

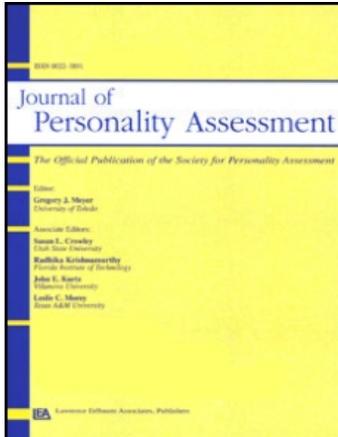
This article was downloaded by: [Columbia University]

On: 11 March 2010

Access details: Access Details: [subscription number 918552111]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Personality Assessment

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775653663>

Inventory of Interpersonal Problems Personality Disorder Scales: Operating Characteristics and Confirmatory Factor Analysis in Nonclinical Samples

Barry L. Stern ^a; Yookyung Kim ^b; Timothy J. Trull ^c; Angela Scarpa ^d; Paul Pilkonis ^b

^a New York Presbyterian Hospital-Cornell Medical Center. ^b Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine. ^c Department of Psychology, University of Missouri-Columbia. ^d Department of Psychology, Virginia Polytechnic Institute and State University.

To cite this Article Stern, Barry L., Kim, Yookyung, Trull, Timothy J., Scarpa, Angela and Pilkonis, Paul(2000) 'Inventory of Interpersonal Problems Personality Disorder Scales: Operating Characteristics and Confirmatory Factor Analysis in Nonclinical Samples', *Journal of Personality Assessment*, 74: 3, 459 – 471

To link to this Article: DOI: 10.1207/S15327752JPA7403_9

URL: http://dx.doi.org/10.1207/S15327752JPA7403_9

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Inventory of Interpersonal Problems Personality Disorder Scales: Operating Characteristics and Confirmatory Factor Analysis in Nonclinical Samples

Barry L. Stern

New York Presbyterian Hospital–Cornell Medical Center

Yookyung Kim

*Western Psychiatric Institute and Clinic
University of Pittsburgh School of Medicine*

Timothy J. Trull

*Department of Psychology
University of Missouri–Columbia*

Angela Scarpa

*Department of Psychology
Virginia Polytechnic Institute and State University*

Paul Pilkonis

*Western Psychiatric Institute and Clinic
University of Pittsburgh School of Medicine*

Research involving clinical samples has demonstrated the utility of a 28-item personality disorder (PD) screening measure (Inventory of Interpersonal Problems–Personality Disorder scale [IIP–PD]) culled from the IIP in the prediction of the presence or absence of a PD (Pilkonis, Kim, Proietti, & Barkham, 1996). This article extends these diagnostic efficiency findings to nonclinical samples and presents additional data regarding the factor structure of the 28 IIP–PD items. Diagnostic efficiency statistics for the IIP–PD scale, calculated using both interview and self-report methods, support the utility of the IIP–PD scale as a screening tool for the presence or absence

of a PD. High specificity estimates indicate that individuals who do not exceed *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) symptom thresholds rarely exceed the IIP-PD cutoff. Furthermore, a high negative predictive power (NPP) estimate derived using an interview-based diagnostic standard suggests that the IIP-PD scale accurately screens out individuals who do not have a PD. Finally, cross-validated confirmatory factor-analytic results involving items composing the 5 IIP PD subscales identified in previous research (Kim, Pilkonis, & Barkham, 1997) suggest that a measurement model with a single second-order factor (general PD) and 5 first-order factors (one representing each PD subscale) provided the best fit to the observed data compared to 2 other competing models.

