with high reliability using a simple prototype-matching procedure, even with prototypes that are relatively unfamiliar to them. In light of its demonstrated reliability, efficiency, and versatility, prototype diagnosis appears to be a viable system for the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* and the 11th edition of the International Classification of Diseases, with exceptional utility for research and clinical practice.

Keywords: prototype diagnosis, personality disorders, prototype reliability, *DSM–5*, *ICD–11*

Despite considerable efforts toward developing and refining diagnostic categories and criteria for the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) and International Classification of Diseases (*ICD*) diagnostic systems, very little research has focused on how best to implement those criteria for making diagnoses in clinical practice. The architects of *DSM–III* aban-

doned the approach taken in *DSM–I* and *–II* (descriptive paragraphs defining disorders, which clinicians diagnosed as present or absent), which lacked empirically derived diagnostic criteria, reliability across clinicians and sites, and formal decision rules for applying the diagnostic categories to individual patients. After it became clear that most diagnoses are not “classical” categories, in

Editor's Note. This article is one of six articles in this special section of *Professional Psychology: Research and Practice* on “Improving Clinical Utility in the Classification of Mental and Behavioral Disorders: Revisions to the *ICD* and the *DSM*.”—MCR

DREW WESTEN is a clinical, personality, and political psychologist and neuroscientist, and professor in the Departments of Psychology and Psychiatry at Emory University. He completed his PhD in clinical psychology at the University of Michigan and formerly taught at the University of Michigan, Harvard Medical School, and Boston University. His areas of specialization are in the field of personality disorder diagnosis and treatment, psychopathology, and political decision making.

JARED A. DEFIFE is a clinical psychology research scientist at Emory University and associate director of the Laboratory of Personality and Psychopathology. He earned his master's and PhD degrees in clinical psychology at Adelphi University and was a clinical fellow at Harvard Medical School. He specializes in the study of personality, mood disorders, and psychotherapy.

BEKH BRADLEY received his PhD in clinical-community psychology from

the University of South Carolina. He is an assistant professor in Department of Psychiatry and Behavioral Sciences; Emory University. He is also the Director of the PTSD treatment team at the Atlanta VAMC. His areas of research and practice include childhood abuse and neglect, PTSD, and risk/resilience following trauma exposure.

MARK J. HILSENROTH received his PhD in clinical psychology from the University of Tennessee and a Diplomate from the American Board of Assessment Psychology. He is currently a professor of psychology at the Derner Institute of Advanced Psychological Studies, Adelphi University. His areas of research interest are personality assessment, training/supervision, psychotherapy process, and treatment outcomes.

PREPARATION OF THIS ARTICLE was supported in part by National Institute of Mental Health Grants MH62377, MH62378, and MH078100. Jonathan Shedler and Drew Westen are the copyright holders of the Shedler-Westen Assessment Procedure (SWAP-II).

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Drew Westen, Emory University, 36 Eagle Row, Atlanta, GA 30322. E-mail: dwesten@emory.edu

which category membership requires all members of a category (in this case, patients) to share a fixed set of defining features, the architects of subsequent editions of the manual switched to the familiar “polythetic” criterion approach, in which a patient can receive a diagnosis by crossing a threshold of features (e.g., five of nine for borderline personality disorder) that are neither necessary nor sufficient for diagnosis (except for some Axis I disorders, such as posttraumatic stress disorder, which require certain features to be present before considering other criteria).

