

Structured Interview of Personality Organization (STIPO): Preliminary Psychometrics in a Clinical Sample

BARRY L. STERN,¹ EVE CALIGOR,¹ JOHN F. CLARKIN,² KENNETH L. CRITCHFIELD,³ SUSANNE HORZ,⁴
VERNA MACCORNACK,¹ MARK F. LENZENWEGER,^{2,5} AND OTTO F. KERNBERG²

¹Columbia University College of Physicians and Surgeons and the Columbia Center for Psychoanalytic Training and Research

²Personality Disorders Institute, Weill Medical College of Cornell University

³Department of Psychology, University of Utah

⁴Department of Psychology, Ludwig-Maximilians-Universitaet Muenchen, Germany

⁵Department of Psychology, State University of New York at Binghamton

In this article, we describe the development and preliminary psychometric properties of the Structured Interview of Personality Organization (STIPO), a semistructured interview designed for the dimensional assessment of identity, primitive defenses, and reality testing, the three primary content domains in the model of personality health and disorder elaborated by Kernberg (1984; Kernberg & Caligor, 2005). Results of this investigation, conducted in a clinical sample representing a broad range of personality pathology, indicate that identity and primitive defenses as operationalized in the STIPO are internally consistent and that interrater reliability for all 3 content domains is adequate. Validity findings suggest that the assessment of one's sense of self and significant others (Identity) is predictive of measures of positive and negative affect, whereas the maladaptive ways in which the subject uses his or her objects for purposes of regulating one's self experience (Primitive Defenses) is predictive of measures of aggression and personality disorder traits associated with cluster B personality disorders. We discuss implications of these findings in terms of the theory-driven and trait-based assessment of personality pathology.

Alongside the empirical elaboration of Axis II of the *Diagnostic and Statistical Manual for Mental Disorders* (4th ed., text revision [DSM-IV-TR]; American Psychiatric Association, 2000) and the host of standardized assessment tools keyed to the DSM criteria sets (e.g., Loranger, 1999; First, Spitzer, Gibbon, Williams, & Benjamin, 1997), other theory-driven approaches to the study of personality disorder have emerged in recent years with accompanying assessment technologies (e.g., Benjamin, 2005; Westen & Shedler, 1999a, 199b). In this report, we present preliminary data on the psychometric development of the Structured Interview of Personality Organization (STIPO), a semi-structured interview developed to assess dimensions of personality pathology within Kernberg's (1984) particular psychodynamic frame of reference. Informed by contemporary object relations theory, Kernberg (1984; Kernberg & Caligor, 2005) has described what is essentially a dimensional model of personality centered on the assessment of three key domains: identity disturbance, primitive psychological defenses, and reality testing (1984; Kernberg & Caligor, 2005). In his clinical writing and teaching, Kernberg characterized the clinical psychopathology of personality disorder according to the individual's standing on these three dimensions as falling into two primary regions: the borderline (BPO) and neurotic (NPO) levels of "personality organization." The purpose of this study was not to validate these putative typological classifications; rather, we sought to establish preliminary reliability and validity of

the underlying clinical dimensions, that is, identity, primitive defenses, and reality testing.

Identity in Kernberg's (1984) model is comprised of the various ways in which individuals experience themselves in relation to others. Normal identity is based on the individual's ability to shift flexibly across various self-representations, resulting in an appraisal of the self that is realistic and integrated, with an ability to tolerate both the positively and negatively imbued qualities of the self and a correspondingly realistic and stable experience of others. In contrast, identity pathology is characterized by inflexible and unstable, poorly integrated, black and white (i.e., "all good" or "all bad") experiences of self and other, with a resulting incoherence in the experience of self and others and a predominance of negative affects as well as instability and conflict in the interpersonal sphere (Kernberg & Caligor, 2005).

Working hand in glove with the construct of identity is the construct of primitive defensive operations, conceived as psychological strategies for the regulation of emotion carried on outside of conscious awareness, which involve the separation of positive and negative sectors of experience (commonly referred to as "splitting"). This separation comes at the expense of the individual's maintaining a distorted, fragmentary, caricatured, and brittle sense of self and others. In addition to the assessment of identity and primitive defenses, the assessment of reality testing, specifically, attunement to social and interpersonal norms and the demonstration of expected tact and empathy (as opposed to the ability to separate self from nonself and internal from external stimuli) is central to the classification of personality organization. In sum, greater identity pathology and the more extensive use of primitive defenses, in the context of grossly intact reality testing, are pathognomonic of severe personality

Received February 21, 2009; Revised July 17, 2009.
Address correspondence to Barry L. Stern, Columbia University College of Physicians and Surgeons, 286 Madison Avenue, Suite 1602, New York, NY 10017; Email: bs2137@columbia.edu

pathology in Kernberg's model and are thought to be associated with the significant dysfunction in adult interpersonal relationships that is characteristic of the personality disorders.

We have pursued a program of instrument development that allows for the reliable and valid assessment of these three principal components of Kernberg's (1984) structural model. An initial study in a nonclinical sample of the psychometric properties of the Inventory of Personality Organization (IPO; Clarkin, Foelsch, & Kernberg, 2001; Lenzenweger, Clarkin, Kernberg, & Foelsch, 2001), a self-report questionnaire assessing the same model of personality, revealed positive associations between the IPO Identity, Primitive Defenses, and Reality Testing scales and indexes of negative affect and aggression, and inverse relations between Identity and Primitive Defenses and a measure of positive affect (Lenzenweger et al., 2001). These findings are broadly consistent with a neurobehavioral view of the personality disorders as being characterized by high levels of trait neuroticism (depression, anxiety, vulnerability) in the context of diminished constraint as expressed in measures of aggression (see Depue & Lenzenweger, 2001, 2005; see also Lenzenweger, Clarkin, Fertuck, & Kernberg, 2004). The inverse relation between the IPO scales and measures of positive affect suggested a diminished capacity for incentive or reward-motivated behavior, perhaps considered as evidence of a "muted approach system" (Lenzenweger et al., 2001). Additionally, significant relations were revealed between the IPO Reality Testing scale and various measures of schizotypy or psychosis proneness. Taken together, these findings provide preliminary support for the three conceptual pillars of Kernberg's (1984) structural model (identity, primitive defenses, and reality testing). The work cited above also highlighted the interconnectedness of identity pathology and primitive defenses as assessed in self-report format. These early results involving the questionnaire-based IPO encouraged us to develop the STIPO, as to our minds, the existence of complementary self-report and interview-based measures would be useful in the continued validation of this particular model of personality organization and in serving various clinical and research needs related to the assessment of disordered personality.