Current research suggests that the prevalence rate for personality disorders (PDs) in adults is between 8 and 13% (Loranger, 1997). Other research has demonstrated the considerable psychosocial sequelae of PDs, including increased risk for criminality (Dowson, 1995), suicide (Paris, 1993; Stone, 1990), and alcohol and drug abuse (Goldman, D'Angelo, & DeMaso, 1993). Furthermore, people with PDs show increased medical utilization for conditions without medical cause (Dowson, 1995; Lillienfeld, Van Volkenburg, Larntz, & Akiskal, 1986), increased occupational or social role difficulties (Links, Mitton, & Steiner, 1990; Stone, 1990) and, by definition, chronic maladjustment in their interpersonal relationships. Finally, the complicating effects of PD diagnoses on the treatment of a variety of Axis I conditions is also well documented (Reich & Vasile, 1993). All of the previous underscore the need for brief screening tools that can accurately assess the presence or absence of a PD so that individuals who are either at risk for a PD or who have an undiagnosed PD can be identified. Given the time pressures associated with much of contemporary psychotherapy, a brief and accurate screening tool for PDs would allow relevant interpersonal and skill-based factors (e.g., distress tolerance or self-regulation skills) to be incorporated into psychotherapeutic treatment in a timely fashion. Similarly, PD researchers would also benefit from a brief Axis II screening tool with strong operating characteristics because they could then focus more comprehensive and expensive interview and assessment efforts on those individuals most likely to meet *Diagnostic and Statistical Manual of Mental Disorders* (DSM) criteria for PDs.

Given the importance of identifying samples in which PDs are likely to be present, researchers have begun to evaluate the effectiveness of self-report "Stage 1" screening tools in identifying individuals with PD diagnoses. Pilkonis et al. (1996) analyzed the 127 items of the Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureno, & Villaseñor, 1988) and developed a brief, 28-item screening index for the presence or absence of a PD (IIP-PD). Derived from a qualitative analysis of the intake interviews of psychiatric outpatients, the IIP provides an index of chronic difficulties in the respondent's interpersonal relation-

ships. Factor analyses of the 28 items composing the IIP-PD screener yielded three subscales that were found to be useful in distinguishing individuals with any PD diagnosis from those without a PD diagnosis: PD1, Interpersonal Sensitivity; PD2, Interpersonal Ambivalence; and PD3, Aggression. Nineteen additional IIP items were identified as being useful in discriminating between individuals who met criteria for a Cluster C diagnosis and those who did not. Factor analyses of these 19 items yielded two subscales: C1, Need for Social Approval, and C2, Lack of Sociability.

Results of this initial validation study (Pilkonis et al., 1996), which was conducted with a sample of psychiatric patients, suggest that a scale composed of the 28 items found on the three PD subscales (hereafter called the *IIP-PD scale*) has good operating characteristics, in particular good positive predictive power (PPP; $> .85$) and high sensitivity (ranging from .71 to .91 depending on the IIP-PD scale cutoff score), in distinguishing between individuals with or without a PD diagnosis. The initial study also found that the two C subscales had low PPP, due primarily to an inability to distinguish between individuals with Cluster B and Cluster C diagnoses.

In a later study of the factor structure of the 47 PD scale items, Kim et al. (1997) examined the relations among the five PD subscales with data collected from five different clinical samples using confirmatory factor analysis (CFA). Among several competing models, the best-fitting model was hierarchical, with one superordinate general PD factor and five first-order factors, one representing each of the five PD subscales.

Although initial findings regarding the PD subscales are promising, the majority of the findings reported to date have involved clinical samples, and no diagnostic efficiency or CFA data have been reported in nonclinical samples. A screening tool with the ability to accurately predict the presence or absence of personality pathology in nonclinical samples, for which prediction would be more difficult given the lower base rates of the various PDs, would be of considerable benefit to PD researchers and diagnosticians. For researchers with access to large nonclinical samples, for example, an accurate screening measure could be used to identify individuals who are likely to meet PD diagnoses and who should thus be selected for more time- and cost-intensive evaluation.

Initial findings in nonclinical, college undergraduate samples (Scarpa et al., 1999) suggest that the five PD subscales have high internal consistency and test-retest reliability and that the factor structures of the 28-item IIP-PD scale and the 19-item Cluster C scale are virtually identical to those found in Pilkonis et al.'s (1996) and Kim et al.'s (1997) studies. Results of Scarpa et al.'s (1999) study also indicate that high IIP-PD scorers reported significantly higher levels of emotional and behavioral problems and higher levels of Axis II symptomatology (based on structured clinical interview results) than low IIP-PD scorers.

