Journal of Personality Disorders, 7(4), 285-303, 1993

THE WISCONSIN PERSONALITY DISORDERS INVENTORY: DEVELOPMENT, RELIABILITY, AND VALIDITY

Marjorie H. Klein, PhD, Lorna Smith Benjamin, PhD, Rochelle Rosenfeld, MA, Catherine Treece, PhD, Jack Husted, MA, and John H. Greist, MD

The Wisconsin Personality Disorders Inventory (WISPI) is a self-report questionnaire derived from an interpersonal perspective on the DSM-III-R personality disorders. Internal consistency for 11 personality disorder scales was very high in a sample of 1230 normals and patients. Two-week test–retest reliability in 80 patients and nonpatients was also high. Interscale correlations were higher than desirable but were reduced by corrections for response bias. A clinician sort of WISPI items to DSM-III criteria and independent coding of items for interpersonal content indicated good validity vis-a-vis both frames of reference. Patients scored higher than nonpatients on most scales, and patients with current clinical diagnoses of any personality disorder scored higher than those with no Axis II disorders. Other indicators of concurrent validity were high to moderate correlations with Personality Disorders Questionnaire, MMPI Personality Disorder Scales, and Millon Clinical Multiaxial Inventory self-reports; and with clinician ratings on a dimensional personality assessment form in a subsample of 146 outpatients.

This paper describes the development of the Wisconsin Personality Disorders Inventory (WISPI) and presents reliability and validity data. Although other self-report inventories for the personality disorders exist or are under development (Clark, 1990; Hyler et al., 1988; Millon, 1982, 1987; Morey, Waugh, & Blashfield, 1985), we believe that instruments that reflect a range of approaches to the conceptualization and assess-

From the University of Wisconsin-Madison. Lorna Benjamin is now at the University of Utah, Department of Psychology.

This work was supported by a grant from the National Institute of Mental Health (MH 32624). We especially wish to acknowledge the contributions of Drs. Stephen Wonderlich, Karla Moras, and Paul Pilkonis; and of Lisa Loder, Paula Machtinger, Mary Mussey, and Stacey Skare to various aspects of this work.

Address correspondence to Marjorie H. Klein, Department of Psychiatry, University of Wisconsin, School of Medicine, Madison, WI 53706.

ment of the personality disorders are needed, and are of the opinion that interpersonal formulations are especially relevant and appropriate (Benjamin, 1993; Klein, 1993).

The American Psyciatric Assoiation's (APA) publication of the new criteria for psychiatric disorders in DSM-III and DSM-III-R (*Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed.; 3rd ed., rev.—APA, 1980, 1987) has been accompanied by an important shift in focus in psychiatric classification toward more detailed, explicit systems with clearly defined behavioral descriptors (APA, 1987). Although this trend offers the potential for significant improvement in reliable classification, conceptual issues remain unclear. For example, although the recommendation for more objective criteria for personality disorders would seem to suggest assessment methods that directly mirror the content of DSM-III (e.g., the PDQ; Hyler et al., 1988) or employ specific behavioral checklists (Livesley, 1986), these approaches may not be sufficiently grounded in any theory of psychopathology to stand the test of time or to able to survive the rapid evolution of the diagnostic nomenclature.

Among various suggestions for a conceptual foundation for the assessment of the personality disorders, the most consistent recommendation has been for the relevance of interpersonal dimensions for Axis II (Kiesler, 1983; McLemore & Benjamin, 1979; Morey, 1985; Widiger & Frances, 1985; Widiger & Kelso, 1983; Wiggins, 1982). Although the details vary, the interpersonal approaches suggested generally incorporate dimensions of affiliation and dominance in circumplex models along the lines introduced by Leary (1957). Arguments for the special relevance of interpersonal constructs for the personality disorders emphasize the role of early family relations and social learning in the development of personality traits and vulnerabilities, and point to the individual's current intimate and social relationships as the context in which disorders become manifest. From such a perspective, it follows that psychosocial treatment and the psychotherapeutic relationship are important mechanisms for change. Thus, in addition to offering a unified conceptual basis for personality diagnosis, an interpersonal approach provides hypotheses about etiology, dynamics, and treatment.

The specific interpersonal perspective we chose to guide the development of the WISPI is Benjamin's (1974, 1984) model, Structural Analysis of Social Behavior (SASB), which has well- developed procedures for measuring and analyzing interpersonal behavior. The model characterizes interpersonal behavior according to three distinctions: (1) focus (on other, on self, or within self); (2) the dimension of affiliation; and (3) the dimension of interdependence, which is orthogonal to affiliation. It specifies predictions about behaviors that are complementary or antithetical in interpersonal relationships and identifies links between early object relations, social learning, and the development of the self-concept (e.g., Benjamin, 1987). Its construct validity, reliability, internal consistency, circumplex structure, and ability to discriminate between clinical groups and between clinical and normal samples have been demonstrated empirically (Benjamin, 1974; 1984).

Recently, Benjamin (1986, 1993) has developed formulations and descriptions of each of the personality disorder categories from this

perspective. These formulations consider the interpersonal context in which the behaviors associated with each personality disorder are hypothesized to have developed and they form the basis of predictions about interpersonal issues that are likely to arise in treatment. For example, in the case of borderline personality disorder, this theory explains how it is that experiences of being victimized in an abusive (often incestuous) family environment were central in shaping the instability and abandonment sensitivity of the adult borderline that is so apparent in treatment.

In developing the WISPI we have followed the general sequential strategy recommended by Jackson (1970) and Millon (1982). Item and scale content was derived directly from the definitions, first in DSM-III and later in DSM-III-R (APA, 1980, 1987). We also drew upon Benjamin's SASB-based formulations of the disorders as a guide, especially in instances of ambiguous or overlapping criteria. Specific steps of validation included: (1) content validation by clinician sort of WISPI items against Axis II categories and by independent ratings of the interpersonal content of each item; (2) assessment of the reliability and internal consistency of items and scales in a large sample of patients and nonpatients; (3) assessment of test–retest reliability in a separate sample of patients and nonpatients; and (4) comparison of WISPI results with other Axis II paper-and-pencil measures and clinical ratings of personality disorder in a patient subsample. Results of validation using factor analysis and structured and clinical interviews will be reported separately.