HYPOTHESES AND PLANNED STATISTICAL ANALYSES

The objective of this preliminary study was to examine the initial psychometric properties of the STIPO scales. First and foremost, we sought to examine the internal consistency and interrater reliability of the STIPO in a diverse clinical sample.¹ Second, we sought to evaluate in a preliminary manner the criterion and convergent relations between the STIPO scales and personality disorder traits as well as other constructs known to be closely associated with personality pathology, for example, negative and positive affect, aggression, and interpersonal adjustment. Constructs related to dysregulation of affect and aggressive dyscontrol were chosen because they are central to the phenomenology of Axis II, particularly Cluster B (e.g., borderline) disorders and because they are also grounded in well-established biological systems (Depue & Lenzenweger, 2005).

Although conceptually distinct in theory, we are mindful of the interdependence of the identity and primitive defenses constructs as well as their statistical association in the initial study

involving the IPO (Lenzenweger et al., 2001). Nonetheless, we sought to examine the independent contributions of each STIPO domain in the prediction of each outcome measure. We hypothesized that individuals characterized by significant identity pathology and the extensive use of primitive defenses (i.e., probable cases of BPO), would exhibit higher levels of negative affect and aggression, high levels of interpersonal distress, and lower levels of positive affect. Consistent with the view that impairments in reality testing in the borderline spectrum occur principally under conditions of what has been termed "peak affect activation or aggression" (Kernberg & Caligor, 2005; as opposed to being trait-like, enduring impairments such as those seen in schizotypic pathology), we did not anticipate that reality testing would contribute significantly to the prediction of measures of affect, aggression, or personality disorder traits above and beyond what is accounted for by the identity and primitive defenses domains. We did anticipate that after controlling for the independent effects of identity and primitive defenses, reality testing would be significantly associated with measures of cognitive impairment.

METHOD

Development of the STIPO

In writing items for the STIPO, we attempted to cover the breadth of each of the three content domains to ensure content validity. Although we attempted to create domains that would be conceptually and empirically distinct, prior research findings and the theory itself predict some covariation and conceptual overlap among the STIPO domains. Thus, item writing required considerable care and discussion of the clinical and theoretical implications of proposed content. Several STIPO items are expanded versions of parallel items contained in the IPO. Although the two measures can be viewed as alternative methods for the assessment of Kernberg's theory of personality organization, the content sampled across the two instruments is somewhat divergent and, as one would expect in a semistructured interview, the STIPO contains follow-up probes for each item designed to enhance accuracy in scoring. Most items were written by J. F. Clarkin, E. Caligor, and B. L. Stern. In the development, testing, and the placement of STIPO items within the various domains, we worked in close consultation with O. F. Kernberg, who reviewed all STIPO items and provided critical feedback.

The STIPO asks two types of questions: (a) those regarding descriptive features that the model suggests are related, but not limited to, pathology of internal object relations (e.g., regular and significant interpersonal conflict, inability to direct consistent and productive energy and attention toward work or school); and (b) those regarding the internal experience that we have come to observe clinically as characteristic of those with personality disorders (e.g., dramatic shifts in the experience or perception of self and/or other, descriptions of self and/or other that lack a sense of depth and reality). As an example, in assessing the erratic behavior characteristic of personality disordered patients that is felt to signify the operation of primitive defenses, the STIPO asks

Do you act in ways that appear to others as unpredictable and erratic? Do people tell you that you behave in contradictory ways, or would you say that people pretty much know what to expect from you in terms of your behavior? Are people regularly surprised by your behavior?

¹Test-retest reliability, an additional and essential component of reliability testing, is currently being investigated in a separate study.

In contrast, an item from the identity domain focusing on a more subtle, intrapsychic experience reads

If you look back over your life the past 5 years would you say that you have a continuous sense of yourself moving through time? Does the passage of time, or the series of events in your life feel like a steady flow, or would you say that it feels choppy or broken up?

To promote reliability among raters and ease of use, we adopted an administration and scoring format similar to that used in the International Personality Disorders Examination (IPDE; Loranger, 1999) and Structured Clinical Interview for the *DSM-IV* [American Psychiatric Association, 1994] Axis II disorders (SCID-II; First et al., 1997). Items begin with a stem question with standardized follow-up probes depending on the response. For most STIPO items, the interviewer is prompted to ask the subject to elaborate on affirmative responses by providing a compelling example, and the interviewer then determines the extent to which the quality being assessed is characteristic of the respondent across time and situation and the extent to which it causes impairment in the respondent's functioning. STIPO items are scored on a 3-point scale similar to the IPDE: 0 = the quality being assessed is not present at all, or, if present, has no impact on respondent's functioning; 1 = the quality being queried is present, but results in only minor impairment in functioning; and, 2 = the quality being assessed is present and reflects significant to severe pathology and impairment in functioning. Additional clarification of these scoring guidelines is provided for most items. In answering the interview questions, respondents are asked to reflect on their life over the past 5 years. STIPO domain scores are calculated by taking the mean of the 0 to 2 ratings for the items in that domain.

We revised the STIPO item pool with items added and deleted several times prior to the start of the study described in this article. Feedback from patients participating in pilot interviews as well as from the interviewers (B. L. Stern, S. Horz, and V. MacCormack) concerning the clarity of the items and the interpretability of responses was one criterion used in deciding whether to retain or delete an item. Feedback from O. F. Kernberg, E. Caligor, and J. F. Clarkin concerning the fit between the item and the targeted construct, content overlap, and the extent to which items might be useful in discriminating between various levels of personality pathology was also considered in developing the item pool. We used a final 100-item version of the STIPO in this study, 34 items of which were included in the three scales examined in this study: Identity (19 items), Primitive Defenses (10 items), and Reality Testing (5 items). The remaining items were grouped into four subscales not examined in this report: Quality of Object Relations, Coping and Rigidity (a scale of mature defenses), Aggression, and Moral Values. After all study data had been collected, three items (two from Identity, and one from Reality Testing) were deleted due to poor performance in a preliminary set of conventional item analyses (alpha-if-item-deleted, corrected-item-total correlations), resulting in a final 31-item set comprising the three scales examined in the subsequent analyses reported. The final item pool, along with item-level statistical data, is presented in Table 1.

Participants

For this preliminary reliability/validity study, we gathered data from clinical samples representing a broad spectrum of personality adjustment and psychopathology. Accordingly, our

sample includes both inpatients and outpatients from two different sites. The first site includes data from 30 inpatients drawn from a unit specializing in personality disorders at a university-based psychiatric hospital in the New York metropolitan area. Data from this site also include 24 outpatients being seen in clinics and private practices of clinicians associated with the same hospital. Unit staff and therapists working in the clinics and private practices were notified that the research opportunity was available; that patients would be compensated with \$25 for their participation; and that, with the patient's consent, feedback related to the STIPO profile could be provided to the therapist. All patients at this site who initiated the research for whom complete data was available were included in the study ($n = 54$). In addition, consecutive applicants for psychoanalytic treatment by candidates at a psychoanalytic training institute affiliated with a major training hospital in the metropolitan New York were also included in the sample ($n = 88$). All applicants to the clinic were required to complete a battery of standardized tests as a condition of their evaluation for psychoanalytic treatment along with the STIPO. Overall, the study sampled patients that were expected to span the range of relevant pathology.

