

Breaking the Ceiling on Risk Assessment: Dispositional Indicators of Risk Exposure (DIRE) Scale

Jonathan Shedler

Luisa E. Marin-Avellan

Olga G. Shechter

Peter Fonagy

Michael Karson

Eric L. Lang

Abstract

Current risk-assessment methods may be approaching a ceiling on accuracy. The domain of personality represents a source of untapped information for enhancing prediction not only of criminality but also of broadly defined misconduct, including breaches of trust and other forms of non-criminal insider threat in organizations. We describe the Shedler-Westen Assessment Procedure (SWAP), a comprehensive method of personality assessment, and the Dispositional Indicators of Risk Exposure (DIRE) scale, a psychometric scale designed to harness implicit and explicit expert knowledge concerning personality and risk. Study 1 examined the convergent validity of the DIRE scale in a national clinical sample of $N = 1,201$ patients. DIRE correlated significantly with a range of risk-related criterion measures, including global maladaptive functioning ($r = .64$), employment trouble ($r = .49$), mental instability ($r = .34$), criminality and violence ($r = .46$), and child/adolescent antisociality ($r = .53$). Study 2 examined the prospective prediction of criminal recidivism in a sample of violent offenders. DIRE was a significant prospective predictor of criminal recidivism over a 1-year period ($r = .37$). We discuss implications for risk assessment in both general and criminal populations.

Authors' note

We express our appreciation to the late Professor Gill McGauley for her contributions to this paper. As a dedicated psychiatrist in forensic services in the UK, she pioneered forensic psychotherapy through clinical work, service creation, postgraduate teaching, and research.



DR. JONATHAN SHEDLER

Jonathan Shedler, PhD is a Clinical Professor in the Department of Psychiatry and Behavioral Sciences at the University of California, San Francisco (UCSF). He is among the leading experts on personality styles and disorders and their treatment. He is co-author of the Psychodynamic Diagnostic Manual (PDM-2) and co-author of the Shedler-Westen Assessment Procedure (SWAP). He lectures internationally and consults to mental health professionals, organizations, and government agencies.

.....

Risk assessment, broadly defined as the prediction of undesirable outcomes (cf. Kraemer et al., 1997), appears to be approaching a ceiling on accuracy (Skeem & Monahan, 2011, p. 41). Contemporary risk-assessment scales comprise items developed to distinguish criminal recidivists from non-recidivists in correctional populations, with the result that the various scales commonly used for risk assessment are largely interchangeable (Kroner, Mills & Morgan, 2005; Yang, Wong & Coid, 2010). Prior research has highlighted four factors related to criminal violence: criminal history, persistent antisocial lifestyle, psychopathic personality (McWilliams & Shedler, 2017; Meloy, 1988), and substance abuse and/or mental health issues (Kroner, Mills, & Morgan, 2005).

The domain of personality, beyond psychopathy, represents a relatively untapped source of information for risk assessment that could enhance prediction not only concerning criminality, violence, and insider threat (Harms, et. al., 2022) but also with respect



The domain of personality, beyond psychopathy, represents a relatively untapped source of information for risk assessment that could enhance prediction not only concerning criminality, violence, and insider threat (Harms, et. al., 2022) but also with respect to misconduct and more broadly defined undesirable outcomes.

.....



Luisa E. Marin-Avellan, PhD is a psychologist, psychotherapist, and psychoanalyst with a deep-rooted interest in personality. Her 10-year research experience in England's forensic psychiatric services explored the link between personality factors and the risk of violent recidivism. Dr. Marin-Avellan now focuses on her private clinical practice in Geneva, Switzerland.

.....

to misconduct and more broadly defined undesirable outcomes. Such undesirable outcomes may include, for example, insider threats in organizations such as compromise of information systems, failure to protect sensitive information, security breaches, misuse of resources, and breaches of trust, whether or not they involve illegal activity.

Many aspects of personality that are conceptually linked to risk are not represented or represented only minimally in risk assessment item pools. The domain of personality has been partially tapped in risk-assessment instruments described as structured professional judgment (SPJ), such as the Historical-Clinical-Risk Management-20 (HCR-20; Webster, et al., 1997), where a clinician considers an array of historical, clinical, and risk factors to render an overall judgment of risk. However, the personality variables included in the HCR-20 address relatively overt (easily observable) aspects of personality functioning such as personality disorder, impulsivity, negative attitudes, and lack of insight. These concepts stay close to the four factors noted above and do not significantly expand the potential item pool for risk assessment. With respect to personality dynamics, these variables can be said to represent relatively surface-level phenomena.

Conceptually, many personality pathways could lead to high-risk behavior (Buss, 1961; Daffern & Howells, 2002). For example, transgression by individuals with psychopathic personalities may be motivated by a desire for power or personal gain, and transgression by individuals with paranoid personalities may be motivated by a misdirected sense of justice and a desire to turn the



DR. OLGA SHECHTER

Olga Shechter, PhD is a project director at the Defense Personnel and Security Research Center (PERSEREC), which is a division of the Defense Personnel Assessment Center (DPAC). She completed her doctoral education in social and personality psychology from University of Wisconsin, Madison. At PERSEREC, Dr. Shechter manages research projects in the areas of military suicide prevention and postvention.



tables on perceived persecutors (e.g., avenger violence), and transgression by individuals with borderline personality pathology may represent the externalization or “exportation” of internal chaos. Analyzing personality dynamics associated with misconduct not only increases the potential for accurate prediction, but also enhances the ability to take effective countermeasures (Heilbrun, 1997; Nicoletti, Spencer-Thomas, & Bollinger, 1999) based on an accurate understanding of motivation and likely precipitating circumstances.

This article describes the *Shedler-Westen Assessment Procedure* (SWAP), an approach to personality assessment that relies on informed clinical observation and judgment, and a risk-assessment scale derived from it, the *Dispositional Indicators of Risk Exposure* (DIRE) scale. Study 1 examines convergent validity of the DIRE scale with respect to a range of risk-related criterion measures in a large national clinical sample, and Study 2 examines the prospective prediction of criminal recidivism in a psychiatrically disturbed correctional population.