In this study, we further examined the validity properties of the five PD subscales, and the 28-item IIP-PD scale in particular, in nonclinical samples. Oper-

ating characteristics of the IIP–PD scale in two independent nonclinical samples at the University of Missouri are presented. These data evaluate the efficiency of the IIP–PD scale in predicting which individuals either meet Axis II diagnostic criteria (as determined by diagnostic interview) or exceed an Axis II symptom count threshold (as determined by a self-report of symptom presence or absence). In addition, CFA data, aggregated from these two samples and cross-validated with two additional nonclinical samples, are presented. CFA permits the construction of alternative a priori models that can be tested empirically (Jöreskog & Sörbom, 1988). We tested the three best-fitting measurement models identified by Kim et al. (1997) as alternative structures to account for relations among the five PD subscales and the items they include. The first model was a hierarchical model in which the five first-order factors (Interpersonal Sensitivity, Interpersonal Ambivalence, Aggression, Need for Social Approval, and Lack of Sociability) were specified. The model further required that these five factors load on a single second-order factor interpretable as a latent construct representing general PD. The second model specified that the five first-order factors be intercorrelated with no second-order structure. The third model specified two second-order factors that were hypothesized to be correlated: one for externalizing PDs (including subscales for Interpersonal Sensitivity, Interpersonal Ambivalence, and Aggression) and a second for internalizing PDs (including subscales for Need for Social Approval and Lack of Sociability).

METHOD

Samples

Four separate IIP databases ($N = 921$ nonclinical university students) were obtained from samples in the United States for the purpose of the CFAs. Four investigators from three cities contributed data. A sample of 145 students enrolled at the University of Pittsburgh Dental School was obtained through the Western Psychiatric Institute and Clinic (M age = 24.12; % female = 38.6), and an additional 461 students were sampled from Eastern Washington University (M age = 21.05; % female = 60.7). Two independent samples were recruited at the University of Missouri–Columbia. The first of these involved 90 students who participated in a 2-year follow-up study of students with features of borderline personality disorder (BPD; M age = 19.02; % female = 56.6). A second sample of 225 students enrolled in an introductory psychology course at the University of Missouri–Columbia also were surveyed (M age = 21.21; % female = 79). There was a small representation of ethnic minority students in each sample; the percentage of White participants was between approximately 83 and 92% for most of the samples.

Data Analyses

To investigate the psychometric properties of the five PD subscales and of the combined 28-item IIP-PD scale in particular, we conducted several analyses. We computed coefficient alpha to assess the internal consistency of all five PD subscales and the IIP-PD scale. For those samples for which Axis II symptom or diagnostic data were available (the two Missouri samples), we conducted three sets of analyses. First, we conducted *t* tests to compare mean PD subscale scores, including the IIP-PD scale, across individuals who either exceeded or failed to exceed the symptom threshold for at least one PD diagnosis; second, we computed correlations between PD subscale scores, including the IIP-PD scale, and symptom counts for each of the *DSM* PDs; and third, we computed operating characteristics for the IIP-PD scale (predicting the presence or absence of any PD). Finally, we conducted the CFA described earlier on the Missouri samples (combined $n = 315$) and cross-validated it with the Pittsburgh and Eastern Washington University samples (combined $n = 606$).

As noted previously, Axis II diagnostic data were available only for individuals in the two Missouri samples. PD diagnoses were derived for the sample of students in the BPD follow-up study ($n = 90$) through the administration of the Structured Interview for *DSM-III-R* Personality (SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989), a structured interview for the *DSM-III-R* (3rd ed., rev.; American Psychiatric Association, 1987) PDs. PD symptoms were assessed by means of self-report in the introductory psychology sample ($n = 225$) through the administration of the Personality Diagnostic Questionnaire-IV (PDQ-IV; Hyler, 1994), a direct translation of *DSM-IV* diagnostic criteria into a *true-false* response format. Considerable research involving earlier generations of the PDQ, as well as recent research involving the PDQ-IV, has demonstrated statistically significant but low agreement with interview measures of PDs as well as a tendency to overdiagnose PDs (Fossati et al., 1998; Hyler, Skodol, Oldham, Kellman, & Doidge, 1992; Yeung, Lyons, Wateraux, Faraone, & Tsuang, 1993).