Exclusionary criteria for the study based on psychiatric illness included the presence of schizophrenia or other psychotic disorders as well as bipolar disorders. Of the participants, 38% were male and 62% female, with ages ranging from 20 to 55 years and a mean age of 32 years. Demographic and clinical variables are described in Table 2. Of the combined sample, 80% had a clinician-assigned primary psychiatric diagnosis, the most commonly assigned diagnosis being an affective disorder. Of note, 24% of the sample had a clinician-assigned primary diagnosis of a personality disorder. An additional estimate of the amount of Axis II personality pathology in the sample can be derived from the Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993; described following); if SNAP-derived *DSM-III-R* (American Psychiatric Association, 1987) personality disorder symptom counts were converted into diagnoses, 56% of the sample would have met criteria for at least one *DSM* personality disorder, suggesting appreciable variability in personality pathology in the sample. Of study participants, 52% were taking some psychotropic medication at the time of evaluation, with the most commonly described medication being antidepressants.²

Inspection of Table 2 also reveals several notable differences between the hospital and analytic center samples. Although the gender and age distribution is largely consistent across the samples, the hospital participants were less well-educated (mean number of years of postsecondary education = 2.34 vs. 4.87 for the psychoanalytic center sample) and, due in part to the presence of a significant number of inpatients in this subsample, evidenced more significant and diverse psychopathology. In contrast to the psychoanalytic-center sample, a greater number of the hospital participants were assigned a primary clinical diagnosis by their clinicians, and the total number of primary diagnoses assigned per patient was higher for hospital study participants. Participants in the hospital sample also were more likely to be taking psychotropic medication, took a greater

²Study participants who reported taking either a mood-stabilizer or anti-psychotic medication were not doing so due to a disclosed bipolar or psychotic condition, which were exclusionary criteria for the study, but rather, we presume, as part of the medical management of other, included conditions.

TABLE 1.—Structured Interview of Personality Organization item stems and scale- and item-level statistics.

Item Stem	<i>M</i> (<i>SD</i>)	Skew	Kurtosis	Correlation		
				<i>i-t</i>	<i>i-pd</i>	<i>i-rt</i>
Identity, $\alpha = .86$	0.94 (0.40)	-.14	-0.45	—	—	—
1. Effectiveness in primary role (work/school); performance in relation to subjective sense of ability	0.91 (0.63)	.24	-0.68	.62	.47	.38
2. Stability in primary role (work/school); leaves of absence, consistent engagement in work/school over time	0.87 (0.76)	.33	-1.26	.52	.43	.37
3. Importance of primary role (work/school); continuity in reported ambition/goals over time	0.91 (0.67)	.22	-1.02	.60	.47	.40
4. Satisfaction and pleasure from primary role (work/school)	0.78 (0.70)	.43	-0.99	.48	.38	.29
5. Presence of and capacity to invest in recreational pursuits, i.e., hobbies, religious, cultural, and/or athletic pursuits	0.99 (0.73)	-.06	-1.16	.50	.25	.27
6. Stability and consistent engagement in the identified recreational pursuits over time	1.06 (0.72)	-.09	-1.05	.55	.29	.27
7. Satisfaction and pleasure derived from recreational pursuits	0.90 (0.70)	.16	-0.92	.46	.23	.22
8. Open-ended self description, coded for fluency, superficiality/depth, and ambivalence	1.10 (0.66)	-.10	-0.66	.54	.39	.45
9. Continuous and consistent experience of the self experiencing the events of one's life through time	0.80 (0.78)	.38	-1.21	.58	.48	.40
10. Consistency in opinions, values, tastes, and preferences over time	0.70 (0.69)	.51	-0.80	.43	.45	.41
11. Consistency in the experience of the self across situation and person	0.89 (0.77)	.19	-1.32	.57	.53	.46
12. The experience of being alone, i.e., does it enhance a sense of groundedness in one's identity, or result in anxiety, confusion about the self	0.70 (0.73)	.54	-0.97	.43	.40	.39
13. The tendency to "lose" one's sense of self in the context of an intimate relationship	0.88 (0.66)	.13	-0.72	.47	.36	.37
14. Self-esteem; stability vs. vulnerability, and valence	1.03 (0.73)	-.06	-1.12	.26	.45	.31
15. Resiliency of sense of self and self esteem to external feedback and influence	1.21 (0.71)	-.35	-0.99	.50	.49	.55
16. Open-ended description of other, coded for fluency, superficiality/depth, and ambivalence	1.09 (0.65)	-.08	-0.59	.51	.40	.43
17. Open-ended description of parent, coded for fluency, superficiality/depth, and ambivalence	1.02 (0.58)	.01	0.07	.43	.23	.34
Primitive Defenses, $\alpha = .85$	0.87 (0.46)	-.18	-0.47	—	—	—
1. Primitive projection, i.e., the tendency to keep information about oneself from others, a consistent attitude of vigilance in one's social relations	0.98 (0.70)	.03	-0.99	.63	.41	.50
2. Unpredictable and erratic behavior; the tendency to exhibit contradictory behaviors to others	0.74 (0.75)	.46	-1.12	.59	.54	.51
3. Idealization/devaluation, i.e., drastic, unpredictable, or frequent shifts in one's desire for or valuation of things (objects, values, goals)	0.79 (0.73)	.35	-1.06	.58	.42	.40
4. Idealization/devaluation, i.e., drastic, unpredictable, and/or frequent shifts in one's experience or valuation of people	0.93 (0.66)	.10	-0.74	.70	.50	.57
5. Black and white, all or nothing thinking; the tendency to view situations in the extreme	1.05 (0.70)	-.08	-0.97	.58	.52	.54
6. Primitive denial, i.e., the tendency to significantly deny or ignore painful yet important situations in one's life	0.97 (0.78)	.04	-1.33	.51	.47	.31
7. Externalization, i.e., the tendency to view one's problems as originating in or caused by others, circumstances, or the outside world generally	0.80 (0.71)	.32	-0.97	.52	.47	.40
8. Projective identification, i.e., the tendency to experience provocation, unjust treatment, disrespect or disregard, and to guard against control or manipulation by others	0.99 (0.69)	.02	-0.85	.60	.50	.57
9. The tendency to retreat from demands of reality to fantasy, daydreaming, avoidance	0.61 (0.65)	.59	-0.63	.47	.36	.39
10. The tendency to react to emotional stress through the development of somatic symptoms	1.00 (0.68)	-.01	-0.80	.30	.25	.20
Reality Testing $\alpha = .69$	0.95 (0.48)	-.07	-0.49	—	—	—
1. Paranoia, i.e., mistrust or suspicion of others, their motivations and truthfulness, including those closest to the oneself	0.67 (0.57)	.13	-0.66	.38	.34	.47
2. Confusion in the assessment of the thoughts, feelings, motivations of others; tendency to misjudge what others are thinking or feeling, to offend others or put others off without knowing why	0.89 (0.74)	.17	-1.15	.52	.46	.49
3. Confusion as to how one is experienced by others, i.e., what people appreciate or dislike about oneself	1.19 (0.70)	-.29	-0.95	.58	.47	.45
4. A tendency to overreact, lose perspective and think clearly under stressful conditions	1.03 (0.61)	-.02	-0.29	.43	.57	.53