Overview of the Shedler-Westen Assessment Procedure (SWAP-200)

The SWAP is a personality-assessment instrument completed by an expert clinical assessor after developing a thorough knowledge of a patient or assessment subject in a professional evaluative context (the instrument is available online at swapassessment.org). The SWAP provides assessors with a “standard vocabulary” for describing and quantifying clinical observation and inference about personality. The vocabulary



DR. PETER FONAGY

Peter Fonagy, OBE is Professor of Contemporary Psychoanalysis and Developmental Science, Head of Division for Psychology and Language Sciences, University College London (UCL); Chief Executive of the Anna Freud National Centre for Children and Families and Executive Clinical Director, UCL Partners Mental Health and Wellbeing Programme. Dr. Fonagy’s clinical and research interests lie in early attachment relationships, social cognition, borderline personality disorder and violence. A central focus has been a research-based psychodynamic therapeutic approach, mentalization-based treatment.



comprises 200 personality-descriptive statements or items, each of which may describe a given person very well, somewhat, or not at all. An assessor describes a person by ranking the SWAP items into eight categories, from most descriptive of the person (scored 7) to not descriptive or irrelevant (scored 0). Thus, the instrument yields a score from 0 to 7 for 200 personality-descriptive variables. The major editions of the SWAP instrument are the SWAP-200 and revised SWAP-II (Shedler, 2022; Shedler, 2015; Shedler & Westen, 2004a, 2004b, 2007; Westen & Shedler, 1999a, 1999b; Westen, Shedler, Bradley, & DeFife, 2012).

The “standard vocabulary” of the SWAP allows an assessor to provide a comprehensive, in-depth psychological description of a patient or assessment subject in a systematic form. SWAP items stay close to the clinical data (e.g., “Tends to get into power struggles,” or “Is capable of sustaining meaningful relationships characterized by genuine intimacy and caring”) and items that require inference or deduction are written in clear, jargon-free language (e.g., “Tends to express anger in passive and indirect ways [e.g., may make mistakes, procrastinate, forget, become sulky, etc.]” or “Tends to see own unacceptable feelings or impulses in other people instead of in him/herself”). Writing items in jargon-free language minimizes unreliable interpretive leaps by assessors and makes the item set useful to clinicians of all theoretical orientations.

The initial SWAP item pool was drawn from a wide range of sources including the clinical literature on personality pathology written over the past 50 years (e.g., Kernberg, 1975, 1984; Kohut, 1971; Linehan, 1993; McWilliams,



DR. MICHAEL KARSON

Michael Karson, PhD, J.D., A.B.P.P. (Clinical) was on the faculty at the University of Denver for 20 years before returning to clinical practice and consultation. He is the sole or senior author of six books, including *Principles of Forensic Report Writing* and *What Every Therapist Needs to Know*. He has consulted on safe termination practices and personnel selection for numerous organizations.



1994; Shapiro, 1965); DSM Axis II diagnostic criteria included in DSM-III through DSM-IV; selected DSM Axis I criteria that could reflect enduring dispositions (e.g., depression and anxiety); research on coping, defense, and affect regulation (e.g., Perry & Cooper, 1987; Shedler, Mayman, & Manis, 1993; Vaillant, 1992; Westen, Muderrisoglu, Fowler, Shedler, & Koren, 1997); research on interpersonal functioning in patients with personality disorders (Westen, 1991; Westen, Lohr, Silk, Gold, & Kerber, 1990); research on personality traits in non-clinical populations (e.g., Block, 1971; John, 1990; McCrae & Costa, 1990); research on personality pathology conducted since the development of DSM Axis II (see, e.g., Livesley, 1995); pilot studies in which observers watched videotaped interviews of patients with personality disorders and described them using draft versions of the SWAP item set; and the clinical experience of the SWAP authors.

Most important, the SWAP item pool was revised and refined through a 12-year iterative revision process that incorporated the feedback of over 2,000 clinician-consultants of all theoretical orientations who used earlier versions of the SWAP instrument to describe their patients. The instrument developers asked each clinician-consultant one crucial question: “Were you able to describe the things you consider psychologically important about your patient?” They added, rewrote, and revised items based on this feedback, then asked new clinician-consultants to describe new patients, repeating this process over many iterations until most clinicians could answer “yes” most of the time. In a sample of 1,201 psychologists and



DR. ERIC LANG

Eric Lang, PhD is the Director of the Department of Defense Personnel and Security Research Center (PERSEREC). Dr. Lang has over 30 years' experience leading social science research to improve the effectiveness, efficiency and fairness of personnel security, insider threat and suitability policies and operations to help DoD, other Federal agencies, and partners in industry, academe, and over a dozen allied countries.



psychiatrists who used the SWAP-II to describe a current patient, 84% “agreed” or “strongly agreed” with the statement “The SWAP-II allowed me to express the things I consider important about my patient’s personality” (fewer than 5% disagreed). The ratings were unrelated to clinicians’ theoretical orientation (Shedler & Westen, 2007).

The SWAP is based on the Q-Sort method, which requires assessors to assign each score a specified number of times (there is a “fixed” score distribution). The fixed score distribution is asymmetric, with 100 items receiving scores of 0 or “not descriptive” and progressively fewer items receiving higher scores (the shape of the fixed distribution mirrors the naturally occurring distribution in the population; for a discussion of this and other psychometric issues, see Westen & Shedler, 2007). Use of a fixed distribution has psychometric advantages and reduces measurement error or “noise” inherent in standard rating scales.¹ The psychometric rationale for the Q-Sort method has been described in detail by Block (1978).

When the SWAP is used in the context of psychotherapy, an experienced clinician can

“
When the SWAP is used in the context of psychotherapy, an experienced clinician can score the instrument after a minimum of 6 clinical contact hours with a patient.

score the instrument after a minimum of 6 clinical contact hours with a patient. When used in a pure assessment context, as in personnel or forensic evaluation, the SWAP can be scored on the basis of the Clinical Diagnostic Interview (CDI), which systematizes and compresses into an approximately 2.5-hour time frame the kind of interviewing skilled clinicians engage in during the initial hours of patient contact to assess personality (Westen, 2004; Westen & Muderrisoglu, 2006; Westen & Weinberger, 2004). The interview does not rely on self-report questions about personality; rather, it elicits narrative accounts of past and present relationship experiences, which provide a psychologically rich data source from which clinically expert assessors can draw reliable and valid inferences about personality. The SWAP can also be scored reliably and validly from other comparably psychologically rich interview sources (e.g., Marin-Avellan, McGauley, Campbell, & Fonagy, 2005).