RESULTS

Descriptive Statistics

On the basis of the SIDP-R interviews conducted in the BPD follow-up sample ($n = 90$), 15.6% of the sample met criteria for at least one PD. The most frequently occurring disorders in this sample were the passive-aggressive (6.7%), paranoid (5.6%), and avoidant (5.6%) PDs. On the basis of a self-report of PD symptoms using the PDQ-IV in the introductory psychology sample ($n = 225$), 68.9% of participants in the introductory psychology sample exceeded the *DSM-IV* symptom

threshold for at least one PD, with the most frequently occurring disorders in this sample being the obsessive–compulsive (54.5 %), paranoid (24.7 %), and avoidant (25.1 %) PDs.

Internal consistency estimates were high for all five PD subscales in each of the four samples. Averaged across all four sites, coefficient alphas for each of the five PD subscales ranged from .84 (Interpersonal Sensitivity) to .89 (Lack of Sociability). Internal consistency estimates for the 28-item IIP–PD scale were also high, ranging from .91 to .93 across the four sites (mean α for the IIP–PD scale = .91).

Mean PD subscale scores for individuals with and without at least one PD (as determined by the SIDP–R structured interview) and for individuals who exceeded or failed to exceed the *DSM* symptom threshold for at least one PD (as determined by the self-report PDQ–IV), are presented in Table 1. In each sample, mean scores for each subscale, including the IIP–PD scale, were significantly higher for individuals who exceeded the *DSM* symptom threshold for at least one PD. Zero-order correlations between the PD subscales, including the IIP–PD scale, and symptom counts for the *DSM* PDs (presented in Table 2) revealed significant correlations at or above the .01 level for at least two of the PD subscales and each PD.

TABLE 1
Mean Scores for the IIP PD Scales (Missouri Samples)

Scale	SIDP–R Sample						PDQ–IV Sample					
	Total Sample ^a		Those With a PD ^b		Those Without a PD ^c		Total Sample ^d		Those Exceeding a PD Symptom Threshold ^e		Those Not Exceeding a PD Symptom Threshold ^f	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
PD1	1.05	0.64	1.75	0.36	0.92**	0.60	1.31	0.69	1.48	0.68	0.88**	0.51
PD2	0.62	0.49	1.11	0.46	0.53**	0.44	0.80	0.61	0.90	0.62	0.57**	0.52
PD3	0.64	0.65	1.58	0.77	0.46**	0.44	0.94	0.77	1.10	0.78	0.75**	0.72
IIP–PD	0.79	0.49	1.48	0.30	0.67**	0.41	1.03	0.56	1.34	1.30	0.69**	0.90
C1	1.25	0.77	1.87	0.79	1.14**	0.71	1.46	0.82	1.62	0.83	1.06**	0.62
C2	0.75	0.77	1.36	0.94	0.64*	0.69	0.97	0.80	1.12	0.84	0.62**	0.58

Note. IIP = Inventory of Interpersonal Problems; PD = personality disorder; SIDP–R Sample = University of Missouri BPD follow-up sample using the Structured Interview for *DSM–III–R* Personality ($N = 90$); PDQ–IV Sample = University of Missouri introductory psychology sample using the Personality Diagnostic Questionnaire–IV ($N = 225$); PD1 = Interpersonal Sensitivity; PD2 = Interpersonal Ambivalence; PD3 = Aggression; C1 = Need for Social Approval; C2 = Lack of Sociability.

^a $N = 90$. ^b $n = 14$. ^c $n = 76$. ^d $N = 223$. ^e $n = 160$. ^f $n = 63$.

* $p < .01$. ** $p < .001$.