The diagnostic procedures used since *DSM-III* have improved research diagnosis (using structured interviews) and made possible the explosion of research since 1980. However, they remain mostly untested against alternative approaches, particularly in clinical practice, and their problems have gradually become clear (for a summary, see Ortigo, Bradley, & Westen, 2010; Westen, Heim, Morrison, Patterson, & Campbell, 2002). For example, diagnostic overlap produces spuriously high estimates of comorbidity (with most patients who receive one personality disorder [PD] diagnosis by structured interview receiving as many as four to six), and for PDs, as with most other disorders, more patients receive not-otherwise-specified (NOS) diagnoses. A lack of diagnostic specificity hinders both research and practice. The problem of comorbidity is related to the proliferation of NOS and new diagnoses because every time researchers place parameters on a category (e.g., number of criteria required for diagnosis), subthreshold or otherwise not-quite-present variants are identified. Equally problematic, the shift to clear diagnostic criteria and cutoffs since *DSM-III* has not improved interrater agreement (reliability) in clinical practice or field trials (Zimmerman, 1994), in part because disorders defined as lists of distinct criteria are difficult to remember, require dichotomous (and unreliable) judgments about each criterion, are cumbersome in clinical practice (and hence tend not to be used), and have not proven useful to practitioners, for whom the question of whether a patient meets four or five criteria for borderline personality disorder is not particularly relevant (Rottman, Ahn, Sanislow, & Kim, 2009). Furthermore, and related to all of these problems, is the now overwhelming evidence that most disorders are distributed continuously rather than categorically in nature, suggesting the importance of considering dimensional approaches for diagnosis for both Axis I and Axis II disorders (e.g., Brown, Chorpita, & Barlow, 1998; Krueger et al., 2002; Widiger & Clark, 1999). Indeed, dimensional diagnosis has been targeted as one of the major research priorities for *DSM-5*, starting with Axis II (see Kupfer, First, & Regier, 2002; Rounsaville et al., 2002), and the use of some form of dimensional system appears almost certain (Skodol & Bender, 2009).

An important question, however, is how to implement dimensional diagnosis. One possibility that has become the norm in PD research is simply to sum the number of diagnostic criteria met for each disorder. The advantage of dimensionalizing current criteria is continuity with the current diagnostic approach. The disadvantage is that clinicians find *DSM* diagnosis cumbersome already (e.g., Jampala, Sierles, & Taylor, 1988). Expecting them to count criteria across dozens of dimensions is thus unrealistic. In fact, a growing body of literature suggests that clinicians find both categorical diagnosis and a dimensionalized symptom-counting approach to be clinically unworkable across several dimensions of clinical utility (Rottman et al., 2009; Spitzer, Shedler, Westen, & Skodol, 2008; Westen, Shedler, & Bradley, 2006).

Elsewhere we have proposed a prototype-matching approach for diagnosis designed to maximize diagnostic accuracy while taking into consideration the cognitive characteristics of human clinicians (Westen & Bradley, 2005; Westen et al., 2002; Westen & Shedler, 2000). Using this procedure, clinicians rate the overall similarity or “match” between a patient and the prototype using a 5-point scale, taking the prototype *as a whole* rather than counting individual symptoms (see Figure 1). Prototypes consist of paragraph-long descriptions of each disorder rather than lists of circumscribed criteria. This format permits inclusion of more and richer diagnostic criteria and allows for organization of criteria in ways that facilitate memory. Rather than memorize symptom lists with arbitrary and variable cutoffs across disorders, diagnosticians can form mental representations of coherent syndromes in which signs and symptoms may be linked by meaningful functional relations (Ahn, 1999).

Prototype diagnosis has several advantages. For example, it generates both categorical and dimensional diagnoses, overcoming a significant limitation of many forms of dimensional diagnosis (Rounsaville et al., 2002): For purposes of communication, ratings of 4 or 5 denote a categorical diagnosis (“caseness”), and a rating of 3 translates to “features” or subthreshold pathology. The method parallels diagnosis in many areas of medicine, where variables such as blood pressure are measured on a continuum but physicians refer to certain ranges as “borderline” or “high.” In addition, a prototype-matching method more closely resembles how the brain actually works. Cognitive research science on classification processes indicates that human thinking naturally relies on forms of cognitive prototype matching (Cantor & Genero, 1986; Horowitz, Post, de Sales French, Wallis, & Siegelman, 1981; Horowitz, Wright, Lowenstein, & Parad, 1981; Kim & Ahn, 2002).