Software-based scoring algorithms combine and weight item scores to derive diagnostic scale scores. SWAP-2 00 generates 37 diagnostic scales organized into three score profiles (Shedler, 2009). The three score profiles provide (1) dimensional scores for DSM-5 personality disorder diagnoses, (2) dimensional scores for an alternative set of personality syndromes identified empirically through SWAP research, and (3) dimensional trait scores derived via factor analysis of the SWAP item set. SWAP also generates a global Psychological Health Index, which measures personality strengths or adaptive resources and capacities (e.g., ego strengths). To facilitate score interpretation, all diagnostic scores are reported as T-scores ($M = 50$, $SD = 10$).²

```
particles = [p for p in psys.particles] if pset.type == 'ALIVE']
filenames = []
if pset.render_type == "OBJECT":
    dupli_ob = pset.dupli_object
    if dupli_ob is not None and draw_dat.instances_write_dupli:
        filepath = [bpath.abspath(draw_dat.path), dupli_ob.name]
        if os.path.exists(bpath.abspath(draw_dat.instance_export_path)):
            filepath = [bpath.abspath(draw_dat.instance_export_path), dupli_ob.name]
        filepath = "".join(filepath)
        dupli_world = dupli_ob.matrix_world.copy()
        transl_inv = Matrix.Translation(-dupli_world.translation)
        dupli_ob.matrix_world = transl_inv * dupli_ob.matrix_world
        filenames.extend(writeDupliObjects(scene, [dupli_ob], filepath, temp))
        dupli_ob.matrix_world = dupli_world
        obj.matrix_world = Matrix.Identity(4) * (unsigned long **) lookup->data;
        writeObject(context, instance_filepath, [obj]);
```

1 One way it does so is by ensuring that raters are “calibrated” with one another. Consider the situation with rating scales, where raters can use any value as often as they wish. Inevitably, certain raters will gravitate toward extreme values (e.g., values of 0 and 7 on a 0–7 scale) and others toward middle values (e.g., values of 4 and 5). Thus, the scores reflect not only the personality characteristics of the subjects but also the calibration of the raters. The Q-Sort method, with its fixed distribution, eliminates this kind of measurement error, because all clinicians must assign each score the same number of times. If the use of a standard item set gives clinicians a common vocabulary, use of a fixed distribution can be said to give them a “common grammar” (Block, 1978).

Median inter-rater reliability of SWAP diagnostic scales is above .80 in all studies to date and is often above .90 (Marin-Avellan, McGauley, Campbell, & Fonagy, 2014; Westen & Muderrisoglu, 2003; Westen & Shedler, 2007). Median test-retest reliability of SWAP-II personality disorder diagnostic scales, over a four-to-six-month interval, is .90 (Blagov, Bi, Shedler, & Westen, 2012). With respect to validity, SWAP diagnostic scales show predicted relations with a wide range of criterion measures, including genetic history variables (e.g., psychotic disorders in first- and second-degree biological), developmental history variables (e.g., childhood physical or sexual abuse), adult life events (e.g., arrests, psychiatric hospitalizations, suicide attempts), employment trouble (e.g., job loss due to interpersonal problems in the workplace), social functioning, global adaptive functioning, response to mental health treatment, and numerous other variables (for reviews, see Blagov et al., 2012; Shedler, 2015; Westen & Shedler, 2007).

Overview of the Dispositional Indicators of Risk Exposure (DIRE) Scale

A SWAP scale for risk assessment was constructed using the same method used to construct SWAP-200 scales for DSM personality disorders (Westen & Shedler, 1999a, 1999b). The method involved tapping the explicit and implicit knowledge of expert clinicians by asking them to use the SWAP-200 to describe a hypothetical, prototypical patient representing a specific personality disorder in its “ideal” or pure form (e.g., a prototypical patient with paranoid personality disorder). The resulting SWAP-200 item scores were then averaged across the clinicians to create a *diagnostic prototype* for each personality disorder—a quantified personality description representing experts’ consensus understanding of the disorder. SWAP-200 diagnostic scale scores measure the resemblance or “match” between an assessment subject and the personality disorder diagnostic prototypes, with higher scores indicating greater resemblance to a diagnostic prototype and more severe personality pathology.

We applied this method to develop a risk-assessment scale, called the *Dispositional Indicators of Risk Exposure* (DIRE) scale. We tapped the explicit and implicit knowledge of experts about personality attributes associated with risk by asking them to use the SWAP-200 to describe a hypothetical, prototypical person who poses maximal risk. In this case, the experts were 20 adjudicators from four U.S. government intelligence agencies. Adjudicators make determinations with respect to granting or revoking security



clearances for sensitive positions such as those requiring access to classified information. The 20 adjudicators were asked to describe a hypothetical, prototypical high-risk individual “capable of endangering the safety of others, compromising important systems, or otherwise undermining national security.” We relied on adjudicators rather than clinical psychologists and psychiatrists because of their extensive experience with security risk. Clinicians are experts in diagnosis, but generally have less direct experience addressing impaired judgment, reliability, or trustworthiness in settings where security breach can be catastrophic. All adjudicators were highly experienced and had expertise in personnel security and counterintelligence.

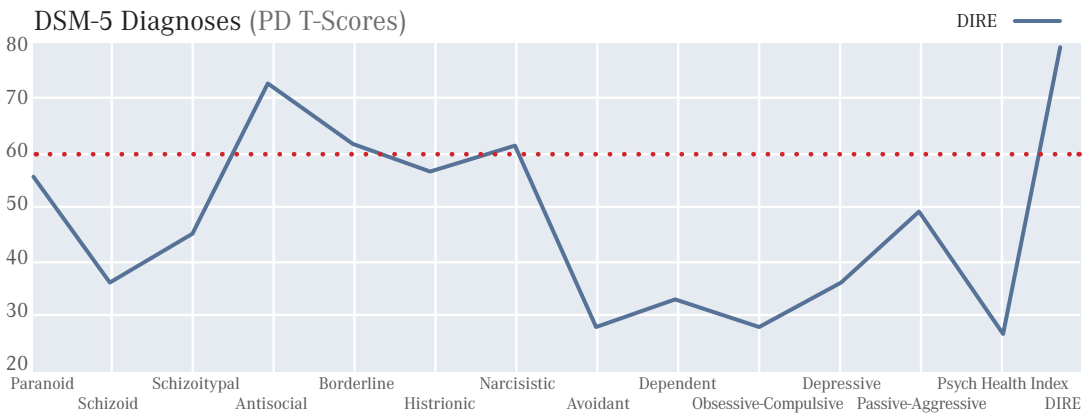
The 20 adjudicators showed high inter-rater reliability in their SWAP-200 descriptions (Cronbach’s alpha = .92, based on intercorrelations among their SWAP-200 descriptions), indicating shared understanding (implicit or explicit) with respect to personality attributes associated with risk. The SWAP-200 descriptions were averaged across the adjudicators to create the DIRE diagnostic prototype representing maximal risk. The DIRE scale measures the degree of resemblance or match between an assessment subject and the DIRE diagnostic prototype, with higher scores indicating greater resemblance and higher risk. DIRE scores are reported as T-scores. As a general interpretive guideline, we have treated DIRE scores of $T \geq 60$ as indicative of unacceptable risk and scores of $T \geq 55$ as danger signs warranting close scrutiny (the cut-points mirror those for personality disorder diagnosis, where $T \geq 60$ warrants a categorical DSM personality disorder diagnosis and $T \geq 55$ warrants a diagnosis of traits or features of a personality disorder). The development and characteristics of the DIRE scale have been described in greater detail elsewhere (Shechter & Lang, 2011).