METHOD

Item Construction. In order to write WISPI items that corresponded to DSM criteria and are also grounded in interpersonal theory, we first used SASB coding procedures to develop interpersonal descriptors for each Axis II criterion in DSM-III and DSM-III-R (Benjamin, 1986, 1993). We also used Benjamin's formulations of the interpersonal dynamics and pathogenesis of each syndrome to guide item construction (see Benjamin, 1986, 1987, 1993). Both specific criterion codes and clinical formulations were then used to write the inventory items corresponding to each Axis II criterion (generally two items per criterion).

Items were worded from the perspective of a respondent in the hope that even socially undesirable attributes could be endorsed by a person with a given disorder if that item accurately captured his or her phenomenology. For example, an item that simply restates a DSM-III criterion such as: "People say I am cold and aloof" (schizoid criterion A1) would be less likely to be endorsed by people with schizoid disorders than an item such as "Because I wall myself off from others, I'm not affected by people" (WISPI item 86). In instances of overlap among disorders, items were written with specific distinctions in mind. The item for paranoid "coldness" is "When I have feelings I keep them to myself because others might use them against me" (WISPI 142); for schizoid coldness, one item is "I don't particularly care how I affect others" (WISPI 120). In sum, the interpersonal approach was used to specify the attitudes, expectations, fears, and/or wishes associated with each disorder, as well as the quality of impairment or distress associated with each. (See Table 1 for examples of items for each of the personality disorder categories.)

This method was used to draft most of the personality disorder items in the WISPI. Selected items from the Chapman Schizotypy Scale (Chapman & Chapman,

Table 1. Sample WISPI Items for Each of the Personality Disorder Categories (Criterion 1)

Paranoid: Certain people will take unfair advantage of me if they get the slightest chance.
Schizoid: Other people's feelings just don't move me one way or the other.
Schizotypal: I often get personal messages from the media (TV, radio, the news) that were sent especially to me.
Histrionic: I demand that people show their appreciation for my efforts to entertain them.
Narcissistic: I feel humiliated and angry when I hear about someone who is more appreciated than I am.
Antisocial: Before I was 15 years old, people were already giving me a hard time for breaking the rules at home or school.
Borderline: I shift back and forth between strong love and strong hate for the people I am closest to.
Avoidant: When people look at me, I am afraid that they will criticize or make fun of me for being strange or weird.
Dependent: Because I don't trust my judgment, I wait to make everyday decisions until the person close to me is there to take over.
Compulsive: I often get so involved in making each detail of a project absolutely perfect that I never finish.
Passive-aggressive: I have agreed to do more at work, home, or school, but I just never get around to it.

1986) and some behavioral descriptors were also added for trial purposes. In addition to items for the 11 personality disorder categories, we included 10 items corresponding to Millon's (1981) more general conception of antisocial personality disorder, the 10-item Marlowe–Crowne Scale (Greenwald & Satow, 1970) for social desirability response bias (MCSD), 3 global items for social adjustment (SAS), and the 36-item introject scale from Benjamin's Intrex Questionnaire (1983). Only results for the 302 WISPI items that correspond to the 11 DSM-III-R categories and the MCSD are covered in this report.

Clinician Sort. To determine the content validity of items with respect to DSM-III-R categories, four postdoctoral level clinicians (average of 9 years of clinical experience) working in pairs, performed the task of sorting the WISPI items into the Axis II categories. As preparation, the clinicians reviewed the Axis II section of the DSM-III and personality disorder cases in the DSM-III Casebook (Spitzer, Skodol, Gibbon, & Williams, 1981). Each WISPI item was presented in random order to be sorted into only 1 of the 11 categories (DSM-III descriptions were available).

SASB Coding of WISPI Items. To determine the correspondence of the WISPI items with the original interpersonal descriptions of each disorder, an experienced SASB coder, Dr. Laura Humphrey, coded each of the WISPI items. Items were presented in the inventory order and Dr. Humphrey was blind to their personality disorder category designation. Following standard SASB coding procedures (Benjamin, Foster, Giat-Roberto, & Estroff, 1986), the coder judged each thought unit within each item for focus, degree of affiliation, and degree of autonomy. This procedure yielded codes for the eight-cluster level of the SASB model (Benjamin, 1984). Next, the SASB codes for all items within each scale were summarized as profiles, defined as the percent distribution of each focus and cluster code within each personality disorder category. Two standards were then used to calculate validity coefficients: (1) within-category profiles of Benjamin's original SASB codes of the Axis II criteria for each personality disorder category that had guided item

construction and (2) within-category profiles of Benjamin's SASB codes of the specific WISPI items.

Subjects. Nonpatient-subjects were recruited from the general public by means of media advertisements, posted announcements, visits to classrooms, and solicitation from visitors to the University Hospital. Patient-subjects were recruited from psychiatry inpatient and outpatient services, community mental health centers, a psychology clinic, and a student counseling center, or by means of posted announcements and staff referral. Informed consent was obtained from all subjects; patients were also asked for permission to review charts and contact therapists for diagnostic and treatment information.

A total of 1230 subjects completed WISPIs; 1059 were given the paper-and-pencil form; 171 the computer form. This total sample is referred to as the "validation sample" in subsequent sections. The 368 patients within the validation sample were identified by self-report (n = 131) or by referral source (n = 237) as currently receiving some form of current mental health care. Among the 862 subjects from the general public who were not currently patients, another 208 reported having received some form of mental health treatment in the past. Similar procedures were used to recruit an additional 80 subjects for the test–retest study. Subject characteristics for each of these samples are summarized in Table 2.

Scale Instructions and Administration. Each of the 360 items in the WISPI is rated on a 10-point scale (1 = "never or not at all true of you" to 10 = "always or extremely true of you"). Respondents are instructed to rate their "usual selves during the past five years or more," that is, to think of how they "typically approach work and other day-to-day responsibilities, their relationships with others and their feelings about themselves."