In a series of recently completed studies of personality, mood, anxiety, eating, and adolescent diagnoses, prototype diagnoses correlated highly with, and had similar correlates to, both categorical and dimensional diagnoses obtained by summing *DSM-IV* criteria or using self-report measures of specific syndromes (Ortigo et al., 2010; Westen et al., 2006). With respect to PDs, across several studies from three different research teams, clinicians who applied different diagnostic approaches to a real patient in their care rated prototype diagnosis substantially more useful, comprehensive, and clinically efficient than *DSM-IV* diagnosis and various other dimensional alternatives (Rottman et al., 2009; Spitzer et al., 2008; Westen et al., 2006).

An important question, however, is whether clinicians can make prototype diagnoses reliably, particularly in everyday practice, where reliability of diagnosis remains poor. The present study was designed to address this question.

Method

Participants

Participants ($N = 65$) were nonpsychotic patients seeking outpatient treatment from a community-based clinic (Hilsenroth, 2007). Patients were 80.0% female, with a mean age of 29.4 years ($SD = 11.6$) and a Global Assessment of Functioning score of 59.3 ($SD = 5.4$); 75.4% were single, with the remainder married, divorced, or widowed. All had at least one Axis I diagnosis ($M = 1.69$), the most common being mood (44.6%), anxiety (23.1%),

Antisocial-psychopathic personality disorder

Patients who match this prototype tend to be deceitful, to lie and mislead people. They take advantage of others, have minimal investment in moral values, and appear to experience no remorse for harm or injury caused to others. They tend to manipulate others' emotions to get what they want; to be unconcerned with the consequences of their actions, appearing to feel immune or invulnerable; and to show reckless disregard for the rights, property, or safety of others. They have little empathy, and seem unable to understand or respond to others' needs and feelings unless they coincide with their own. Individuals who match this prototype tend to act impulsively, without regard for consequences; to be unreliable and irresponsible (e.g., failing to meet work obligations or honor financial commitments); to engage in unlawful or criminal behavior; and to abuse alcohol. They tend to be angry or hostile; to get into power struggles; and to gain pleasure or satisfaction by being sadistic or aggressive toward others. Patients who match this prototype tend to blame others for their own failures or shortcomings, and to believe their problems are caused by external factors. They have little psychological insight into their own motives, behavior, etc. They may repeatedly convince others of their commitment to change but then revert to previous maladaptive behavior, often convincing others that "this time is really different."

1	little or no match (description does not apply)	
2	some match (patient has <i>some features</i> of this disorder)	
3	moderate match (patient has <i>significant features</i> of this disorder)	Features
4	good match (patient <i>has</i> this disorder; diagnosis applies)	Diagnosis
5	very good match (patient <i>exemplifies</i> this disorder; prototypical case)	

Figure 1. Prototype diagnosis of antisocial-psychopathic personality disorder.

and adjustment (10.8%) disorders. The majority of patients had a PD diagnosis (63.1%), approximately half with Cluster B and half with either Cluster A or C diagnoses. After complete description of the study to the participants, written and informed consent was obtained.

The clinician-raters who conducted the psychological assessment, feedback sessions, and prototype ratings were advanced doctoral students enrolled in an accredited clinical psychology PhD program. Each clinician-rater received a minimum of 3.5 hr of supervision per week (1.5 hr individual, 2 hr group) by a licensed clinical psychologist on the therapeutic assessment model/process, scoring/interpretation of assessment measures, presentation/organization of collaborative feedback, clinical interventions (for the clinicians who conducted the assessment procedure, who in all cases were the patients' psychotherapists), and review of videotaped case material.

Procedure

Each patient was assigned to a member of a psychotherapy treatment team in an ecologically valid manner based on clinician availability and caseload. Patient evaluations were conducted using standard clinical interviewing methods traditionally used in private practice but included three meetings totaling approximately

4.5 hr (as well as one independent patient appointment to complete a battery of self-report measures). The procedures were standardized more than is typically the case in clinical practice given that this is both a training clinic and a research clinic focusing on naturalistic psychotherapy research. The assessment procedure was videotaped and included both systematic clinical interviewing about the patient's life history and symptoms (Westen & Muderisoglu, 2003, 2006) as well as collaborative feedback (Finn & Tonsager, 1992, 1997; Fischer, 1994). Further details of the measures, methodology, and procedures used in this assessment process have been described more fully elsewhere (Hilsenroth, 2007; Peters, Hilsenroth, Eudell-Simmons, Blagys, & Handler, 2006).