Examination of the SWAP items weighted heavily in the DIRE diagnostic prototype highlighted three personality syndromes associated with risk (Shechter & Lang, 2011). One syndrome is psychopathy, characterized by lack of an internalized value system, readiness to use and exploit others, deceitfulness, power seeking for its own sake, lack of remorse, sadism, impulsivity, thrill seeking, and externalization of blame (note that these descriptors refer to personality dynamics, not criminality or antisocial behavior). The second syndrome is what has been termed *malignant narcissism* in the clinical literature—a variant of narcissistic personality in which grandiosity, entitlement, and self-importance are suffused with aggression and shade into exploitation (e.g., Kernberg, 1975, 1984). Malignant narcissism is one of three subtypes of narcissistic personality identified empirically in prior research conducted with the SWAP (Russ, Shedler, Bradley, & Westen, 2008). The third syndrome is *borderline personality organization* (Clarkin, Yeomans, & Kernberg, 2006; Kernberg, 1975, 1984; McWilliams, 1994), characterized by affect dysregulation, unstable attachments, and unstable identity.

Clinical theory and experience suggest that these syndromes represent different pathways to risk. Individuals with psychopathic personality styles may transgress for personal gain or for the thrill of manipulating others and getting away with it. Individuals with malignantly narcissistic personality styles may transgress because they do not believe that rules created for lesser beings apply to them, or out of rage and desire for revenge when they feel slighted or devalued. Individuals with borderline personality are unstable and unpredictable (e.g., their attitudes, values, and loyalties are subject to unexpected change). Consequently, the person you are dealing with today may not be the person you are dealing with tomorrow. Additionally, individuals with borderline personality are prone to recreate internal emotional chaos in the external interpersonal world, fueling animosity, discord, and dysfunction in organizational settings (e.g., Clarkin, Yeomans, & Kernberg, 2006; Linehan, 1993).

It is also possible to look at the DIRE prototype through the lens of DSM-IV/DSM-5 personality disorder diagnoses. Recall that the DIRE diagnostic prototype is a SWAP description of a hypothetical person representing maximal risk. Figure 1 shows the SWAP-200 DSM-IV/DSM-5 personality disorder score profile for this hypothetical high-risk person. The score profile indicates how the person would be diagnosed with respect to DSM-IV/DSM-5 personality disorders by a consensus of expert clinicians (who do not limit themselves to DSM diagnostic criteria when making DSM personality disorder diagnoses; see Shedler & Westen, 2004b).

FIGURE 1. DSM-IV/DSM-5 Personality Disorder Score Profile



The recommended cut-point for making a categorical DSM diagnosis is a SWAP-200 scale score of $T \geq 60$ (Shedler & Westen, 2007).³ Figure 1 shows three DSM personality disorder scales with score elevations above this threshold (indicated by the red horizontal line). In DSM terms, the hypothetical, maximally high-risk individual would therefore be diagnosed with antisocial personality disorder, narcissistic personality disorder, and borderline personality disorder. Also noteworthy is the low score of $T=27$ on the Psychological Health Index, which is nearly 2.5 standard deviations below the normative sample mean. Thus, severe personality pathology and deficits in adaptive psychological resources represent risk in their own right, independent of specific personality disorder(s).

The SWAP-200 and DIRE scale underwent initial field testing at U.S. government facilities where clinical psychologists perform psychological evaluations of personnel who require high-level security clearances for positions involving access to sensitive information. The SWAP-200 was added to a rigorous psychological assessment protocol (comprising interviews and a battery of tests including the Minnesota Multiphasic Personality Inventory [MMPI-2-RF; Ben-Porath & Tellegen, 2008] and Personality Assessment Inventory (PAI; Morey, 2009)) to evaluate clinical utility under real-world conditions. Utility was evaluated via structured surveys and debriefing interviews at the conclusion of the field trial. Participating clinicians reported that the SWAP-200 and DIRE scale was more effective than existing tools for assessing personality and for making legally defensible recommendations with respect to risk (Shechter & Lang, 2011).

³ For SWAP-200, a T-score of 50 indicates average functioning in a reference sample of patients with DSM personality disorder diagnoses. A T score of 60 represents an elevation of one standard deviation relative to a reference sample of patients with DSM personality disorder diagnoses.

Study 1: Convergent Validity of DIRE Scale

Overview

This study examines the convergent validity of the DIRE scale by examining correlations between the scale and a range of risk-related criterion measures recorded by participating psychologists and psychiatrists in a large national clinical sample.

Method

We contacted a random national sample of psychiatrists and psychologists with at least 5 years' experience post-training, selected from the membership rosters of the American Psychological Association and American Psychiatric Association, and asked them to use the SWAP-II to describe "an adult patient you are currently treating or evaluating who has enduring patterns of thoughts, feeling, motivation or behavior—that is, personality patterns—that cause distress or dysfunction." To obtain a sample with a broad spectrum of personality functioning, we emphasized that patients need not have a DSM personality disorder diagnosis but did need to meet the following inclusion criteria: > 18 years of age, not currently psychotic, and known well by the clinician (using the guideline of > 6 clinical contact hours but < 2 years). To obtain a random selection of patients from clinicians' practices, we instructed clinicians to consult their calendars to select the last patient they had seen during the previous week who met the study criteria. Each clinician provided informed consent, contributed data describing one patient, and received \$200 in compensation. The sample and data-collection methods have been described in prior publications (Russ et al., 2008; Westen & Shedler, 2007; Westen, Shedler, & Bradley, 2006; Westen et al., 2012).