TABLE 2
Correlations Between IIP PD Scales (Missouri Samples) and PD Symptom Counts

Sample	PD1		PD2		PD3		IIP-PD		C1		C2	
	SIDP-R	PDQ-IV										
Paranoid	.46***	.32***	.40***	.35***	.43***	.21**	.52***	.35***	.29**	.18**	.38***	.32***
Schizoid	.30**	.11	.28***	.31***	.40***	.13	.38***	.22***	.23*	.07	.47***	.36***
Schizotypal	.44**	.28***	.37***	.24***	.36**	.13	.47***	.26***	.27*	.26***	.41***	.40***
Obsessive-compulsive	.40**	.30***	.41***	.17**	.41***	.12	.49***	.25***	.30**	.24***	.39***	.12
Dependent	.46***	.52***	.13	.17**	.24*	.08	.36***	.32***	.44***	.55***	.48***	.51***
Avoidant	.58***	.55***	.37***	.27***	.40**	.14*	.56***	.42***	.48***	.60***	.68***	.63***
Histrionic	.23	.38***	.30**	.21***	.37***	.31***	.34**	.35***	.16	.17**	.14	.08
Borderline	.19	.58***	.62*	.35***	.43***	.43***	.43**	.56***	.00	.45***	.17	.46***
Narcissistic	.36***	.33***	.36***	.33***	.39***	.35***	.44***	.39***	.28	.19**	.21*	.18**
Antisocial	.13	.08	.31**	.24***	.41***	.29***	.31**	.22**	-.04	.02	-.04	.03

Note. IIP = Inventory of Interpersonal Problems; PD = personality disorder; PD1 = Interpersonal Sensitivity; PD2 = Interpersonal Ambivalence; PD3 = Aggression; C1 = Lack of Sociability; C2 = Need for Social Approval; SIDP-R = University of Missouri borderline personality disorder follow-up sample using the Structured Interview for DSM-III-R Personality ($N = 90$); PDQ-IV = University of Missouri introductory psychology sample using the Personality Diagnostic Questionnaire-IV ($N = 225$).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Diagnostic Efficiency of the 28-Item IIP-PD Scale

Diagnostic efficiency estimates for the SIDP-R and PDQ-IV samples, as well as those reproduced from the initial validation study involving the 28-item IIP-PD scale (Pilkonis et al., 1996), are presented in Table 3. In general, diagnostic efficiency estimates from the nonclinical SIDP-R sample (a lower base rate sample) revealed higher sensitivity, specificity, and NPP, and somewhat lower PPP estimates, than those derived at similar cutoff points in the validation sample (a higher base rate, clinical sample; Pilkonis et al., 1996). It thus appears that the IIP-PD scale can effectively screen out individuals who do not meet diagnostic criteria for a PD in nonclinical samples. The lower estimates for PPP are likely a function of disorder base rates across SIDP-R sample from this study and Pilkonis et al.'s (1996) validation samples.

TABLE 3
Operating Characteristics of the 28-Item IIP-PD Scale Predicting
Presence or Absence of a Personality Disorder

<i>Cutoff and Sample</i>	<i>EFF</i>	<i>SENS</i>	<i>SPEC</i>	<i>PPP</i>	<i>NPP</i>
0.70					
PDQ-IV	0.70	0.78	0.52	0.79	0.49
SIDP-R	0.70	1.00	0.64	0.34	1.00
Pilkonis (1996)	0.81	0.91	0.40	0.85	0.55
0.90					
PDQ-IV	0.65	0.61	0.76	0.86	0.45
SIDP-R	0.83	1.00	0.80	0.48	1.00
Pilkonis (1996)	0.74	0.81	0.47	0.85	0.39
1.10					
PDQ-IV	0.62	0.52	0.85	0.89	0.43
SIDP-R	0.89	0.93	0.87	0.57	0.99
Pilkonis (1996)	0.70	0.71	0.67	0.89	0.38
1.30					
PDQ-IV	0.53	0.38	0.89	0.89	0.38
SIDP-R	0.89	0.64	0.93	0.64	0.93
Pilkonis (1996)	—	—	—	—	—
1.50					
PDQ-IV	0.46	0.26	0.92	0.89	0.34
SIDP-R	0.89	0.64	0.93	0.64	0.34
Pilkonis (1996)	—	—	—	—	—

Note. IIP-PD = Inventory of Interpersonal Problems–Personality Disorder; EFF = diagnostic efficiency (i.e., percentage agreement, or hit rate); SENS = sensitivity; SPEC = specificity; PPP = positive predictive power; NPP = negative predictive power; PDQ-IV = introductory psychology sample from the University of Missouri using the Personality Diagnostic Questionnaire-IV ($N = 225$); SIDP-R = borderline personality disorder follow-up sample from the University of Missouri using the Structured Interview for *DSM-III-R* Personality ($N = 90$). Pilkonis (1996) refers to data reported in Pilkonis, Kim, Proietti, and Barkham (1996).