For the present study, treating clinicians diagnosed each patient's personality pathology at the end of this assessment procedure. External raters consisted of the same pool of clinicians and in some cases the study supervisor (none provided video-ratings for their own patients).

Measures

Clinicians and independent evaluators diagnosed the patient using a version of the PD prototype rating system depicted in Figure 1. Prototypes were empirically derived from data provided by a large national sample of experienced clinicians who used a

200-item Q-sort instrument (Shedler-Westen Assessment Procedure; SWAP-200) to describe a specific PD patient in their care, by applying a statistical procedure (Q factor analysis) to the data to identify empirically distinct diagnostic groupings (Westen & Shedler, 1999). The procedure generated seven primary diagnoses: dysphoric, antisocial, schizoid, paranoid, histrionic, obsessional, and narcissistic. The dysphoric diagnosis included five subtypes: avoidant, high-functioning depressive, emotionally dysregulated (borderline), and hostile-oppositional (a variant of passive-aggressive PD).

We constructed paragraph-long prototypes of each diagnosis by weaving together the items (criteria) most empirically descriptive of each into paragraph form (Westen et al., 2006), grouping functionally or thematically similar items for ease of clinical use (see Figure 1). The advantage of using these empirically derived diagnostic prototypes for the present purposes is that they were designed to be nonredundant; hence, poor discriminability among the prototypes would not be attributable to comorbidity inherent in the current Axis II criterion sets.

Results

Table 1 presents the correlations between clinician and external evaluator prototype ratings for the primary empirically derived disorders, with the exception of dysphoric (whose five subtypes were rated separately as prototypes). Table 2 presents correlations for the five dysphoric subtypes. (The rating system at the time of this study had 7 scale points rather than 5; however, for the present study we collapsed the data to 5 scale points for continuity with other research. The findings were essentially identical when we analyzed them using all 7 scale points.)

As can be seen from the two tables, clinicians were able to make highly reliable and discriminating judgments, with median interrater reliabilities of $r = .72$ and $.74$ for the primary PDs and dysphoria subtypes, respectively. Median correlations off the diagonal, which represent correlations between disorders designed to be relatively low in overlap, were $.17$ and $.11$, respectively. In other words, two independent clinicians tended to see patients much the same way, agreeing on the extent to which they matched the same prototypes and producing low correlations between unrelated diagnoses.

Discussion

The data provide strong evidence for the interrater reliability of prototype diagnosis of PDs in clinical practice. Whereas field trials

and inter-interview studies comparing diagnoses made by different structured interviews and questionnaires for the same PDs administered days or a few weeks apart have shown low correlations and even lower kappa coefficients indicating convergent diagnoses (Clark, Livesley, & Morey, 1997; Pilkonis et al., 1995; Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991), we found high correlations in the range of $r = .70$ between two assessments made by independent assessors from naturalistic clinical hours. Prototype diagnoses also demonstrated strong differentiation across disorders (what might be called discriminant interrater reliability), something not seen in previous research. Although elsewhere we have discussed whether prototype matching might be useful for research diagnosis or for diagnosis of Axis I disorders as well (Westen & Bradley, 2005; Westen et al., 2002), what the data here suggest is that prototype diagnosis offers a promising alternative method for personality diagnosis in clinical practice.

The major limitation of the study is that the sample was relatively small and the clinicians were relatively inexperienced and drawn from the same clinical training pool. Two considerations, however, mitigate these limitations. First, the limitations would favor null findings. For example, inexperienced clinicians would likely have more difficulty using diagnoses other than the more familiar *DSM-IV* categories as well as the *DSM-IV* diagnostic approach, and their limited clinical experience would render them less likely to converge on diagnostic impressions following an interviewing procedure that is relatively open-ended, focusing on the patient's life history as a way of exploring ongoing and enduring personality dynamics. Second, the effects were large and significant, even with this sample size, and the small magnitude of the correlations off the diagonal (i.e., between unrelated or minimally related diagnoses) relative to those on the diagonal (showing diagnostic agreement) clearly demonstrated that even inexperienced clinicians could make highly specific diagnostic judgments when evaluating the same patient using the kinds of data experienced clinicians collect over the course of initial interviews in clinical practice (which the assessment procedure was intended to simulate and standardize).