The inventory was administered in two forms, a paper-and- pencil form with mark-sense answer sheets and a computer interview version (Erdman, Klein, & Greist, 1985).

Scoring. Summary scores used for this report are the means of the ratings of all (nonmissing) items for each scale.

Other Measures and Assessments. All subjects taking the WISPI were given a brief demographic and background questionnaire. In addition, 146 of the patient-subjects were given various diagnostic and symptomatic assessments: (1) Millon Clinical Multiaxial Inventory-I (MCMI—Millon, 1982), (2) Personality Disorders Questionnaire (PDQ—Hyler, et al.. 1988), and (3) Symptom Checklist 90-R (Derogatis, 1983). Therapists familiar with each patient were asked to complete an Axis II diagnostic checklist, to list all Axis I and Axis II diagnoses, and to rate each patient on the Personality Assessment Form (PAF—Shea, Glass, Pilkonis, Watkins, & Docherty, 1987). These 146 subjects are referred to as the "clinical subsample."

Test-Retest Study. A total of 80 subjects—40 patients, 40 nonpatients—were recruited as already described. Half in each group were assigned to one of four test–retest sequences that counterbalanced the two forms of the WISPI, paper-and-pencil (PP) and computer interview (CI): PP-PP, CI-CI, PP-CI, and CI-PP. Retests occurred within 2 weeks.

Analyses. The SPSS RELIABILITY program was used to evaluate the internal consistency of the 11 scales. MANOVA and ANOVA procedures were used to compare patients and nonpatients on all items and scale means. Kappas were used to calculate the reliability of the clinicians' sorting of the items against the Axis II categories. For validity coefficients, correlations were corrected for attenuation (Guilford, 1954), using test–retest reliabilities where available, for the MCMI-I (Millon, 1982), MMPI (Hurt, Clarkin, & Morey, 1990), and PDQ (Hurt, Hyler, Frances, Clarkin, & Brent, 1984); and interrater reliabilities for SASB and PAF ratings (P. Pilkonis, personal communication, November 1990). In analyses of the

Table 2. Demographic Characteristics of Validation and Reliability Study Samples

	Validation study						Reliabilty study			
	Never patients (n = 654)		Former patients (n = 208)		Current patients (n = 368)		Nonpatients (n = 40)		Patients (n = 40)	
Variable	n	%	n	%	n	%	n	%	n	%
Sex										
Male	215	33	56	27	98	27	6	15	10	25
Female	439	67	152	73	269	73	33	83	30	75
NA	0		0		1		1		0	
Education										
High school	27	4	10	5	56	15	5	13	4	10
Some college	458	70	131	63	185	50	17	43	18	45
College	125	19	41	20	70	19	8	20	3	8
Postcollege	32	5	18	9	44	12	10	25	15	28
NA	11		8		13		0		0	
Age										
Range	16–86		18–60		17–82		18–56		18–75	
Mean	24.4		27.6		31.8		30.9		34.7	
SD	9.1		9.7		11.1		10.6		13.3	
NA	8		7		7		0		0	

Note.. NA = not available.

11 WISPI scales, alphas were adjusted by dividing $p = .05$ by 11 for a corrected A = .005.

RESULTS

CONTENT VALIDITY

Clinician Sort. Overall Kappas for each pair's sorting of all of the WISPI items against the 11 personality disorder categories were .81 and .90. Kappas for specific categories ranged from .57 (histrionic) to .96 (passive–aggressive) for pair A; from .73 (narcissistic) to .98 (passive–aggressive and dependent) for pair B. The median kappa for pair A was .82; for pair B, .89. Pair B's somewhat better reliability may have been due to the fact that they had longer preparation time and somewhat more clinical experience (10 vs. 8 years). When pair A missorted histrionic items, they were most frequently placed in borderline or dependent categories. This is not surprising, considering the overlap in content among these categories as defined by DSM-III.

SASB Ratings. Correlations between the profiles of the SASB ratings of WISPI items made blindly by Humphrey's with Benjamin's original ratings of the items averaged .68 and ranged from .43 (schizotypal) to .97 (paranoid). The lower correlations were largely due to the two coders' consistently choosing different but closely related categories: the same cluster location but a different focus (e.g., schizoid) or adjacent clusters of the same focus (e.g., histrionic and narcissistic). Correlations between Humphrey's item rating profiles and the profiles of Benjamin's ratings of the Axis II criteria averaged .69 and ranged from .41 (narcissistic) to .93 (paranoid). When corrected for attenuation, using the Benjamin-Humphrey item correlations to adjust Humphrey's item ratings and the Benjamin criteria and item correlations to adjust Benjamin's criteria ratings, the average corrected correlation between Humphrey's item ratings and Benjamin's conception of the DSM criteria was .79; the lowest correlation was .68 for the narcissistic scale.

Reliability and Internal Consistency. Table 3 summarizes the scale means, medians, variances, ranges, and several measures of internal consistency for the full validation sample. Means were highest for two Cluster C scales (avoidant, compulsive) and lowest for antisocial and schizotypal scales. Histrionic, paranoid, and dependent scales had the greatest ranges. Table 4 summarizes results of the test–retest study. Test–retest coefficients averaged .88 and ranged from .71 (schizoid) to .94 (dependent). Test–retest coefficients for the two methods of test administration were in the same ranges. Repeated-measures MANOVAS of scale scores were significant for Times, $F(11, 62) = 3.08$, $p < .001$, but not for Methods, $F(11, 62) = 1.18$, $p < .322$. Univariate paired t-tests for the Time variable showed that retest means were significantly lower than the initial test means for 7 of the 11 WISPI scales. Exceptions were schizotypal, narcissistic, borderline, and avoidant, all scales in the middle of the profile for this sample. In most respects, however, the overall profile remained the same over time.