Dependent Measures

In addition to completing the SWAP-II, each participating clinician completed the Clinical Data Form (CDF), a clinician-report form that gathers extensive data on demographic, diagnostic, etiological, and adaptive functioning variables. CDF life event and developmental history variables show strong agreement (cross-method validity) with independent data collected via patient self-report (DeFife, Drill, Nakash, & Westen, 2010), and adaptive functioning variables assessed via the CDF (e.g., Global Assessment of Functioning [GAF]) show high validity with respect to ratings by independent observers (DeFife et al., 2010; Dutra, Campbell, & Westen, 2004; Westen et al., 1997).

Thirty CDF variables were chosen a priori by experts in personnel security as outcome or criterion variables, because they directly addressed specific undesirable events and outcomes (e.g., violence, criminality, domestic abuse, employment termination due to interpersonal problems, psychiatric hospitalization) or because of their conceptual and empirical link to risk (e.g., childhood and/or adolescent antisociality).

The criterion variables covered a wide spectrum of specific undesirable life events and behaviors as well as indicators of severe mental instability. Multiple measures of adaptive functioning provided a broad-based assessment of psychiatric stability/instability. They included the DSM-IV Global Assessment of Functioning (GAF) scale as well as clinician ratings of chronic level of personality functioning (high-functioning to severe pathology). Other items addressed quality and stability of social relationships and occupational functioning. The CDF variables also included items addressing historical life events of potential relevance to risk assessment (rated dichotomously as “no/unsure” or “yes”). These items addressed psychiatric history (i.e., suicide attempts, self-mutilation, psychiatric hospitalization), criminality and violence (e.g., arrest within the past 5 years, violence in the past 5 years, being a perpetrator in an abusive domestic relationship), or severe interpersonal or occupational problems (e.g., job loss within the past 5 years due to interpersonal conflict in the workplace). Other items addressed childhood and adolescent behaviors and events that are empirically and conceptually linked to psychopathy or antisociality (e.g., fire setting, animal torture, physical fights, stealing, violent/armed crime, running away from home, substance abuse, school trouble, sexual promiscuity).

Results

Sample Characteristics

The sample was $N = 1,201$ patients, 53.2% female, 73.1% seen in private practice settings (with the remainder seen in a range of settings from outpatient clinics to forensic units), 82.7% White (with the remainder Black and/or Hispanic), with a mean age of 42.3 ($SD = 12.3$) years. Patients spanned all social classes. GAF scores spanned a broad range of functioning, ranging from 10 to 93 ($M = 57.9$, $SD = 10.8$). One third of the sample had had at least one psychiatric hospitalization, one fourth had a history of suicide attempt(s), and one in ten had been arrested during the previous 5 years. Clinician respondents were highly experienced ($M = 19.8$ years' practice experience, $SD = 9.2$ years) and diverse in theoretical orientation.

Construction of Composite Outcome Indices

To create reliable criterion measures and facilitate data interpretation, we constructed five composite scales or indices from the CDF variables. Item selection for the composite scales was guided by a principal components analysis of the 30 CDF variables, which yielded five conceptually coherent factors (technically, components), described below. We created a composite index for each factor by averaging the CDF variables with the highest loadings on each factor, after first standard scoring the CDF variables (i.e., transforming them to create score distributions with $M = 0$, $SD = 1$). This procedure ensures equal weighting of the items comprising a composite index. We reversed the direction of scoring of variables as needed so that higher scores always indicated maladaptive behavior or impairment.

- 1. Adult Maladaptive Functioning** provides a global measure of impaired functioning across multiple life domains. The scale comprises Global Assessment of Functioning (GAF) scores, ratings of overall personality functioning, ratings of social and occupational functioning, history of self-mutilation, history of psychiatric hospitalization, being arrested during the previous 5 years, committing a violent crime during the previous 5 years, losing a job during the past 5 years due to interpersonal problems in the workplace, and being the perpetrator in an abusive domestic relationship. Scale reliability (Cronbach's alpha) is $\alpha = .76$.
- 2. Employment Trouble** measures maladaptive functioning in employment settings. The scale comprises two variables—rating of occupational functioning and job loss in the past 5 years due to interpersonal problems in the workplace. Scale reliability is $\alpha = .54$.
- 3. Mental Instability** measures severe mental health problems. The scale comprises history of suicide attempts and history of psychiatric hospitalization. Scale reliability is $\alpha = .71$.
- 4. Forensic Risk/Violence** measures criminality and violence. The scale comprises arrest in the past 5 years, committing a violent crime in the past 5 years, and being the perpetrator in an abusive domestic relationship. Scale reliability is $\alpha = .50$.
- 5. Childhood/Adolescent Psychopathy** measures childhood/adolescent behaviors conceptually and empirically related to the constructs of psychopathy and/or antisociality. The scale comprises childhood/adolescent fire setting, animal torture, running away from home, substance abuse, physical fights, school conduct problems, school performance, lying, stealing, violence, arrests, and age at first intercourse (reverse scored). Scale reliability is $\alpha = .74$.

Convergent Validity

Table 1 lists the correlations between the DIRE scale and the five CDF composite indices. All correlations involving DIRE were in the predicted direction, statistically significant ($p < .001$ for all variables), and moderate to large in magnitude. The highest observed correlation was between DIRE and Adult Maladaptive Functioning, $r = .64$, $p < .001$.

For more fine-grained detail, Table 2 lists the correlations between the DIRE scale and the individual CDF variables that constitute the composite indices. Within the five content domains, CDF variables are listed in descending order by magnitude of correlation. The DIRE scale showed statistically significant relations ($p < .001$ for all variables), in the expected direction, with all 30 criterion variables.