A second limitation is that clinicians were rating empirically derived diagnoses rather than prototypes of the current Axis II disorders. Three considerations, however, limit this concern. First, as with the first limitation, this would also reduce interclinician agreement, rendering the findings more conservative, given that clinicians were matching patients to diagnoses with which they were unfamiliar. Second, prior research has found that the four

Table 1
Interrater Reliability for Empirically Derived Personality Disorder Diagnoses (N = 65)

Clinician rating	External evaluator rating					
	Antisocial	Paranoid	Schizoid	Obsessional	Histrionic	Narcissistic
Antisocial	.74⁺	.39 ⁺	.22	.05	.10	.28*
Paranoid	.35**	.77⁺	.42 ⁺	.22	.20	.19
Schizoid	.08	.40 ⁺	.76⁺	.44 ⁺	-.28*	.07
Obsessive	-.11	.20	.31*	.63⁺	-.15	.08
Histrionic	.27*	.11	-.25*	-.04	.72⁺	.33**
Narcissistic	.39**	.19	.17	.01	.36**	.56⁺

Note. Correlations in bold type along the hypotheses represent (convergent) reliability coefficients.

⁺ Significant at the .001 level. * Significant at the .05 level. ** Significant at the .01 level.

Table 2
Interrater Reliability for Empirically Derived Dysphoric Disorder Subtype Diagnoses (N = 65)

Clinician rating	External evaluator rating				
	Avoidant	High-functioning depressive	Emotionally dysregulated	Dependent	Hostile-oppositional
Avoidant	.78⁺	.31*	-.01	.13	.27*
High-functioning depressive	.38**	.70⁺	.05	.05	.05
Emotionally dysregulated	.11	-.18	.71⁺	.37**	.43 ⁺
Dependent	-.06	-.21	.25*	.77⁺	-.04
Hostile-oppositional	.18	-.20	.34**	.02	.73⁺

Note. Correlations in bold type along the hypotheses represent (convergent) reliability coefficients.

⁺ Significant at the .001 level. * Significant at the .05 level. ** Significant at the .01 level.

empirically derived disorders tested here that resemble Cluster B disorders have similar correlates as the *DSM-IV* Cluster B PDs and hence are likely to be a reasonable proxy for them.

Third, although *DSM-5* appears increasingly likely to include prototype matching as an approach to dimensionalizing PD diagnosis, it is unlikely to use the current diagnoses precisely as they are configured now, given their many limitations, such as comorbidity (Skodol & Bender, 2009). Thus, the approach tested here, with rich, empirically derived prototypes, is at least as likely as a version of the current diagnoses woven into prototype form to resemble prototype diagnosis in *DSM-5*. Indeed, the prototypes tested here provide a model of the kind of prototypes that might be useful in *DSM-5*, in that they are empirically derived rather than clinically constructed by committee. Spitzer and colleagues (2008) used these empirically derived PD prototypes in their comparative study of clinical utility and found that clinicians preferred even these unfamiliar prototypes over prototypes derived from the current Axis II disorders because of their clinical richness of description. Westen and colleagues (2006) compared the four prototypes most comparable to the *DSM-IV* Cluster B disorders (antisocial-psychopathic, emotionally dysregulated, histrionic, and narcissistic) as well as prototypes of the four Cluster B disorders to *DSM-IV* diagnoses and found similar results. One of the advantages of prototype diagnosis is that it can also include more, and clinically richer, criteria than the eight to nine criteria per disorder included in the current diagnostic system, because clinicians do not have to make independent judgments on each criterion; rather, they make a single prototypicality judgment on each diagnosis taken as a gestalt.