Note. *N* = 142. Correlation *i-t* denotes adjusted item-to-scale correlation; Correlation *i-id* denotes item-identity domain; Correlation *i-pd* denotes item-primitive-defenses domain; Correlation *i-rt* denotes item-reality-testing domain.

number of psychotropic medications, and had a higher total number of psychiatric hospitalizations than their counterparts in the analytic center sample.

Measures

In addition to the STIPO, participants referred from both sites completed two self-report questionnaires that served as the primary validity measures in our study. The IPO (Clarkin et al.,

2001) is an 83-item self-report questionnaire designed, like the STIPO, to assess the content domains central to Kernberg's theory of personality. The IPO consists of three primary clinical scales (Identity, Primitive Defenses, and Reality Testing) and two secondary scales (Aggression and Moral Values). The IPO Identity, Primitive Defenses, and Reality Testing scales demonstrate sound internal consistency (.81, .88, and .88, respectively, in the original Lenzenweger et al. [2001] validation study; and

TABLE 2.—Demographic and clinical variables grouped by sample.

Variable	Hospital ^a				Psychoanalytic Clinic ^b				Statistical Significance
	Range		<i>M</i>	<i>SD</i>	Range		<i>M</i>	<i>SD</i>	
	Min	Max			Min	Max			
Age	20	55	32.52	8.74	20	55	32.85	8.01	.31
Years of postsecondary education	-4	8	2.34	2.61	0	9	4.87	2.32	.04
No. of psychiatric hospitalizations	0	15	2.08	2.70	0	3	.14	.49	.001
Total no. of psychotropic medications	0	6	2.44	1.48	0	5	.46	.85	.001
Total no. of psychiatric diagnoses	0	4	1.76	1.00	0	3	1.05	.77	.01
			<i>N</i>	%			<i>N</i>	%	
Gender									.15
Male			17	32%			38	43%	
Female			37	68%			50	57%	
Primary diagnosis									
Affective disorder			29	54%			51	58%	.80
Anxiety disorder			5	9%			21	24%	.05
Substance abuse disorder			11	20%			8	9%	.03
Eating disorder			2	4%			4	5%	.91
Personality disorder			32	71%			2	2%	.001
Medications									
Antidepressant			43	80%			27	31%	.001
Anxiolytic			10	19%			1	1%	.001
Sleep medication			9	17%			1	1%	.001
Antipsychotic			31	57%			1	1%	.006
Benzodiazepine			9	17%			3	3%	.001
Mood stabilizer			17	32%			3	3%	.001

Note. Statistical significance was determined for dimensional variables by *t* test and for categorical variables by chi-square test.
^a*N* = 54. ^b*N* = 88.

.91, .89, and .89, respectively, in this study) and short-term test-retest reliability (Foelsch et al., 2000). The primary IPO scales have also been shown to be associated with measures of negative affect and aggression (Lenzenweger et al., 2001).

The SNAP (Clark, 1993) is a 375-item, true-false, self-report questionnaire designed to assess traits associated with personality disorders. The SNAP item pool includes the *DSM-III-R* (American Psychiatric Association, 1987) diagnostic criteria for the personality disorders as well as traits derived from other conceptualizations of personality disorder and some related Axis I traits. The measure provides 12 trait scales (e.g., Entitlement, Mistrust, Aggression, Propriety) and three temperament scales reflecting affective states (Positive Temperament, Negative Temperament, and Disinhibition). The SNAP items can also be scored to assess the diagnostic criteria for the personality disorders found in the *DSM-III-R*,³ and these scores demonstrate strong correlations with interview-based measures of Axis II. The SNAP scales have demonstrated adequate internal consistency and test-retest reliability (Clark, 1993).

Additional questionnaires were available for participants referred from the hospital; these include a well-established, multidimensional measure of aggression, the Buss-Durkee Inventory (Buss & Durkee, 1957), and the Inventory of Interpersonal Problems–Personality Disorder scales (IIP-PD; Pilkonis, Kim, Proietti, & Barkham, 1996; Stern, Kim, Trull, Scarpa, & Pilko-

nis, 2000), a 28-item personality disorder screen derived from the Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureno, & Villaseñor, 1988), which was developed as a brief measure of personality-disorder proneness and was included in this study as an index of interpersonal distress.

Interviewers

Three women and two men with appreciable diagnostic experience served as interviewers for this study. Interviewers included one psychiatrist, one clinical psychology doctoral trainee (author, S. Horz), one clinical psychologist/psychoanalytic institute trainee (author, B. L. Stern), a senior psychoanalyst (author, V. MacCornack), and a licensed clinical social worker. In addition to clinical expertise in the diagnosis of personality disorders, each of the interviewers was well versed in Kernberg's (1984) psychoanalytic theory and experienced in the treatment of personality disordered patients. All interviewers received training in the administration and scoring of the STIPO, which involved didactic training, videotape review, and the observation and scoring of live interviews conducted by B. L. Stern. The trainees' scores for videotaped and live interviews were reviewed at the completion of the interviews as a further didactic tool, and we discussed and corrected discrepancies in scoring. Additionally, each of the interviewers conducted several practice interviews before interviewing study participants. The practice interviews were observed by B. L. Stern, and critical feedback was provided.

Statistical Analyses

Our test of convergent correlations between STIPO domains and indexes of personality disorder traits and personality

³The SNAP personality disorder symptom indices are based on *DSM-III-R* criteria; the SNAP was revised to reflect changes in *DSM-IV* after data collection for this study began. Dimensional SNAP symptom indexes for *DSM-III-R* and *DSM-IV* are highly correlated, with convergence coefficients ranging from .81 to 1.00 and an average of .95 (L. A. Clark, personal communication, September 25, 2005).