Table 5 summarizes the pattern of interscale correlations for the

Table 3. Means, Standard Deviations (SD) and Internal Consistency Coefficients for WISPI Scales in the Validation Sample (N = 1230)

Scale	Item n	Range	Median	Mean	*SD*	Alpha	Mean item-total r	Mean interitem r
Paranoid	37	1.00–9.00	3.04	3.21	1.27	.937	.519	.288
Schizoid	22	1.09–7.77	2.64	2.87	1.10	.847	.419	.208
Schizotypal	33	1.00–7.36	2.00	2.33	1.11	.923	.500	.272
Histrionic	26	1.00–8.04	2.87	3.03	1.08	.876	.438	.219
Narcissistic	28	1.00–7.86	2.93	3.10	1.19	.906	.496	.273
Antisocial	35	1.00–6.14	1.57	1.84	0.84	.908	.458	.232
Borderline	20	1.00–7.65	2.73	2.98	1.29	.884	.498	.282
Avoidant	29	1.00–9.35	3.31	3.68	1.75	.958	.650	.442
Dependent	24	1.00–8.92	2.75	3.06	1.35	.924	.568	.350
Compulsive	19	1.05–7.95	3.37	3.52	1.23	.841	.431	.221
Passive-aggressive	29	1.00–7.30	2.67	2.82	1.07	.889	.450	.227
Average			2.72	2.95	1.21	.899	.493	.274

Table 4. Means, Standard Deviations, and Paired *t*-test Comparisons of Time 1 and Time 2 WISPI Scores, and Test-retest Coefficients the Reliability Sample

	Time 1 (n = 80)		Time 2 (n = 80)			Test-retest *rs*	
Scale	*M*	*SD*	*M*	*SD*	*p*	Time: T1/T2	Method: C/PP
Paranoid	2.87	1.29	2.67	1.41	.003	.911	.905
Schizoid	2.70	1.01	2.43	0.94	.001	.712	.564
Schizotypal	2.11	1.19	2.05	1.27	.290	.933	.927
Histrionic	2.68	1.01	2.48	0.99	.001	.871	.895
Narcissistic	2.57	1.05	2.49	1.17	.243	.862	.876
Antisocial	1.58	0.54	1.51	0.60	.039	.879	.876
Borderline	2.61	1.34	2.50	1.34	.125	.881	.836
Avoidant	3.54	1.85	3.41	2.05	.140	.920	.903
Dependent	2.77	1.47	2.58	1.54	.003	.936	.915
Compulsive	3.12	1.30	2.87	1.41	.005	.832	.775
Passive-aggressive	2.44	1.01	2.24	1.01	.000	.897	.824
Overall *F*		3.065 df 11/69			.002		
Average *r*						.876	.845

Note. Adjusted critical *value for* A = .05 is $p < .005$.

validation sample. Interscale correlations are quite high: The overall average was .62, and averages within specific scales ranged from .48 (schizoid) to .69 (paranoid).

CONCURRENT VALIDATION

Discrimination Between Patients and Nonpatients. In order to see if the WISPI scales discriminated between patients and nonpatients, responses of patients and nonpatients were compared in the validation and reliability samples, respectively (Table 6). Comparison by MANOVA for all scales indicated an overall significant difference between patients and nonpatients in both the validation, $F(11, 1185) = 23.93$, $p < .001$, and the reliability samples, $F(11,,62) = 4.852$, $p < .001$. In univariate ANOVAS, the validation sample patients scored significantly higher than nonpatients on 8 of the 11 scales. Patients did not differ from nonpatients on histrionic, narcissistic, or antisocial scales. These differences held for both males and females (analyses not shown). In the smaller reliability sample, patients and nonpatients did not differ on the schizoid, histrionic, and antisocial scales.

In the clinical subsample where additional diagnostic data were available, it was possible to compare the 62 patients who had no chart diagnosis of personality disorder (PD) with those with either a history of any PD (n = 21) or any current PD diagnosis (n = 52). Results of three-group MANOVAs, ANOVAs, and post hoc comparisons are summarized in Table 7. Consistent with results for the full sample, the PD groups did not differ on histrionic, narcissistic and antisocial scales, and differences for the dependent scale also did not reach the corrected p value. For all 7 of the remaining scales where there were significant differences, the patients without PDs scored lowest; and the patients with any current PD, the highest. For four comparisons, the group with a PD-history was intermedi-

Table 5. WISPI Interscale Correlations for Validation Sample (N = 1230)

Scale	Par	Szd	Stp	His	Nar	Asp	Bor	Avo	Dep	Com	Pas
Schizoid	.58										
Schizotypal	.73	.53									
Histionic	.67	.29	.60								
Narcissistic	.77	.41	.64	.82							
Antisocial	.60	.47	.68	.58	.58						
Borderline	.74	.44	.74	.73	.69	.65					
Avoidant	.69	.53	.58	.41	.53	.38	.68				
Dependent	.64	.44	.59	.58	.58	.44	.75	.81			
Compulsive	.72	.49	.54	.53	.66	.38	.59	.66	.61		
Passive-aggressive	.79	.54	.71	.69	.78	.69	.71	.60	.64	.63	
Average *r*	.69	.48	.63	.59	.64	.56	.66	.56	.58	.57	.67

Note. All correlations are significant at $p < .001$.

Table 6. Means, Standard Deviations, and Analysis of Variance Comparisons of Nonpatients and Patients in the Validation and Reliability Samples

	Validation sample					Reliability sample (Time 1)				
	Nonpatient (n = 852)		Patient (n = 368)			Nonpatient (n = 40)		Patient (n = 40)		
Scale	*M*	*SD*	*M*	*SD*	*p*	*M*	*SD*	*M*	*SD*	*p*
Paranoid	3.08	1.15	3.50	1.48	.000	2.39	1.25	3.35	1.15	.001
Schizoid	2.73	0.99	3.22	1.26	.000	2.64	1.12	2.77	0.89	.591
Schizotypal	2.26	1.04	2.50	1.23	.001	1.73	1.11	2.48	1.17	.004
Histrionic	3.01	1.03	3.08	1.19	.309	2.24	0.95	2.94	1.00	.019
Narcissistic	3.07	1.14	3.18	1.29	.115	2.14	0.90	3.00	1.01	.000
Antisocial	1.80	0.82	1.91	0.89	.036	1.42	0.38	1.73	0.63	.011
Borderline	2.77	1.12	3.48	1.50	.000	1.99	1.09	3.23	1.30	.000
Avoidant	3.26	1.42	4.64	2.04	.000	2.54	1.30	4.54	1.78	.000
Dependent	2.78	1.12	3.72	1.60	.000	2.07	0.79	3.47	1.66	.000
Compulsive	3.37	1.10	3.88	1.42	.000	2.60	1.19	3.65	1.20	.000
Passive-aggressive	2.75	1.00	2.98	1.20	.001	1.99	0.77	2.89	1.04	.000
Overall F		(12, 1217) = 28.612			.001		F(11, 68) = 3.829			.001

Note. Adjusted critical p value for A = .05 is p = .005.