Finally, a major question often raised about a prototype-matching approach to diagnosis is whether it is a “throwback” to *DSM-II* (i.e., a return to paragraph-length diagnoses). That question, however, misses the point. Prototype diagnosis has the parsimony of *DSM-II* diagnosis but lacks its disadvantages. Although the format of the diagnostic prototypes may superficially resemble the format of the diagnostic paragraphs in the first two editions of the *DSM*, this approach to diagnosis differs from the early diagnostic manuals in several key respects: (a) The diagnostic criteria (and in this case, the diagnoses themselves) are entirely empirically derived, not rationally or clinically derived, as in *DSM-I* and *-II*; (b) the diagnoses are not laden with causal clinical hypotheses of the 1930s and 1940s; and (c) most important, clinicians are not making idiosyncratic dichotomous characterizations of patients as either having or not having a disorder, which would

likely be as unreliable as dichotomous judgments about prototypes. Rather, clinicians are taking into account all available data and making a judgment of *the extent to which* the patient matches an empirically derived prototype.

Clinicians are reluctant to implement the existing Axis II diagnostic system with its laundry list of symptoms, cumbersome algorithms, overlapping criteria, and descriptive vagaries. Prototype matching, on the other hand, allows for rich descriptions of personality constructs without an exorbitant clinical effort. Using a prototype system, clinicians could briefly and efficiently (within 1 or 2 min) make an Axis II diagnosis, generating a diagnostic profile that indicates for each disorder both the extent to which the patient resembles the prototype and whether the patient matches the prototype strongly enough to receive a categorical diagnosis. Empirically, the results of this study generated extremely high estimates of cross-clinician reliability and lower cross-correlations with unrelated disorders than we have seen in any PD study to date. The prototype diagnostic system used in this study offers clinically rich diagnostic descriptions that are not only reliably observable across clinicians but also highly discriminative. Narrative diagnostic descriptions allow for improved treatment planning and clinical training while reliable and distinctive diagnoses increase the efficiency of clinical communication and coordination across providers. Indeed, clinicians find prototype diagnosis preferable to alternative approaches across a range of clinical utility variables, including comprehensiveness, ease of implementation, enhancement of treatment planning, and clarity of communication with mental health providers as well as patients (Rottman et al., 2009; Spitzer, First, & Skodol, 2006).

At this point, given the consistent evidence of the validity, clinical utility (First et al., 2004), and now interclinician reliability of prototype diagnosis for PDs, we would recommend that *DSM-5* incorporate prototype matching as the primary method of diagnosing personality constellations, given the likelihood that the Personality and Personality Disorders Work Group appears headed toward maintaining a constellational approach that is likely to be supplemented by other approaches, such as trait diagnosis (Skodol & Bender, 2009). We would also recommend, based on these and other data on Axis I disorders (Ortigo et al., 2010), that the ICD-11 consider prototype diagnosis for all disorders for clinical practice, and that the architects of both the ICD-11 and *DSM-5* undertake research to test whether prototype matching may be a workable approach for clinical practice and research for all clinical disorders, not only PDs.