TABLE 1. Correlations of DIRE scale with Clinical Data Form composite indices (N = 1,201)

Composite Scale	DIRE
Adult Maladaptive Functioning	.64*
Employment Trouble	.49*
Mental Instability	.34*
Forensic Risk/Violence	.46*
Childhood/Adolescent Psychopathy	.53*

* $p < .001$, two-tailed

TABLE 2. Correlations of DIRE scale with individual Clinical Data Form items (N = 1,201)

CDF Item	DIRE	CDF Item	DIRE
Adult Maladaptive Functioning		Employment Trouble	
Overall personality functioning	-.47*	Employment functioning	-.46*
Employment functioning	-.46*	Lost job due to interpersonal problems in past 5 years	.39*
Quality of friendships	-.44*	Childhood/Adolescent Psychopathy	
Lost job due to interpersonal problems in past 5 years	.39*	School trouble	.42*
Arrested in past 5 years	.36*	School performance	-.40*
Violent crime in past 5 years	.35*	Physical fights	.37*
The perpetrator in an adult abusive relationship	.34*	Chronic lying	.37*
Prior psychiatric hospitalization	.32*	Substance abuse	.35*
GAF	-.31*	Stealing	.35*
Suicide history	.28*	Age at first intercourse	-.32*
Self-mutilation	.20*	Arrest	.31*
Mental Instability		Running away frequency	.29*
Past suicide attempt	.28*	Promiscuity	.26*
Prior psychiatric hospitalization	.32*	Violent crime	.26*
Forensic Risk/Violence		Animal torture	.21*
Arrested in past 5 years	.36*	Fire setting	.17*
Violent crime in past 5 years	.35*		

* $p < .001$, two-tailed

Discussion of Study 1

The SWAP harnesses reliable clinical observations and inference with respect to personality processes, which are largely lacking from risk-assessment measures beyond global psychiatric disturbance and relatively overt markers of psychopathy. The DIRE scale, derived from the SWAP instrument, shows strong correlations with a wide range of past and current high-risk behaviors and undesirable outcomes, suggesting that the personality features encompassed by the DIRE scale are valid predictors of undesirable outcomes and risk, including, but not limited to, criminality.

A limitation of the study is that the same clinicians who completed the SWAP completed the CDF and knew the subjects' history. Although SWAP items address dynamic psychological and personality variables rather than risky events and outcomes, knowledge of an assessment subject's history may have influenced the scoring of SWAP items, or alternatively, knowledge of current personality functioning may have influenced the scoring of CDF items.

A number of considerations mitigate these concerns. With respect to the CDF, research shows high convergent validity of clinician-rated CDF variables with independent data sources; many of the variables code objective events which leave little room for interpretation; and where clinician- and self-report data occasionally diverged with respect to historical events (e.g., childhood sexual abuse), clinicians were appropriately conservative in their ratings and followed instructions to code "no/unsure" when uncertain (Westen et al., 1997; DeFife et al., 2010; Dutra et al., 2004). With respect to SWAP scores, research consistently shows high inter-rater reliability between treating clinicians and independent research interviewers who score the SWAP on the basis of the CDI or other personality-oriented research interviews that do not address life history (Marin-Avellan, McGauley, Campbell, & Fonagy, 2005; Westen & Muderrisoglu, 2003; Westen & Shedler, 2007; Westen et al., 2012). It therefore appears that SWAP diagnostic scores reflect, as intended, reliable clinical observations and inferences drawn from the here-and-now interaction between clinician and subject. Had the SWAP-II been scored by research interviewers with little or no knowledge of the subjects' history, the SWAP-II diagnostic scale scores would have been largely unchanged (for further discussion, see Westen & Shedler, 2007).

The magnitude of the correlations between DIRE and the composite and individual criterion variables suggests that the DIRE functions as intended, as a measure of personality-related risk. An additional advantage of SWAP

over a history review for risk assessment is that it provides a comprehensive, in-depth assessment of personality that can inform intervention and risk management strategies in ways that a generic tabulation or sum of risky past behaviors cannot.

Study 2: Predictive Validity of DIRE

Overview

This study examined the prospective relation between the DIRE scale and criminal recidivism in a sample of psychiatrically disturbed criminal offenders during a 1-year period of living in the community. Predictive validity of DIRE is compared with that of two standard risk-assessment measures for prediction of criminality and violence with mentally disordered offenders, the HCR-20 and the Hare Psychopathy Check List: Screening Version (PCL-SV; Hart, Cox & Hare, 1995).

Method

All offenders were assessed with the SWAP-200 at baseline. All had been convicted of violent crimes and were either living in the community or about to be discharged into the community. They were followed for a 1-year period. The outcome measure was criminal recidivism, defined as criminal offense(s) during the 1-year follow-up period.

Sample

The initial sample consisted of $N = 35$ psychiatrically disturbed male criminal offenders in England who had been convicted of at least one violent offense and were currently living in the community or about to be discharged from a secure psychiatric facility into the community, having been detained under the 1983 Mental Health Act for England and Wales (Department of Health, 1983). As part of their sentence, they were required to be under the supervision of a forensic psychiatrist, and many were also in treatment with a forensic clinical psychologist. Exclusion criteria were a diagnosis of schizophrenia or impaired intelligence. Outcome data are reported for $N = 31$ offenders for whom follow-up data were available.

All offenders had previously been diagnosed with one or more DSM-IV personality disorders (mean = 1.4), the most prevalent being antisocial personality disorder, as well as a variety of other lifetime or current mental health conditions, the most common of which were lifetime alcohol and/or drug abuse (69%) and current mood or anxiety disorder (17%).

The offenders had an average of 21 criminal convictions, the most common being minor violence (75%) (e.g., assault, affray, actual bodily harm, child cruelty) and crimes against the person (68%) (e.g., harassment, menacing). The offenders had served an average of 3.2 years in prison ($SD = 4.5$). Ninety percent of the sample started to offend between the ages of 10 and 24, and 60% started to engage in violent criminal behavior between the ages of 14 and 24. Mean age was 38 ($SD = 9$). The sample characteristics have been described in greater detail elsewhere (Marin de Avellan, 2010).

Assessment

The SWAP-200 was completed by the offender's treating psychiatrist or psychologist at the start of the study. Clinicians who contributed SWAP-200 data had a minimum of 2 years' experience working with forensic psychiatric patients. A clinical researcher (not involved with offenders' treatment) separately scored the HCR-20 and PCL-SV based on chart review as well as brief conversations with treating clinicians when necessary to clarify or verify information. The HCR-20 (Webster et al., 1997) contains 20 items designed to structure and systematize professional judgments about risk with mentally disordered offenders. The HCR-20 items were combined to create a scale score for research use. The PCL-SV (Hart et al., 1995) is a 12-item version of the Hare Psychopathy Checklist—Revised (Hare, 2003), a standard risk-assessment instrument in forensic and correctional populations, originally developed to assess offenders convicted of violent crimes. The PCL-SV correlates sufficiently highly with the parent test ($r = .95$; Guy & Douglas, 2006) to be considered empirically interchangeable.



Outcome

The outcome measure was criminal recidivism, defined as criminal arrests(s) during a 1- year period living in the community (0 = no reported offense, 1 = one or more arrests).

Results

Ten (32.3%) of the 31 offenders for whom follow-up data were available recidivated during the follow-up period, with eight arrested for violent offenses.