Diagnostic efficiency estimates from the nonclinical PDQ-IV sample revealed sensitivity estimates that are lower than those found in both the SIDP-R sample and Pilkonis et al.'s (1996) validation samples. Although such a finding suggests that the indicator in question (here, the IIP-PD scale) may not accurately identify those individuals meeting diagnostic criteria, these data are perhaps better interpreted as evidence that the PDQ-IV is poorly suited to the task of assigning Axis II diagnoses. Said differently, instead of suggesting that the IIP-PD scale misses too many individuals who are truly diagnostic for a PD, an alternative explanation is that the number of individuals exceeding *DSM* symptom thresholds for diagnosis when assessed by the PDQ-IV is artificially inflated because of biases associated with the self-assessment of PD symptoms (for a review of this issue, see Zimmerman, 1994; for a quantitative analysis of PDQ-IV validity, see Fossati et al., 1998). The high estimates of PPP derived from the PDQ-IV sample are also likely a function of the inflated diagnostic base rate in this sample.

Consistent with results from this study's SIDP-R sample, higher specificity estimates in the PDQ-IV sample indicate that the IIP-PD scale cutoff, when it is in the 1.1-to-1.3 range, is rarely exceeded by individuals who are below the *DSM* symptom threshold for a PD diagnosis as assessed by the PDQ-IV.

Confirmatory Factor Analysis

Data analysis. To evaluate the appropriateness of the multivariate normality assumption, we examined the skewness and kurtosis of each of the measured variables. The distributions of each scale variable showed moderate deviations from univariate normality with consistent positive skew, suggesting that participants tended to endorse lower scores on each item included in the PD subscales. Given this result, assessment of overall model fit was based on the Satorra-Bentler scaled statistic ($S-B\chi^2$) and its related corrected comparative fit index (CFI*), which corrects for non-normality (Satorra & Bentler, 1988). We also evaluated the fit of each model using the ratio of $S-B\chi^2$ to its degrees of freedom ($S-B\chi^2/df$) because of the large sample size. If this ratio is less than 5, the fit between the model and the data is assumed adequate, with smaller ratios indicating a better fit (Wheaton, Muthen, Alwin, & Summers, 1977).

Cross-validation. To test the stability of the results, we estimated the fit of each of the three CFA models using two combined samples (the combined Missouri samples [$n = 315$] and the combined Pittsburgh and Eastern Washington University samples for validation purposes [$n = 606$]). Samples were combined to increase the total sample size for analyses and to obtain greater stability in parameter estimates.

TABLE 4
Measures of Model–Data Fit

Model	df	$S-B\chi^2$		$S-B\chi^2/df^a$		CFI* ^b	
		Study Sample ^c	Validation Sample ^d	Study Sample ^c	Validation Sample ^d	Study Sample ^c	Validation Sample ^d
Model 1: Single second-order factor	85	58.65	116.32	0.69	1.37	1.00	0.99
Model 2: Five correlated first-order factors	80	297.04	595.22	3.71	7.44	0.91	0.92
Model 3: Two correlated second-order factors	84	120.37	156.02	1.43	1.86	0.99	0.98

^a $S-B\chi^2/df = S-B\chi^2/\text{degrees of freedom ratio}$. ^bCorrected comparative fit index based on $S-B\chi^2$ fit for the null model. ^c $n = 315$. ^d $n = 606$.

Model–data fit. Recall that the three best-fitting models of the seven tested by Kim et al. (1997) as alternative structures to account for relations among the IIP–PD subscales and the items they include were tested in this study. The measures of fit for the three models tested (i.e., $S-B\chi^2$ values, $S-B\chi^2/df$ ratios, and CFI*s) showed a consistent rank order: Model 1, with a single, second-order factor and five first-order factors, was the preferred model, followed by Model 3, with two correlated second-order factors, followed by Model 2, with five correlated first-order factors (see Table 4). For the three best-fitting models, all corrected comparative fit indexes exceeded .91. These results were replicated with the independent, combined validation sample.