TABLE 3.—Intercorrelations among STIPO domains and measures of affect, aggression, cognitive impairment, and interpersonal distress with STIPO domain score internal consistency estimates on the diagonal.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Identity	.86														
2. Primitive Defenses	.68	.85													
3. Reality Testing	.64	.67	.69												
4. SNAP Positive Temperament	-.31	-.17	-.21	—											
5. SNAP Negative Temperament	.56	.51	.58	-.24	—										
6. IPO Aggression	.39	.47	.44	-.05	.61	—									
7. SNAP Aggression	.36	.48	.39	.02	.60	.74	—								
8. SNAP Disinhibition	.42	.48	.34	-.07	.39	.45	.40	—							
9. Buss–Durkee Irritability	.57	.61	.60	-.27	.73	.63	.75	.51	—						
10. IPO Reality Testing	.39	.44	.45	-.03	.58	.62	.44	.43	.53	—					
11. SNAP Eccentric Perceptions	.29	.36	.34	.03	.51	.56	.45	.49	.37	.71	—				
12. IIP–PD	.57	.63	.59	-.22	.69	.75	.75	.67	.78	.67	.55	—			
13. SNAP Cluster A Index	.49	.50	.52	-.28	.62	.65	.57	.45	.70	.65	.67	.74	—		
14. SNAP Cluster B Index	.39	.54	.41	-.13	.61	.67	.69	.70	.62	.56	.59	.69	.53	—	
15. SNAP Cluster C Index	.36	.37	.45	-.34	.61	.46	.35	.16	.50	.41	.38	.50	.72	.30	—

Note. *N* = 142. STIPO = Structured Interview of Personality Organization; SNAP = Schedule for Nonadaptive and Adaptive Personality; IPO = Inventory of Personality Organization; IIP–PD = Interpersonal Problems–Personality Disorder scales. Bold font indicates STIPO domain internal consistency statistics.

features involved a multiple regression approach advocated by Darlington (1990). In addition to reporting traditional Pearson correlation statistics (*r*), which do not account for covariation among predictor variables (herein, the STIPO domain scores), we report the semipartial correlation statistic, which was derived from the simultaneous entry of all three STIPO domains into a multiple regression model. This procedure statistically isolates the relatively unique contribution of each independent variable (Identity, Primitive Defenses, and Reality Testing) in the prediction of validity measures. We conducted separate regression analyses, all following this model, for each of the selected validity measures, for example, separate analyses for positive and for negative affect, for each measure of aggression, and for each of the personality disorder trait clusters. All of the reported regression analyses involved a two stage procedure, with demographic variables entered simultaneously as a block in a first step (age, gender, education, and recruitment site), and the three STIPO domain scores entered simultaneously as a block in the second step. We did not use a stepwise selection method (e.g., stepwise, forward, backward) in the variable entry process.

RESULTS

Reliability

Internal consistency, as assessed by Cronbach’s alpha, for each of the three STIPO domains, is relatively high for the Identity (.86; 17 items) and Primitive Defenses (.85; 10 items) domains and therefore acceptable for further research. The internal consistency for the Reality Testing (.69; 4 items) domain falls just short of the .70 benchmark established by Nunnally and Bernstein (1994), but it is most likely a function of the scale’s brief length. Intercorrelations among the STIPO domains can be found in Table 3, and suggest that, consistent with our expectation, Identity, Primitive Defenses, and Reality Testing cannot be conceived of as statistically independent domains (they are clearly thought of as interactive constructs, albeit relatively distinct in conceptual terms).

A total of 16 STIPOs (approximately 30% of the hospital sample) were attended by two interviewers, one administering and scoring the interview and the second observing and scoring but not administering the interview (each served as interviewer

for approximately half of the interviews and as the reliability coder for the other half). Although each of the two interviewers had adequate training in the theoretical underpinnings of the STIPO and in the administration of the interview, they were the interviewers with the least clinical experience from among the interviewer cohort. We calculated interrater reliability based on this subset of interviews for scores on each of the three STIPO domains. We report the intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979) for the three STIPO domains along with descriptive data on the interrater reliability subsample in Table 4. The ICCs for each STIPO domain were acceptable: Identity, .96; Primitive Defenses, .97; and Reality Testing, .72.

Initial Criterion Validation

IPO. One step in the construct validation process involves testing relations between the target construct and measures of the same construct using different methods. In the case of Kernberg’s (1984) theory of personality organization, the interview-based STIPO and the IPO, a questionnaire, are parallel measures. Results revealed correlations in the *p* < .001 range for each of the three domains assessed in both instruments: Identity = .57; Primitive Defenses = .56; and Reality Testing = .45.

Affect. Following Lenzenweger et al. (2001), we sought to link the STIPO domains to measures of negative and positive affect. Inspection of the bivariate correlations revealed that all associations between the measures of affect and the three STIPO domains were statistically significant at or below the

TABLE 4.—Interrater reliability for STIPO domains.

STIPO Domain	Range	<i>M</i>	<i>SD</i>	ICC
Identity	0.65–1.79	1.21	.33	.96
Primitive Defenses	0.4–1.6	1.12	.35	.97
Reality Testing	1.0–1.75	1.32	.26	.72

Note. *N* = 16. STIPO = Structured Interview of Personality Organization; ICC = intraclass correlation coefficients.

TABLE 5.—Incremental validity analysis of STIPO domains in predicting measures of affect, aggression, cognitive impairment, and personality disorder traits after controlling for demographic variables.

	R^2 Block 1 Demographic Variables	ΔR^2 Block 2 STIPO Domains	Identity		Primitive Defenses		Reality Testing	
			Semipartial (Part) Correlation	p	Semipartial (Part) Correlation	p	Semipartial (Part) Correlation	p
Affect								
SNAP Negative Temperament	.08*	.36***	.21	.003	.05	.464	.23	.001
SNAP Positive Temperament	.03	.12***	-.21	.016	.04	.611	-.10	.240
Aggression								
IPO Aggression	.14*	.23***	.08	.249	.16	.031	.14	.065
SNAP Aggression	.06	.21***	.07	.372	.20	.015	.09	.271
SNAP Disinhibition	.11**	.21***	.12	.107	.26	.001	-.07	.376
Buss–Durkee Irritability	.04	.42***	.14	.225	.16	.158	.16	.165
Cognitive impairment								
IPO Reality Testing	.06	.24***	.13	.096	.15	.053	.10	.176
SNAP Eccentric Perceptions	.03	.12***	.08	.328	.09	.276	.09	.274
Personality disorder traits								
SNAP Cluster A Index	.06	.29***	.16	.03	.10	.193	.18	.016
SNAP Cluster B Index	.11*	.23***	.04	.629	.29	.001	.02	.762
SNAP Cluster C Index	.09*	.17***	.05	.547	.05	.566	.22	.006
IIP–PD	.07	.42***	.18	.107	.13	.246	.15	.171

Note. $N = 142$. STIPO = Structured Interview of Personality Organization; SNAP = Schedule for Nonadaptive and Adaptive Personality; IPO = Inventory of Personality Organization; IIP–PD = Interpersonal Problems–Personality Disorder scales. All Block 1 demographic variables (age, gender, education, and recruitment site) were entered simultaneously; all Block 2 STIPO domains (Identity, Primitive Defenses, and Reality Testing) were also entered simultaneously. Buss–Durkee and IIP–PD data are based on the hospital sample only ($n = 54$). Analyses in which we did not control for demographic variables and recruitment site, i.e., in which step 1 was eliminated, were broadly comparable. Bold font indicates semipartial correlations that are statistically significant at or below the $p < .05$ level.
 $p < .05$. ** $p < .01$. *** $p < .001$.