As expected, the mean DIRE score was significantly elevated in the offender sample, with a sample mean of $T = 59.2$ ($SD = 7.1$, range = 47.1 to 72.2), or approximately one standard deviation above the mean of the clinical reference sample. DIRE prospectively and significantly predicted criminal recidivism, $r = .37$ ($p < .05$, two-tailed). To facilitate communication to individuals who may lack a statistical background, the relation between DIRE and recidivism can also be expressed in percentage terms: The probability of criminal recidivism increases by 12.5 percentage points for each 5-point (half a standard deviation) increase in the DIRE score. Receiver operating characteristic (ROC) analysis showed that the DIRE scale discriminated offenders who did and did not recidivate, with an area under the ROC curve of .72. A DIRE cut-score of $T = 59$ correctly identified 80% of offenders who recidivated (sensitivity) and 60% of those who did not recidivate (specificity).

DIRE was a somewhat better predictor of criminal recidivism than the HCR-20 ($M = 22.8$, $SD = 6.7$), which yielded an area under the ROC curve of .65 and a positive but nonsignificant correlation with recidivism, $r = .27$, ns. DIRE was approximately equivalent in prediction to the PCL-SV ($M = 12.3$, $SD = 5.4$), which yielded an area under the ROC curve of .74 and a correlation with recidivism of $r = .42$ ($p < .05$, two-tailed). Findings for the three risk-assessment measures are summarized in Table 3.

TABLE 3. Predictive validity of three measures of risk			
Scale	Mean (SD)	Area under ROC curve	Correlation with Recidivism
DIRE	59.2 (7.1)	.72	.37*
HCR-20	22.8 (6.7)	.65	.27
PCL-SV Total	12.3 (5.4)	.74	.42*

**p < .05, two-tailed*

Discussion of Study 2

This study puts the DIRE scale to a stringent test, due to the use of a psychiatrically disturbed offender sample in which all subjects had DSM personality disorder diagnoses. The study inclusion criteria imposed considerable range restriction on the prediction side of the prediction equation, since all subjects had severe personality pathology and clinically diagnosed personality disorders. Such range restriction

necessarily has the statistical effect of attenuating the relation between the DIRE scale and any outcome measure. Use of a correctional population also places the DIRE scale at a disadvantage relative to the two comparison risk-assessment instruments, both of which were developed to discriminate *within* offender samples. The study sets a far more difficult task for DIRE than assessment and prediction in a general (non-correctional) population, such as in personnel screening, where the majority of subjects do not have criminal histories or diagnosed personality disorders. However, even within this truncated group of criminal offenders with personality disorder diagnoses, DIRE was roughly equivalent to the PCL-SV as a predictor of recidivism and a slightly better predictor than the HCR-20.



General Discussion

One way to break through the current “ceiling” (Skeem & Monahan, 2011, p. 41) on the accuracy of risk prediction is to increase the breadth and depth of dynamic personality constructs (vs. static historical events) addressed by assessment and prediction methods. One advantage of the SWAP and DIRE scale over the Psychopathy Checklist is that the SWAP, for roughly the same expenditure of assessor time and effort, provides a comprehensive assessment of personality and a broad array of psychological information which can inform intervention and risk management strategies.

In non-forensic populations—for example, among individuals who pass background checks for government and private sector positions—there is every reason to believe that DIRE will outperform current risk-assessment procedures that were developed in correctional populations and designed for use specifically when there is a known prior offense. The vast majority of members of the general population do not have criminal backgrounds or show overt signs of antisociality. Where risk-assessment instruments that assume a prior offense may give a “free pass,” DIRE has the potential to identify personality styles and syndromes that could pose risk, including both criminal and non-criminal insider threats in organizations that conduct background checks to screen for criminality and other static variables (past events and behavior) indicative of risk. Even in a truncated, range-restricted sample of psychiatrically disturbed criminal offenders with DSM personality disorder diagnoses, DIRE performed as well as or slightly better than standard, widely used risk-assessment methods. ✓

“

Where risk-assessment instruments that assume a prior offense may give a “free pass,” DIRE has the potential to identify personality styles and syndromes that could pose risk.

.....

.....

REFERENCES

- Ben-Porath, Y. S. & Tellegen A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form) manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Blagov, P., Bi, W., Shedler, J. & Westen, D. (2012). The Shedler-Westen Assessment Procedure (SWAP): Evaluating psychometric questions about its reliability, validity, and impact of its fixed score distribution. *Assessment*, 19, 370-382.
- Block, J. (1971). *Lives through time*. Berkeley, CA: Bancroft.
- Block, J. (1978). *The Q-Sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press.
- Buss, A. H. (1961). *The psychology of aggression*. New York, NY: Wiley.
- Clarkin, J. F., Yeomans, F. E., & Kernberg, O. F. (2006). *Psychotherapy for borderline personality: Focusing on object relations*. Washington, DC: American Psychiatric Publishing.
- Daffern, M., & Howells, K. (2002). Psychiatric inpatient aggression: A review of structural and functional assessment approaches. *Aggression and Violent Behavior*, 7, 477-497.
- DeFife, J. A., Drill, R., Nakash, O., & Westen, D. (2010). Agreement between clinician and patient ratings of adaptive functioning and developmental history. *American Journal of Psychiatry*, 167, 1472-1478. doi: 10.1176/appi.ajp.2010.09101489
- Department of Health (1983). *Mental Health Act*. London, HMSO
- Dutra, L., Campbell, L., & Westen, D. (2004). Quantifying clinical judgment in the assessment of adolescent psychopathology: Reliability, validity, and factor structure of the Child Behavior Checklist for Clinician-Report. *Journal of Clinical Psychology*, 60, 65-85.
- Guy, L. S. & Douglas, K. S. (2006). Examining the utility of the PCL:SV as a screening measure using competing factor models of psychopathy. *Psychological Assessment*, 18, 225-230.
- Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist (2nd ed.)*. Toronto, ON, Canada: Multi-Health Systems
- Harms, P. D., Marbut, A., Johnston, A. C., Lester, P. & Fezzey, T. (2022). Exposing the darkness within: A review of dark personality traits, models, and measures and their relationship to insider threats. *Journal of Information Security and Applications*. 71. 103378. 10.1016/j.jisa.2022.103378. <https://doi.org/10.1016/j.jisa.2022.103378>
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Manual for the Psychopathy Checklist: Screening Version (PCL: SV)*. Toronto, Canada: Multi-Health Systems.
- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, 21, 347-359.
- John, O. (1990). The big five factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York, NY: Guilford Press.
- Kernberg, O. (1975). *Borderline conditions and pathological narcissism*. Northvale, NJ: Jason Aronson.
- Kernberg, O. (1984). *Severe personality disorders*. New Haven, CT: Yale University Press.
- Kohut, H. (1971). *The analysis of the self: A systematic approach to the treatment of narcissistic personality disorders*. New York, NY: International Universities Press.
- Kraemer, H., Kazdin, A., Offord, D., Kessler, R., Jensen, P., & Kupfer, D. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, 54, 337-343.
- Kroner, D., Mills, J., & Morgan, B. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law & Psychiatry*, 28, 360-374.
- Linehan, M.M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press.
- Livesley, W.J. (1995). *The DSM-IV personality disorders*. New York, NY: Guilford Press.
- Marin de Avellan, L. E. (2010). *Systematizing clinician judgment of personality and risk: Validation of the SWAP-200 with forensic patients*. Unpublished manuscript. University College London, London, UK.
- Marin-Avellan, L., McGauley, G., Campbell, C., & Fonagy, P. (2014). The validity and clinical utility of structured diagnoses of antisocial personality disorder with forensic patients. *Journal of Personality Disorders*, 28, 500-517.
- Marin-Avellan, L., McGauley, G., Campbell, C., & Fonagy, P. (2005). Using the SWAP-200 in a personality-disordered forensic population: Is it valid, reliable and useful? *Criminal Behaviour and Mental Health*, 15, 28-45.
- McCrae, R., & Costa, P. (1990). *Personality in adulthood*. New York, NY: Guilford Press.
- McWilliams, N. (1994). *Psychoanalytic diagnosis: Understanding personality structure in the clinical process*. New York, NY: Guilford Press.
- McWilliams, N. & Shedler, J. (2017). *Psychodynamic Diagnostic Manual-2 (PDM-2): Personality Syndromes*. NY: Guilford Press.
- Meloy, J. R. (1988). *The psychopathic mind: Origins, dynamics, and treatment*. Northvale, NJ: Jason Aronson.
- Morey, L. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