Table 7. Means, Standard Deviations, and Analysis of Variance Comparisons of Patients in the Clinical Subsample by Diagnosis of Personality Disorder

	No PD[a] (*n* = 62)		PD History[a] (*n* = 21)		Current PD[a] (*n* = 52)		
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p* value
Paranoid	3.18[a]	1.30	3.69[ab]	1.22	4.11[b]	1.72	.004
Schizoid	3.01[a]	1.08	3.88[b]	1.04	3.69[b]	1.30	.001
Schizotypal	2.11[a]	0.96	2.58[ab]	0.98	3.16[b]	1.48	.001
Histrionic	2.91	1.07	3.01	0.99	3.18	1.21	.449
Narcissistic	2.94	1.16	2.98	0.76	3.40	1.41	.115
Antisocial	1.85	0.88	1.59	0.37	1.97	0.84	.188
Borderline	3.09[a]	1.22	3.35[a]	1.18	4.19[b]	1.52	.001
Avoidant	4.52[a]	1.82	5.81[b]	1.75	5.95[b]	2.00	.000
Dependent	3.73	1.62	4.13	1.19	4.67	1.81	.011
Compulsive	3.64[a]	1.26	4.49[ab]	1.53	4.68[b]	1.57	.001
Passive-aggressive	2.75[a]	1.04	2.96[ab]	0.71	3.44[b]	1.28	.005
Overall $F(22, 242) = 2.882$							.001

Note. Adjusted critical *p* value for A = .05 is *p* = .005.
[a]Means with different superscripts are significantly different at the .05 level for Tukey's HSD.

ate and not significantly different from either group. For the avoidant and schizoid scales, the PD-history patients resembled the current-PD patients; for the borderline scale, the PD-history group resembled the non-PD patients.

Other Self-Report Personality Disorder Measures. Concurrent validity was assessed in the clinical subsample by comparing the WISPI with the results of two other personality disorder inventories, the MCMI-I and the PDQ (Table 8). Uncorrected correlations between the same scales of the WISPI and MCMI-I averaged .39 and ranged from -.26 (compulsive) to .68 (dependent). The average of all off- diagonal correlations between WISPI and MCMI-I scales was .11. When corrected for attenuation (Guilford, 1954), coefficients averaged .43. To compute correlations between the PDQ and the other measures, we calculated PDQ dimensional scores as the percent of items within each scale rated "true." This approach weighted all items equally and thus was similar to the scoring conventions for the WISPI, MCMI, and other personality disorder measures. Uncorrected correlations between WISPI and PDQ averaged .69 and ranged from .37 (schizoid) to .79 (dependent). The average off-diagonal correlation was .37. When corrected for attenuation, WISPI-PDQ coefficients averaged .93.

Clinician Ratings. In addition to self-report measures, 103 of the patients in the clinical subsample were also rated by treatment staff on the PAF, using a 5-point scale of the extent to which a given individual resembled brief descriptions of each of the PD prototypes. Results for all 103 patients who were rated are summarized in the first column of Table 9. Although all but one of the uncorrected correlations were significant, the average was low (mean of .33; range from .21 [schizoid] to .49 [borderline]). The average for corrected coefficients was .51. However, since not all treatment staff were equally familiar with the patients they

Table 8. Raw and Corrected Correlations of WISPI with MCMI and PDQ Scores in the Clinical Subsample (N = 146)

	WISPI/MCMI		WISPI/PDQ	
Scale	r	Corrected r	r	Corrected r
Paranoid	.38**	.43	.66**	.90
Schizoid	.48**	.62	.37**	—[a]
Schizotypal	.43**	.48	.72**	.99
Histrionic	.32**	.36	.68**	1.00[b]
Narcissistic	.14	.16	.54**	1.00[b]
Antisocial	.10	.11	.79**	.98
Borderline	.57**	.66	.67**	.90
Avoidant	.79**	.87	.75**	.94
Dependent	.68**	.77	.79**	1.00[b]
Compulsive	-.26*	-.32	.57**	.72
Passive-aggressive	.50**	.56	.67**	.83
Average r	.39	.43	.69	.93
Average off-diagonal r	.11		.37	

[a] Test-retest r for Schizoid scale not reported in Hurt et al. (1984).
[b] Corrected coefficients are > 1.00 when test-retest reliability for one or both measures is low.
*$p < .01$
**$p < .001$

were asked to rate, the correlations were repeated on a restricted sample of 33 patients for whom the therapists' reported levels of confidence for their ratings had been good or excellent. As is apparent in Table 9, the uncorrected and corrected correlations of these more "confident" ratings with WISPI scores were somewhat higher for the majority of the scales.