$p < .05$ level (see Table 3⁴). This indicates that, as predicted, greater levels of personality pathology on the STIPO are positively associated with negative affect and inversely associated with positive affect. Moving beyond the bivariate correlations, a more rigorous regression approach sought to test the relatively independent (or unique) contribution of each STIPO domain to the prediction of measures of affect. By entering all predictors (STIPO domain scores) simultaneously and calculating the semipartial (part) correlations, we sought to determine the relatively unique contribution of each of the three STIPO domain predictors (Darlington, 1990). We ran separate regression analyses based on this general approach for each measure of affect tested. The semipartial (part) correlations that resulted from these multiple regression analyses as well as the total variance explained by this model in the measures of affect are summarized in Table 5. The basic regression model consisted of the three STIPO domains entered directly and simultaneously as a second step after first controlling for the effects of demographic variables and resulted in an increment of variance of 36% for negative affect ($p < .001$) and 12% for positive affect ($p < .001$). The Identity domain was positively associated with negative affect and inversely related to positive affect, suggesting a moderate link between the STIPO Identity domain and measures of affect. In follow-up regression analyses involving the Identity domain, although speculative, we sought to determine which groups of items within the Identity domain most were responsible for the observed relations to measures of affect. These analyses suggested that the association between negative affect and Identity is driven largely by groups of items reflect-

ing the stability and consistency of the subject's sense of self across time, place, and situation. In contrast, these same post hoc regression analyses suggested that STIPO Identity domain items reflecting the capacity to invest oneself in one's primary role (work or school) and recreational pursuits appear to be responsible for the relation between the STIPO Identity domain and positive affect. The STIPO Reality Testing domain also was revealed to be a significant predictor of negative affect.

Aggression. Following the data-analytic model described previously, we first examined the bivariate correlations between each measure of aggression and the three STIPO domains, all of which were statistically significant at or below the $p < .05$ level (see Table 3). We next entered the three STIPO domains simultaneously as a second step into the multiple regression model described previously and executed this model for each of the four measures of aggression. The overall model resulted in significant increments in variance explained by the STIPO domains over and above the demographic variables entered as a first step, ranging from 21% (SNAP Disinhibition and SNAP Aggression; $p < .001$) to 42% (Buss–Durkee Irritability; $p < .001$) for each of the four measures. For three of the four measures (IPO Aggression, SNAP Aggression, and SNAP Disinhibition), the STIPO Primitive Defenses domain was found to uniquely contribute significant portions of variance explained. Neither the Identity nor the Reality Testing scales uniquely contributed to the prediction of measures of aggression.

Cognitive impairment. Bivariate correlations between each of the STIPO domains and the two measures of cognitive impairment were positive in valence and statistically significant at or below the $p < .05$ level (see Table 3). When the STIPO domains were tested in the multiple regression model specified previously, a statistically significant increment in variance

⁴Note that the correlations among the independent and dependent variables in Table 3 are presented simply for descriptive purposes and do not represent a series or family of a priori tests.

explained over and above the demographic variables was observed: 24% for the IPO Reality Testing scale ($p < .001$) and 12% for the SNAP Eccentric Perceptions scale ($p < .01$).

Personality disorder traits. Given Kernberg's (1984; Kernberg & Caligor, 2005) hypothesis that most of the personality disorders described in the *DSM* system (American Psychiatric Association, 1987, 1994) are characterized by pathology of identity and the extensive use of primitive defenses, it would be important to test the association between the STIPO domains and measures of *DSM* personality disorder traits as well as the types of interpersonal difficulties characteristic of the personality disorders. To aggregate the personality disorder trait data derived from the SNAP into meaningful composites, we computed scores from the personality disorder trait indexes that correspond to the Cluster A (odd, eccentric), Cluster B (erratic, impulsive), and Cluster C (anxious, avoidant) personality disorders. For this study, these composites are the simple sums of the traits for each of the personality disorders in Clusters A, B, and C. We added traits that appear under multiple diagnoses to a summary index only once. In addition to the personality disorder (PD) cluster composites, which served as dependent variables in our regression analyses, we also conducted regression analyses using as a dependent measure the total score from the IIP-PD, a measure of interpersonal dysfunction that has been shown to be highly predictive of personality disorder diagnoses (Pilkonis et al., 1996; Stern et al., 2000). As expected, a review of bivariate correlations between each of the three STIPO domains and each of the four personality disorder trait indexes (three SNAP-based Cluster scores and the IIP-PD total score) revealed statistically significant associations in all cases (see Table 3). We carried out the regression model specified previously, which involved the simultaneous entry of the three STIPO domains as predictor variables after controlling for demographic variables, for each of the personality disorder clusters and the IIP-PD total score to determine the unique contribution of the three STIPO domains to prediction of each of these validity measures. These regressions resulted in a significant increment in variance explained after controlling for demographic variables for each of the three personality disorder cluster indexes: 29% for Cluster A ($p < .001$), 23% for Cluster B ($p < .001$), and 17% for Cluster C ($p < .001$). A significant increment of 42% of variance was also explained by the STIPO scales in the IIP-PD scores after controlling for demographic variables ($p < .001$). We observed a varied pattern of unique contribution by the STIPO domains to the prediction of the personality disorder traits, with the STIPO Identity and Reality Testing domains each uniquely predicting variance in Cluster A, the Primitive Defenses domain uniquely contributing to the explanation of variance in Cluster B, and the Reality Testing domain uniquely predicting variance in Cluster C.

DISCUSSION

Overview

The objective of this study was to conduct an initial evaluation of the basic psychometric properties (reliability and validity) of the STIPO, a semistructured interview designed to assess a psychoanalytically oriented model of personality health and disorder. Although best viewed in the spirit of exploration and discovery rather than justification and confirmation, results of

this investigation, conducted in a clinical sample representing a broad range of personality pathology, are encouraging and suggest that the STIPO can be administered in a reliable fashion and that its component scales evidence a degree of convergent and discriminant validity.