REFERENCES

- Nicoletti, J., Spencer-Thomas, S., & Bollinger, C. (1999). *Violence goes to college*. Springfield, IL: Charles C. Thomas.
- Perry, J. C., & Cooper, S. H. (1987). Empirical studies of psychological defense mechanisms. In R. Michels & J. O. J. Caenar (Eds.), *Psychiatry*. Philadelphia, PA: J. B. Lippincott.
- Russ, E., Shedler, J., Bradley, R., & Westen, D. (2008). Refining the construct of narcissistic personality disorder: Diagnostic criteria and subtypes. *American Journal of Psychiatry*, 165, 1473-1481.
- Shapiro, D. (1965). *Neurotic Styles*. New York, NY: Basic Books.
- Shechter, O. G. & Lang, E. L. (2011). Identifying personality disorders that are security risks: Field Test results. *Defense Personnel Security Research Center*, Technical Report 11-05.
- Shedler, J. (2022). The personality syndromes. In R. Feinstein (Ed.), *Personality Disorders*. Oxford: Oxford University Press.
- Shedler, J. (2015). Integrating clinical and empirical perspectives on personality: the Shedler-Westen Assessment Procedure (SWAP). In S. Huprich (Ed.), *Personality disorders: Toward Theoretical and Empirical Integration in Diagnosis and Assessment*. Washington, DC: American Psychological Association.
- Shedler, J. (2009). *Guide to SWAP-200 Interpretation*. Denver, CO: SWAPassessment LLC.
- Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychology*, 48, 1117-1131.
- Shedler, J., & Westen, D. (2004a). Dimensions of personality pathology: An alternative to the Five Factor Model. *American Journal of Psychiatry*, 161, 1743-1754.
- Shedler, J., & Westen, D. (2004b). Refining personality disorder diagnoses: Integrating science and practice. *American Journal of Psychiatry*, 161, 1350-1365.
- Shedler, J., & Westen, D. (2007). The Shedler-Western Assessment Procedure (SWAP): Making personality diagnosis clinically meaningful. *Journal of Personality Assessment*, 89, 41-55.
- Skeem, J. L. & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, 20, 38-42. doi: 10.1177/0963721410397271.
- Vaillant, G. (Ed.). (1992). *Ego mechanisms of defense: A guide for clinicians and researchers*. Washington, DC: American Psychiatric Association Press.
- Webster, C., Douglas, K., Eaves, D., & Hart, S. (1997). *HCR-20: Assessing risk for violence (Version 2)*. Vancouver, Canada: Simon Fraser University.
- Westen, D. (1991). Social cognition and object relations. *Psychological Bulletin*, 109, 429-455.
- Westen, D. (2004). *Clinical Diagnostic Interview: Unpublished Manual*. Emory University, Atlanta, GA. Retrieved from www.psychsystems.net.
- Westen, D., Lohr, N., Silk, K.R., Gold, L., & Kerber, K. (1990). Object relations and social cognition in borderlines, major depressives, and normals: A Thematic Apperception Test analysis. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 355-364.
- 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. *American Journal of Psychiatry*, 163, 1285-1287.
- Westen, D., Muderrisoglu, S., Fowler, C., Shedler, J., & Koren, D. (1997). Affect regulation and affective experience: Individual differences, group differences, and measurement using a Q-sort procedure. *Journal of Consulting & Clinical Psychology*, 65, 429-439.
- Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, Part 1: 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 2: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273-285.
- Westen, D., & Shedler, J. (2007). Personality diagnosis with the Shedler-Western Assessment Procedure (SWAP): Integrating clinical and statistical measurement and prediction. *Journal of Abnormal Psychology*, 116(4), 810-822.
- Westen, D., Shedler, J., & Bradley, R. (2006). A prototype approach to personality diagnosis. *American Journal of Psychiatry*, 163, 846-856.
- Westen, D., Shedler, J., Bradley, B., DeFife, J. (2012). An empirically derived taxonomy for personality diagnosis: Bridging science and practice in conceptualizing personality. *American Journal of Psychiatry*, 169, 273-284.
- Westen, D., Waller, N. G., Shedler, J., & Blagov, P. S. (2014). Dimensions of Personality and Personality Pathology: Factor Structure of the Shedler-Western Assessment Procedure-II (SWAP-II). *Journal of Personality Disorders*. 28. 281-318. 10.1521/pedi_2012_26_059.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist*, 59, 595-613.
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136, 740-767.