DISCUSSION

The WISPI appears to have good psychometric properties. Alpha coefficients (mean = .90) were close to those reported for the comparable MCMI-I scales (mean = .85—Millon, 1982) and identical to those for the MCMI-II (mean = .90—Millon, 1987). They were much higher than internal consistency coefficients reported for the PDQ (mean = .68—Hyler et al., 1988) and MMPI PD scales (mean = .76—Morey, Waugh, & Blashfield, 1985). Other measures of internal consistency also indicated substantial cohesiveness for all 11 WISPI scales: Item-total correlations averaged .49. Only 14 of the 311 items had item-total correlations below the .30 level. Mean interitem correlations ranged from .21 (schizoid) to .44 (avoidant) and averaged .27. Test-retest reliabilities over a 2-week period also were high (mean = .88) and were not influenced by method of administration. These can be compared with an average for the MCMI-I of .86 (Millon, 1982); an average of .76 for the MMPI PD scales for 3-week retest interval (Hurt, Clarkin, & Morey, 1990), and with an average of .58, with a very wide range (from -.25 for passive-aggressive to .75 for compulsive) for the PDQ (Hurt et al., 1984).

On the other hand, the interscale correlations for the WISPI were much higher than those reported for the other available self- report measures of

Table 9. Raw and Corrected Correlations of WISPI with PAF Ratings in the Clinical Subsample

Scale	WISPI/all PAF (n = 103)		WISPI/confident PAF (n = 33)	
	r	Corrected r	r	Corrected r
Paranoid	.23*	.35	.39**	.62
Schizoid	.21	.35	.06**	.10
Schizotypal	.40**	1.00[a]	.51**	1.00[a]
Histrionic	.41**	.50	.44**	.54
Narcissistic	.27*	.38	.45**	.63
Antisocial	.36**	.51	.59**	.84
Borderline	.49**	.59	.58**	.70
Avoidant	.28*	.33	.37*	.44
Dependent	.37**	.66	.26	.46
Compulsive	.26*	.42	.31	.45
Passive-aggressive	.40**	.53	.51**	.68
Average r	.33	.51	.41	.59
Average off-diagonal r	.19		.24	

[a] Corrected coefficients are > 1.00 when test-retest reliability for one or both measures is low.
*$p < .01$.
**$p < .001$.

personality disorder. Interscale correlations for the 11 PD scales in common with the WISPI averaged .11 for the MCMI-I and .20 for the MCMI-II (Millon, 1982, 1987). The interscale correlations for the MMPI PD scales reported by Morey, Waugh, and Blashfield (1985) were slightly higher (average = .26 for complete scales; average = .28 for nonoverlapping scales).

The pattern of interscale correlations for the WISPI is undoubtedly subject to many of the same factors that influence interscale correlations for other methods, such as the high rate of comorbidity within Axis II (Blashfield & Breen, 1989; Mellsop, Varghese, Joshua, & Hicks, 1982; Morey, 1988b), overlap between categories as defined in DSM-III-R (Widiger, Frances, Spitzer, & Williams, 1988), and the effect of general aspects of impairment that cross-cut all categories. In addition, there are several differences between this initial version of the WISPI and these two self-report measures that may account for the even higher interscale correlations for the WISPI. First, the WISPI items are rated dimensionally whereas both the MCMI and MMPI require true-false responses. Second, the MCMI and MMPI scoring systems increase separation between certain scales by reverse scoring of some items. Third, some of the the MCMI scores are corrected for response bias, and extreme tendencies for denial or complaining are flagged for other scales.

Several steps were taken to assess the possible effect of various response biases on the interscale correlations for the WISPI. First, we computed the matrix of interscale part- correlations, controlling for MCSD scores. This correction reduced the correlations slightly; the average interscale part- correlation was reduced from .62 (range from .38 to .82) to .54 (range from .26 to .77) in the full validation sample. Next, to explore the extent to which interscale correlations may have also been influenced by a general tendency to complain or report symptoms, scores on the Global

Symptom Intensity (GSI) scale of the SCL-90-R were used to correct the interscale correlations in the clinical subsample. When this was done, the average of the correlations fell from .52 (range from .12 to .78) to .35 (range from .01 to .72). These are closer to, but still not as low as those reported for the MCMI and MMPI.

Another way to control for response bias is to adjust scores for each individual's tendency to respond at a particular level on the response scale, that is, to calculate scores that reflect the variation within each individual. For this purpose, ipsatized scores were defined within each subject as the deviation (score) of each subject's 11 PD scale scores from the overall mean of that subject's scores. This approach yielded much lower interscale correlations: average of -.09, range from -.48 (histrionic with avoidant) to .38 (avoidant with dependent). This demonstrates that the WISPI PD profiles are different for different individuals when the general response level is controlled. In future applications of the WISPI, we recommend a two-part scoring and interpretation procedure, using the within- individual z scores to establish which specific scales can be considered to be significant peaks within a given profile, then using z scores based on the normative sample to determine the level of endorsement of qualifying scales relative to the normative sample. To define cases, we suggest setting the critical value of z at +1.96. An individual whose z score for a given scale is equal to or larger than this value would have a score that is significantly greater than the normative sample at $p < .05$.

The problem of interscale correlation can also be reduced by applying the inclusion and exclusion criteria recently developed by Benjamin (1993). For a diagnosis of borderline personality disorder, for example, Benjamin would require that a respondent (1) endorse items that reflect fears of abandonment that lead to coercive attempts to force others to give nurturance and lead to self-sabotage for doing well or being happy (inclusion criteria), and (2) not endorse items in schizoid, schizotypal, paranoid, and narcissistic categories concerned with long term comfort with autonomy (exclusion criteria). These necessary and exclusionary criteria are being incorporated into scoring procedures for the WISPI and can be used for scoring that more closely approximates DSM-III-R categories.

The results of the clinician sort and the independent coding of items for interpersonal content indicate good content validity for the WISPI vis-a-vis both the Axis II category descriptions and the formulations derived from interpersonal theory. This suggests that fidelity to DSM-III-R can be achieved without sacrificing theoretical relevance.

The concurrent validity of the WISPI was assessed by comparing WISPI scores of various nonpatient and patient groups and, within the patient subsample, by comparing WISPI results with the PDQ and MCMI-I responses and with clinician PAF ratings. In both the large validation and smaller reliability study samples, patients had higher WISPI profiles overall, and scored higher than nonpatients on 8 of the 11 scales. In both samples, patients and nonpatients were equally moderate on histrionic scales and equally low on antisocial scales. In the larger validation sample, patients and nonpatients also had similar scores on the narcissistic scale;

in the reliability sample, it was the schizoid scale that failed to discriminate patients from nonpatients. The finding that patients and nonpatients could not be discriminated on histrionic and narcissistic scales is consistent with reports by Millon for the MCMI-I (1982, p. 11) and by Morey, Blashfield, Webb, and Jewell (1988) for the MMPI PD scales. Means for histrionic and narcissistic are generally similar in both normals and patients for all three self-report measures.