Reliability

The STIPO Identity and Primitive Defenses scales demonstrated adequate internal consistency, and the considerably shorter Reality Testing scale (four items) approached conventional standards of internal consistency. The relatively high internal consistency (Cronbach's alpha) values for the Identity and Primitive Defenses domains, in conjunction with the item-total correlations, suggest that the component domains of the STIPO tap relatively homogenous content. All three scales proved to be adequate by contemporary psychometric standards of inter-rater reliability (Nunnally & Bernstein, 1994), supporting the integrity of the STIPO scoring system, that is, the scores of multiple raters rating the same interview converge. In an ongoing study, we are examining the short- and intermediate-term test-retest reliability of the STIPO domains.

Initial Criterion-Related Validity

Overall, and broadly consistent with the findings of the initial IPO validation study (Lenzenweger et al., 2001), our data suggest that the STIPO Identity and Primitive Defenses domains explain variation in constructs closely related to personality disorders, including personality disorder traits themselves. In terms of criterion-related validity, the overall regression model testing the relative contribution of the three STIPO domains to the prediction of affect, aggression, cognitive disturbance, and personality disorder trait scores suggests that the component dimensions of the STIPO are related to the various constructs in different ways; said differently, despite significant STIPO domain intercorrelations, our regression model is suggestive of discriminant validity. For example, whereas the assessment of one's representation of self and significant others (Identity) appears to be strongly related to measures of both positive and negative affect, the use of primitive defenses appears to be related to measures of aggression and the interpersonal dysfunction characteristic of the Cluster B personality disorders. These results suggest greater specificity and nuance in the pattern of intercorrelation than we had predicted.

Central to Kernberg's (1984) view of personality pathology is the notion that increased negative affect and decreased positive affect are experienced in a generalized state of behavioral dysregulation and instability. In light of Kernberg's theory, it makes sense that the use of primitive defenses, strategies that in many cases involve the control of others into conformity with a distorted and rigid internal model of self and other, is associated with measures of aggression and dyscontrol and Axis II traits associated with Cluster B, both of which assess the interpersonal distress and volatility characteristic of the BPO-spectrum personality disorders.

It may also be the case that because the Primitive Defenses domain (and not the Identity domain) contains items reflecting the tendency of such strategies to provoke intense feelings of frustration and anger in others, that these items are more directly tracked by measures of aggressive dyscontrol and the types of overt behavioral difficulties catalogued in the Cluster

B of personality disorders (here through the SNAP-derived personality disorder Cluster B index). The failure to link Primitive Defenses to personality disorder traits associated with Cluster C—which includes disorders characterized by a manifestly less aggressive, more passive and anxious quality—is consistent with Kernberg's theory (1984; Kernberg & Caligor, 2005) in which several of the Cluster C disorders are viewed as higher level personality disorders outside the realm of BPO.

The contribution of the STIPO Reality Testing domain to the prediction of variance in our validity measures is inconsistent in this study. The Reality Testing domain was associated in this study with elevated personality disorder traits from Clusters A and C. This finding suggests that impairments in the social aspects of reality testing are associated with the eccentricity and social deficits of individuals characterized by the Cluster A disorders and the anxieties characteristic of the avoidant and dependent PD features that make up Cluster C. The lack of a significant association between the STIPO Reality Testing subscale and both the SNAP Eccentric Perceptions and IPO Reality Testing scales is likely due to the fact that the STIPO Reality Testing scale focuses entirely on the impairments in social reality testing and attention to social cues, whereas the SNAP and IPO scales include items related to frank reality testing impairments such as visual or auditory hallucinations and referential thinking. It thus appears, based on these preliminary data, that the STIPO Reality Testing scale does not assesses a construct akin to the more glaring, positive-symptom-like, subclinical psychotic experiences.

Study Limitations

Several features of this study's design are noteworthy, as they may limit somewhat the generalizability of our findings. First, with the exception of the STIPO, we employed exclusively self-report questionnaires in this study. The construct validity of the STIPO would be advanced significantly by demonstrating associations with personality disorder assessment methods other than self-reports. The use of "best estimate" diagnostic procedures such as the longitudinal, expert, all data methodology in which multiple sources of information are pooled across sources in making Axis II diagnoses (Pilkonis et al., 1995; Pilkonis, Heape, Ruddy, & Serrao, 1991) and of clinician-rated assessments of the personality features and clinical themes characteristic of the personality disorders (Shedler, Westen Assessment Procedure; Westen & Shedler, 1999a, 1999b) would be particularly useful in testing the ability of the STIPO to predict features of Axis II and in determining the extent to which variance due to shared assessment method influences these results.

In terms of reliability, further clarification of the structure of the STIPO is clearly warranted. Exploratory factor analysis performed on the initial IPO validation data set revealed a two-factor structure, with Identity and Primitive Defenses highly intercorrelated and loading onto the same factor (Lenzenweger et al., 2001). Although neither exploratory nor confirmatory factor analytic procedures could be employed in this study given the limited sample size, such procedures are clearly indicated given the high intercorrelation between the STIPO Identity and Primitive Defenses scales. Additionally, interrater reliability in this study was calculated based on a limited subgroup of the hospital sample, a subgroup whose scores on the three STIPO scales were uniformly higher on average across the three STIPO

domains than others in the sample (from both the hospital and psychoanalytic clinic sample); replication of the interrater reliability findings with a more robust and diverse sampling is thus indicated.

This study did not account for the presence of Axis I pathology. Although reports on the influence of Axis I pathology on the stability of Axis II assessment are variable, an examination of the mutual influence of Axis I pathology and the dimensions assessed in the STIPO would be important. Interviewers for this study were all highly trained, experienced in the assessment of severe personality disorders, and well-versed in psychoanalytic theory; replication of these results with interviewers at various levels of training and with those unfamiliar with Kernberg's (1984) theory of personality, would suggest that our findings are not influenced by the theory guiding our research and that the STIPO could be more broadly exported in the research community. Finally, test-retest reliability of STIPO domains, which assess features of personality that are presumed to be stable in the short term and longer term, is currently being investigated in a separate study directed by author B. L. Stern.