Hierarchical model comparisons. We performed all possible hierarchical comparisons between nested models, and these direct comparisons favored Model 1 in each case ($S-B\chi^2$ for the Model 2 vs. Model 1 comparison was 238.39 for the study sample and 478.9 for the validation sample; $S-B\chi^2$ for the Model 3 vs. Model 1 comparison was 61.72 for the study sample and 39.70 for the validation sample). All $S-B\chi^2_{\text{diff}}$ statistics were statistically significant beyond $p < .001$. On the basis of these results and those presented in Table 4, it is clear that a measurement model with five first-order factors and a single second-order factor (Model 1) provided the best fit to the nonclinical sample data.

DISCUSSION

Diagnostic efficiency analyses from two nonclinical samples, using either interview or self-report methods to assess PD symptoms, support the utility of the

28-item IIP-PD scale as a screening tool for the presence or absence of a PD. Individuals who failed to exceed the *DSM* symptom threshold for at least one PD rarely scored above the IIP-PD scale cutoff score (i.e., high specificity in both the SIDP-R and PDQ-IV samples), particularly when the cutoff score was in the 1.1-to-1.3 range. Furthermore, the IIP-PD scale performed especially well in a low-base-rate sample at screening out individuals who did not exceed the *DSM* symptom threshold for at least one PD (high NPP in the SIDP-R sample). These findings should be encouraging to the PD researcher working in nonclinical settings or in settings where the base rate for PDs is relatively low; most individuals who do not meet diagnostic criteria, and thus who would be of lesser interest to the PD researcher, are very unlikely to exceed the IIP-PD scale cutoff when that cutoff is in the 1.1-to-1.3 range. In general, on the basis of the diagnostic efficiency estimates derived from these nonclinical samples, we feel that the 1.1 cutoff score provides the greatest diagnostic and clinical utility.

CFA results involving the PD subscales in these nonclinical samples indicated that a measurement model with a single second-order factor and five first-order factors provided the best fit to the observed data compared to two other competing models. These results were supported by cross-validation. Establishing a higher order structure for the PD subscales has considerable theoretical relevance, suggesting that it may be useful for some purposes to use a single latent construct to represent general personality pathology. On the basis of this psychometric evidence, the five PD subscales are distinct but related subdomains that converge on this broader, latent PD construct.

The results also shed light on several methodological considerations related to the assessment of PD symptoms. When disorder base rates are lower, as in the SIDP-R sample described previously, PPP necessarily suffers; that is, prediction of disorder status is more challenging for any indicator, such as the IIP-PD scale, given the fact that fewer people have the disorder. Given the low base rates of the disorder in such a population, one must be careful to examine other data prior to passing diagnostic judgment based on the indicator in question (here, the IIP-PD scale).

These findings also have implications for the validity of the IIP-PD scale across PD clusters. Seventy-nine percent of participants in the original validation study (Pilkonis et al., 1996) were diagnosed with a PD, 67% of whom were classified with a primary diagnosis in Cluster B. Cluster membership in both the SIDP-R sample (diagnostic base rate = 15.6%) and in the PDQ-IV sample (diagnostic base rate = 68.9%) was more diverse, with more people endorsing symptoms related to paranoid PD and disorders in Cluster C. Although one might expect a greater proportion of Cluster C disorders to be found in nonclinical samples, the implication of the findings across studies conducted to date is that the IIP-PD scale appears to be a good screening tool for PDs across all three clusters.

One limitation of this study involves the use of the PDQ-IV as an index of "caseness" when assessing the diagnostic efficiency of the IIP-PD scale. Although

a correlational–dimensional approach to the validation of the IIP–PD scale yielded promising results in this study (see Table 2), the extremely high rate of diagnostic cases in this nonclinical sample (approximately 70%) raises serious concerns about the PDQ–IV’s utility for purposes of diagnostic assessment and its validity more generally.

A second limitation of these findings is that only self-report data were used in the CFA; thus, common-method variance may account, in part, for the results. Future research involving multitrait validation designs is clearly indicated. For example, we propose to investigate similarities and discrepancies between self- and informant reports on each of the PD subscales.