The most likely explanation for the lack of discrimination for the three Cluster B scales is the composition of the patient sample; they were outpatients, not severely ill, predominantly with Axis I diagnoses of depression and/or anxiety disorders. Consistent with these Axis I diagnoses, three of the scales from the "anxious-avoidant" cluster (avoidant, compulsive, and dependent) were the peak scores in the patient sample. High rates for Cluster C disorders in outpatient samples have also been reported for the MCMI (Millon, 1982), the MMPI-PD (Morey, 1988a), the PAF (Pilkonis & Frank, 1988), and for the WISPI in other psychiatric outpatient samples (K. Moras, personal communication, January 1992; P. Pilkonis, personal communication, January 1989). Also, the interpersonal submissiveness that is characteristic of these Cluster C diagnoses is antithetical to the more active demanding behavior that characterizes the Cluster B disorders (Benjamin, 1993). The lack of differentiation between patients and nonpatients on histrionic and narcissistic scales may also reflect the presence of untreated cases in the normal population tested and the low prevalence of antisocial disorders in both samples. Finally, it is possible that the DSM-III and/or the self-report procedures may fail to discriminate adequately between traits and disorder in these areas.

The pattern of results was similar when patients with current and past personality disorders were compared with patients in the clinical subsample with Axis I disorders only. Patients with current personality disorders had higher WISPI scores than patients with no personality disorder; patients with histories of any personality disorder were generally intermediate. The same three scales, histrionic, narcissistic, and antisocial, that failed to differentiate patients from normals also did not discriminate within the patient group, probably for the same reasons as already enumerated. This is also consistent with Morey and Smith's (1988) hypothesis that moderate levels on certain personality scales such as histrionic and narcissistic in some clinical samples may reflect health. Despite these discrepancies, this preliminary version of the WISPI shows good discriminatory power for clinical and nonclinical samples for most scales.

Concurrent validity was also assessed within the clinical subsample by comparing WISPI results with self-reports on the MCMI-I and PDQ and with clinician ratings on PAF dimensions. Within the self-report domain, the WISPI was more closely associated with the PDQ than with the MCMI-I. The WISPI-MCMI and WISPI-PDQ correlations were almost identical to those reported by Dubro, Wetzler, and Kahn (cited in Morey, 1988a) for the MCMI-I, MMPI PD, and PDQ in a mixed sample of psychiatric inpatients and medical controls: The median correlation between MMPI PD and MCMI-I was .37 (range from .51, avoidant, to -.37, compulsive); for MMPI PD and PDQ the median correlation was .64 (range from .81, dependent, to

.05, narcissistic). Correlations between MCMI-I and MMPI PD reported by Streiner and Miller (1988) were slightly higher (median = .43, range from .66, narcissistic to -.38, compulsive), and Morey and Le Vine (1988) found even stronger associations between these two measures in a sample largely composed of psychiatric outpatients (median = .68, range from .78, schizotypal to -.31, compulsive). Considering the fact that the PDQ was specifically written to tap DSM-III constructs, whereas the MCMI-I reflects Millon's more complex multidimensional theory, it is understandable that closer associations with the PDQ are found for both the WISPI and MMPI PD scales than for the MCMI. This suggests that our effort to make the WISPI faithful to DSM-III-R theory has been successful. Finally, when clinician PAF ratings were used as the validation criterion, coefficients were somewhat lower but still within a reasonable range, especially when the analysis was limited to the more confident therapists' ratings.

In sum, these studies of this initial version of the WISPI provide evidence for good content validity of the inventory as a whole and suggest that the internal consistency and reliability of scales for specific personality disorders are very high, even when paper-and-pencil and computer methods of administration are compared. Although interscale correlations were larger than is desirable, the use of ipsatized scores and the application of Benjamin's necessary and exclusionary criteria may correct this problem. In general, the WISPI scales discriminated between nonpatients and patients; within a patient group, those with current personality disorders scored highest on most scales. Finally, within the clinical subsample where other measures of personality disorders were available, the WISPI correlated well with the PDQ, and moderately with the MCMI-I and clinician ratings on the PAF dimensions. This is further indication that we were successful in our efforts to design a measure of Axis II that adheres to DSM-III criteria while being grounded in interpersonal theory.

Although there are other personality disorder self-report measures in current use, we believe that this new inventory has much to offer clinical practice and research. It was developed within the framework of a theoretically coherent, clinically grounded conception of the personality disorders. While it also conforms to DSM-III-R, the structure provided by SASB would allow researchers to test hypotheses about the interpersonal features of the personality disorders. These results for the initial version of the WISPI are promising, but further steps of development and validation are necessary. Results of the present study have been used to refine and shorten the WISPI to 224 items, and this revised version is currently being evaluated in patient and nonpatient samples.

REFERENCES

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review, 81*, 392–425.

Benjamin, L. S. (1983). *INTREX users man-*

ual, Pts. I and II. Madison, WI: INTREX Interpersonal Institute.

Benjamin, L. S. (1984). Principles of prediction using structural analysis of social behavior (SASB). In R. A. Zucker, J. Aranoff & A. J. Rabin (Eds.), *Personality and prediction of behavior* (pp. 121–174). New York: Academic Press.

Benjamin, L. S. (1986). Adding social and intrapsychic descriptors to Axis I of DSM-III. In T. Millon & G. L. Klerman (Eds.), *Contemporary issues in psychopathology* (pp. 599–638). New York: Guilford Press.

Benjamin, L. S. (1987). Use of the SASB dimensional model to develop treatment plans for personality disorders. I: Narcissism. *Journal of Personality Disorders, 1*, 43–70.