Conclusion and Clinical Relevance

The purpose of this study was to test the preliminary reliability and initial construct validity of a semistructured interview designed to assess Kernberg's theory of personality health and disorder. In terms of reliability, our goal was to show that this elaborate theory can be clearly structured in an interview and that agreement among skilled raters can be obtained; strong internal consistency for the STIPO Identity and Primitive Defenses scales was indeed observed, and sound interrater reliability was obtained for all three STIPO scales. In terms of preliminary construct validity, we hoped to demonstrate relations between the STIPO Identity, Primitive Defenses, and Reality Testing domains and constructs closely linked to personality disorder, that is, affect, aggression, and personality disorder trait indexes. Despite significant correlations among the STIPO domains, a pattern of results consistent with discriminant validity emerged in this preliminary study, with the STIPO Identity domain linked to measures of affect and the STIPO Primitive Defenses domain linked both to measures of aggression and Axis II Cluster B traits. Insofar as this level of discriminant validity was not predicted, and is inconsistent with prior research involving the IPO (Lenzenweger et al., 2001), the meaning and strength of these findings will be further elaborated only through replication of these results in other studies with diverse samples.

Given that Kernberg's theoretical and clinical contributions are widely cited in the personality disorder literature, as well as the prevalence of personality pathology in the population (i.e., approximately 10%; Lenzenweger, 2008), the STIPO may be of interest to researchers interested in the empirical relation between psychoanalytically informed constructs, contemporary trait models of normal and disordered personality and their neurobehavioral underpinnings, and current personality disorder nosology. The beginnings of work to map these relationships appear in these results and in the results of earlier work involving a questionnaire method for the assessment of the same model of personality (Lenzenweger et al., 2001). These assessment findings are amplified by recent data from a randomized clinical trial of various treatments for patients with borderline personality disorder, which demonstrated

the effectiveness of Transference-Focused Psychotherapy (Caligor, Clarkin, & Kernberg, 2007; Clarkin, Yeomans, & Kernberg, 2006), a manualized psychodynamic psychotherapy deriving from Kernberg's theory of personality organization, in treating borderline pathology (Clarkin, Levy, Lenzenweger, & Kernberg, 2004, 2007). Thus, although no single study can establish the construct validity of a proposed theoretical framework (Cronbach & Meehl, 1955), preliminary assessment and treatment-outcome data across several studies have constituted building blocks in the growing corpus of empirical work that in combination support the initial construct validation of Kernberg's (1984) theory of personality organization.

ACKNOWLEDGMENTS

This study was supported in part by grants from the International Psychoanalytic Association to principal investigator B. L. Stern and E. Caligor and a grant from the American Psychoanalytic Association to E. Caligor. We thank Jill Delaney and Jorge Cassab for their contributions to this work and Steven Roose, Andrew Skodol, and Paul Pilkonis for their reviews of an earlier version of this manuscript.

REFERENCES

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Benjamin, L. S. (2005). Interpersonal theory of personality disorders: The structural analysis of social behavior and interpersonal reconstructive therapy. In J. F. Clarkin & M. F. Lenzenweger (Eds.), *Major theories of personality disorder* (2nd ed., pp. 157–230). New York: Guilford.
- Buss, A., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting and Clinical Psychology, 21*, 343–348.
- Caligor, E., Clarkin, J. F., & Kernberg, O. F. (2007). *Handbook of dynamic psychotherapy for higher level personality pathology*. Washington, DC: American Psychiatric Publishing.
- Clark, L. A. (1993). *Schedule for Nonadaptive and Adaptive Personality (SNAP)*. Minneapolis: University of Minnesota Press.
- Clarkin, J. F., Foelsch, P. A., & Kernberg, O. F. (2001). *The Inventory of Personality Organization*. White Plains, NY: The Personality Disorders Institute, Weill College of Medicine of Cornell University.
- Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2004). The Personality Disorders Institute/Borderline Personality Disorder Research Foundation randomized controlled trial for borderline personality disorder: Rationale, methods, and patient characteristics. *Journal of Personality Disorders, 18*, 52–72.
- Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry, 164*, 922–928.
- Clarkin, J. F., Yeomans, F., & Kernberg, O. F. (2006). *Psychotherapy of borderline personality: Focusing on object relations*. Washington, DC: American Psychiatric Publishing.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
- Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Depue, R. A., & Lenzenweger, M. F. (2001). A neurobehavioral dimensional model of personality disorders. In W. J. Livesley (Ed.), *The handbook of personality disorders* (pp. 136–176). New York: Guilford.
- Depue, R. A., & Lenzenweger, M. F. (2005). A neurobehavioral dimensional model of personality disturbance. In J. F. Clarkin & M. F. Lenzenweger (Eds.), *Major theories of personality disorder* (2nd ed., pp. 391–453). New York: Guilford.
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis II Disorders, Patient Edition (SCID-II)*. Washington, DC: American Psychiatric Press.
- Foelsch, P. A., Clarkin, J. F., Kernberg, O. F., Somavia, J., Nomandin, L., & Lenzenweger, M. F. (2000). *The Inventory of Personality Organization: Initial Psychometric Properties*. Unpublished manuscript, available from the Personality Disorders Institute, Weill Medical College of Cornell University, White Plains, New York.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.
- Kernberg, O. F. (1984). *Severe personality disorders*. New Haven, CT: Yale University Press.
- Kernberg, O. F., & Caligor, E. (2005). A psychoanalytic theory of personality disorders. In J. F. Clarkin & M. F. Lenzenweger (Eds.), *Major theories of personality disorder* (2nd ed., pp. 114–156). New York: Guilford.
- Lenzenweger, M. F. (2008). Epidemiology of personality disorders. *Psychiatric Clinics of North America, 31*, 395–403.
- Lenzenweger, M. F., Clarkin, J. F., Fertuck, E. A., & Kernberg, O. F. (2004). Executive neurocognitive functioning and neurobehavioral systems indicators in borderline personality disorder: A preliminary study. *Journal of Personality Disorders, 18*, 421–438.
- Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The Inventory of Personality Organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment, 13*, 577–591.
- Loranger, A. W. (1999). *International Personality Disorder Examination (IPDE) manual*. Odessa, FL: Psychological Assessment Resources.
- Nunnally, J., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- 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 interviews for personality disorders. *Archives of General Psychiatry, 52*, 1025–1033.
- Pilkonis, P. A., Heape, C. L., Ruddy, J., & Serrao, P. (1991). Validity in the diagnosis of personality disorders: The use of the LEAD standard. *Psychological Assessment, 3*, 46–54.
- Pilkonis, P. A., Kim, Y., Proietti, J. M., & Barkham, M. (1996). Scales for personality disorders developed from the Inventory of Interpersonal Problems. *Journal of Personality Disorders, 10*, 355–369.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
- Stern, B. L., Kim, Y., Trull, T. J., Scarpa, A., & Pilkonis, P. A. (2000). Inventory of Interpersonal Problems Personality Disorder Scales: Operating characteristics and confirmatory factor analysis in nonclinical samples. *Journal of Personality Assessment, 74*, 459–471.
- Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry, 156*, 258–272.
- Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part II: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry, 156*, 273–285.