References

- Ahn, W. (1999). Effect of causal structure on category construction. *Memory & Cognition*, 27, 1008–1023.
- Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the *DSM-IV* anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, 107, 179–192.
- Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM-IV* (pp. 233–256). New York: Guilford Press.
- Clark, L. A., Livesley, W. J., & Morey, L. (1997). Personality disorder assessment: The challenge of construct validity. *Journal of Personality Disorders*, 11, 205–231.
- Finn, S., & Tonsager, M. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy: Providing psychological test feedback to clients. *Psychological Assessment*, 4, 278–287.
- Finn, S. E., & Tonsager, M. E. (1997). Information-gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment*, 9, 374–385.
- First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *The American Journal of Psychiatry*, 161, 946–954.
- Fischer, C. (1994). *Individualized psychological assessment*. Hillsdale, NJ: Erlbaum.
- Hilsenroth, M. (2007). A programmatic study of short-term psychodynamic psychotherapy: Assessment, process, outcome, and training. *Psychotherapy Research*, 17, 31–45.
- Horowitz, L. M., Post, D. L., de Sales French, R., Wallis, K. D., & Siegelman, E. Y. (1981). The prototype as a construct in abnormal psychology: II. Clarifying disagreement in psychiatric judgments. *Journal of Abnormal Psychology*, 90, 575–585.
- Horowitz, L. M., Wright, J. C., Lowenstein, E., & Parad, H. W. (1981). The prototype as a construct in abnormal psychology: I. A method for deriving prototypes. *Journal of Abnormal Psychology*, 90, 568–574.
- Jampala, V., Sierles, F., & Taylor, M. (1988). The use of *DSM-III-R* in the United States: A case of not going by the book. *Comprehensive Psychiatry*, 29, 39–47.
- Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, 131, 451–476.
- Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Lacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, 111, 411–424.
- Kupfer, D. J., First, M. B., & Regier, D. A. (Eds.). (2002). *A research agenda for DSM-V*. Washington, DC: American Psychiatric Association.
- Ortigo, K. M., Bradley, B., & Westen, D. (2010). An empirically based prototype diagnostic system for *DSM-V* and *ICD-11*. In T. Millon, R. Krueger, & E. Simonsen (Eds.), *Contemporary directions in psychopathology: Scientific foundations of the DSM-V and ICD-11* (pp. 374–390). New York: Guilford Press.
- Peters, E., Hilsenroth, M., Eudell-Simmons, E., Blagys, M., & Handler, L. (2006). Reliability and validity of the Social Cognition and Object Relations Scale in clinical use. *Psychotherapy Research*, 16, 606–614.
- 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.
- Rottman, B. M., Ahn, W. K., Sanislow, C. A., & Kim, N. S. (2009). Can clinicians recognize *DSM-IV* personality disorders from five-factor model descriptions of patient cases? *The American Journal of Psychiatry*, 166, 427–433.
- 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. Kupfer, M. B. First, & D. A. Regier (Eds.), *A research agenda for DSM-V* (pp. 1–29). Washington, DC: American Psychiatric Association.
- Skodol, A. E., & Bender, D. S. (2009). The future of personality disorders in *DSM-V*? *The American Journal of Psychiatry*, 166, 388–391.
- Skodol, A. E., Oldham, J. M., Rosnick, L., Kellman, D., & Hyler, S. (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., First, M. B., & Skodol, A. E. (2006). *Clinical acceptability of five dimensional systems for personality diagnosis: A pilot study*. Unpublished manuscript, Department of Psychiatry, Columbia University.
- Spitzer, R., Shedler, J., Westen, D., & Skodol, A. (2008). Clinical utility of five dimensional systems for personality diagnosis: A “consumer preference” study. *The Journal of Nervous and Mental Disease*, 196, 356–374.
- Westen, D., & Bradley, R. (2005). Prototype diagnosis of personality. In S. Strack (Ed.), *Handbook of personology and psychopathology* (pp. 238–256). New York: Wiley.
- Westen, D., Heim, A. K., Morrison, K., Patterson, M., & Campbell, L. (2002). Classifying and diagnosing psychopathology: A prototype matching approach. In L. Beutler & M. Malik (Eds.), *Rethinking the DSM: Psychological perspectives* (pp. 221–250). Washington, DC: American Psychological Association.
- Westen, D., & Muderrisoglu, S. (2003). Reliability and validity of personality disorder assessment using a systematic clinical interview: Evaluating an alternative to structured interviews. *Journal of Personality Disorders*, 17, 350–368.
- Westen, D., & Muderrisoglu, S. (2006). Clinical assessment of pathological personality traits. *The American Journal of Psychiatry*, 163, 1285–1287.
- Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part 2: Toward an empirically based and clinically useful classification of personality disorders. *The American Journal of Psychiatry*, 156, 273–285.
- Westen, D., & Shedler, J. (2000). A prototype matching approach to diagnosing personality disorders toward *DSM-V*. *Journal of Personality Disorders*, 14, 109–126.
- Westen, D., Shedler, J., & Bradley, R. (2006). A prototype approach to personality disorder diagnosis. *The American Journal of Psychiatry*, 163, 846–856.
- Widiger, T. A., & Clark, L. A. (1999). Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin*, 126, 947–963.
- Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives of General Psychiatry*, 51, 225–245.

Received January 19, 2010

Revision received July 26, 2010

Accepted August 10, 2010 ■