An additional question of interest is the factorial invariance of the measurement model supported in these analyses with nonclinical samples. Can this structure be replicated across qualitatively different subgroups identified on the basis of sex, age, or ethnicity? To examine factorial invariance of the IIP–PD scales, CFA procedures could be used to examine differences in parameter estimates and in goodness-of-fit indexes for competing models across such subgroups.

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.
- Dowson, J. H. (1995). *Personality disorders: Recognition and clinical management*. Cambridge, England: Cambridge University Press.
- Fossati, A., Maffei, C., Bagnato, M., Donati, D., Donini, M., Fiorilli, M., Novella, L., & Ansoldi, M. (1998). Brief communication: Validity of the Personality Diagnostic Questionnaire–4+ in a mixed psychiatric sample. *Journal of Personality Disorders, 12*, 172–178.
- Goldman, S. J., D’Angelo, E. J., & DeMaso, D. R. (1993). Psychopathology in the families of children and adolescents with borderline personality disorder. *American Journal of Psychiatry, 150*, 1832–1835.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. A. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.
- Hyder, S. E. (1994). *Personality Diagnostic Questionnaire–IV (PDQ–IV)*. New York: New York State Psychiatric Institute.
- Hyder, S. E., Skodol, A. E., Oldham, J. M., Kellman, H. D., & Doidge, N. (1992). Validity of the Personality Diagnostic Questionnaire–Revised: A replication in an outpatient sample. *Comprehensive Psychiatry, 33*, 73–77.
- Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS.
- Kim, Y., Pilkonis, P. A., & Barkham, M. (1997). Confirmatory factor analysis of the personality disorder subscales from the Inventory of Interpersonal Problems. *Journal of Personality Assessment, 69*, 284–296.

- Lillienfeld, S., Van Volkenburg, C., Larntz, K., & Akiskal, H. (1986). The relationships of histrionic personality disorder to antisocial personality and somatization disorder. *American Journal of Psychiatry*, *143*, 718–722.
- Links, P. S., Mitton, J. E., & Steiner, M. (1990). Predicting outcome for borderline personality disorder. *Comprehensive Psychiatry*, *31*, 490–498.
- Loranger, A. W. (1997). *Assessment and diagnosis of personality disorders: The International Personality Disorder Examination (IPDE)*. Cambridge, England: Cambridge University Press.
- Paris, J. (1993). Management of acute and chronic suicidality in patients with borderline personality disorder. In J. Paris (Ed.), *Borderline personality disorder* (pp. 373–384). Washington, DC: American Psychiatric Press.
- Pfohl, B., Blum, N., Zimmerman, M., & Stangl, D. (1989). Structured Interview for *DSM-III-R* Personality (SIDP-R). Iowa City: University of Iowa, Department of Psychiatry.
- 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.
- Reich, J. H., & Vasile, R. G. (1993). Effect of personality disorders on the treatment outcome of Axis I conditions: An update. *Journal of Nervous and Mental Disease*, *181*, 475–484.
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *American Statistical Association 1988 Proceedings of the Business and Economic Sections* (pp. 308–313). Alexandria, VA: American Statistical Association.
- Scarpa, A., Luscher, K. A., Kadee, J., Smalley, M. S., Pilkonis, P. A., Kim, Y., & Williams, W. C. (1999). Screening for personality disorders in a nonclinical population. *Journal of Personality Disorders*, *13*, 345–360.
- Stone, M. H. (1990). *The fate of borderline patients*. New York: Guilford.
- Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (pp. 84–136). San Francisco: Jossey-Bass.
- Yeung, A. S., Lyons, M. J., Waternaux, C. M., Faraone, S. V., & Tsuang, M. T. (1993). Empirical determination of thresholds for case identification: Validation of the Personality Diagnostic Questionnaire-Revised. *Comprehensive Psychiatry*, *34*, 384–391.
- Zimmerman, M. (1994). Diagnosing personality disorders: A review of issues and research methods. *Archives of General Psychiatry*, *51*, 225–245.

Barry L. Stern
 New York Presbyterian Hospital–Cornell Medical Center
 Macy Villa
 21 Bloomingdale Road
 White Plains, NY 10605

Received April 21, 1999
 Revised November 3, 1999