Benjamin, L. S. (1993). *An interpersonal approach to the DSM personality disorders*. New York: Guilford Press.

Benjamin, L. S., Foster, S. W., Giat-Roberto, L., & Esteroff, S. E. (1986). Breaking the family code: Analyzing videotapes of family interactions by Structural Analysis of Social Behavior. In L. S. Greenberg & W. M. Pinsoff (Eds.), *The psychotherapeutic process: A research handbook* (pp. 391–438). New York: Guilford Press.

Blashfield, R. K., & Breen, M. J. (1989). Face validity of the DSM-III-R personality disorders. *American Journal of Psychiatry, 146*, 1575–1579.

Chapman, L. J., & Chapman, J. P. (1986). Psychosis proneness. In M. Alpert (Ed.), *Controversies in schizophrenia* (pp. 157–174) New York: Guilford Press.

Clark, L. A. (1990). *Preliminary manual for the Schedule of Normal and Abnormal Personality (SNAP): A dimensional approach to the assessment of personality and its disorders*. Dallas, TX: Southern Methodist University.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual-II* (2nd ed.). Towson, MD: Clinical Psychometric Research.

Erdman, H. P., Klein, M. H., & Greist, J. H. (1985). Direct patient computer interviewing. *Journal of Consulting and Clinical Psychology, 53*, 760–773.

Greenwald, H. J., & Satow, Y. (1970). A short social desirability scale. *Psychological Reports, 27*, 131–135.

Guilford, J. P. (1954). *Psychometric methods. New York: McGraw-Hill.*

Hurt, S. W., Clarkin, J. H., & Morey, L. C. (1990). An examination of the stability of the MMPI personality disorder scales. *Journal of Personality Assessment, 54*, 16–23.

Hurt, S. W., Hyler, S. E., Frances, A., Clarkin, J. F., & Brent, R. (1984). Assessing borderline personality disorder with self-report, clinical interview, or semistructured interview. *American Journal of Psychiatry, 141*, 1228–1231.

Hyler, S. E., Reider, R. O., Williams, J. B., Spitzer, R. L., Hendler, J., & Lyons, M. (1988). The Personality Diagnostic Questionnaire: Development and preliminary results. *Journal of Personality Disorders, 2*, 229–237.

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology: Vol. 2* (pp. 61–96). New York: Academic Press.

Kiesler, D. J. (1983). The 1982 Interpersonal Circle: A taxonomy for complementarity in human transactions. *Psychological Review, 90*, 185–214.

Klein, M. H. (1993). Issues in the assessment of personality disorders. *Journal of Personality Disorders, 7*, Supplement, 18–33.

Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald Press.

Livesley, W. J. (1986). Trait and behavioral prototypes of personality disorder. *American Journal of Psychiatry, 143*, 728–732.

McLemore, S., & Benjamin, L. S. (1979). Whatever happened to interpersonal diagnosis? A psychological alternative to DSM-III. *American Psychologist, 34*, 17–34.

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

Millon, T. (1981). *Disorders of personality: DSM-III: Axis II*. New York: Wiley.

Millon, T. (1982). *Millon Clinical Multiaxial Inventory Manual* (2nd ed.). Minneapolis MN: National Computer Systems.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory-II: MCMI-II* (2nd ed.). Minneapolis, MN: National Computer Systems.

Morey, L. C. (1985). An empirical comparison of interpersonal and DSM-III approaches to classification of personality disorders. *Psychiatry, 48*, 538–364.

Morey, L. C. (1988a). *The MMPI Personality Disorder Scales: A manual and guide to interpretation*. Nashville, TN: Vanderbilt University.

Morey, L. C. (1988b). Personality disorders in DSM-III and DSM-III-R: Convergence, coverage, and internal consistency. *American Journal of Psychiatry, 145*, 573–577.

Morey, L. C., Blashfield, R. K., Webb. W. W.,

& Jewell, J. (1988). MMPI scales for DSM-III personality disorders: A preliminary validation study. *Journal of Clinical Psychology, 44*, 47–50.

Morey, L. C. & Le Vine, D. J. (1988). A multitrait-multimethod examination of Minnesota Multiphasic Personality Inventory (MMPI) and Millon Clinical Multiaxial Inventory (MCMI). *Journal of Psychopathology and Behavioral Assessment, 10*, 333–344.

Morey, L. C., & Smith, M. R. (1988). Personality Disorders. In R. Greene (Ed.), *The MMPI: Use with specific populations* (pp. 110–158). New York: Grune & Stratton.

Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for DSM-III personality disorders: Their derivation and correlates. *Journal of Personality Assessment, 49*, 245–251.

Pilkonis, P., & Frank, E. (1988). Personalty pathology in recurrent depression: Nature, prevalence, and relationship to treatment response. *American Journal of Psychiatry, 145*, 435–441.

Shea, T., Glass. D., Pilkonis, P., Watkins, J., & Docherty, J. (1987). Frequency and implications of personaltiy disorders in a sample of depressed outpatients. *Journal of Personality Disorders, 1*, 27–42.

Spitzer, R. L., Skodol, A. E., Gibbon, M., & Williams, J. B. (1981). *DSM-III casebook*. Washington DC: American Psychiatric Association.

Streiner, D. L., & Miller, H. R. (1988). Validity of the MMPI scales for DSM-III personality disorders: What are they measuring? *Journal of Personality Disorders, 2*, 238–242.

Widiger, T. A., & Frances, A. (1985). The DSM-III personality disorders: Perspectives from psychology. *Archives of General Psychiatry, 42*, 615–623.

Widiger, T. A., Frances, A., Spitzer, R. L., & Williams, J. B. W. (1988). The DSM-III-R personality disorders: An overview. *American Journal of Psychiatry, 145*, 786–795.

Widiger, T. A., & Kelso, K. (1983). Psychodiagnosis of Axis II. *Clinical Psychology Review, 3*, 491–510.

Wiggins, J. S. (1982). Circumplex models of interpersonal behavior in clinical psychology. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 183–221). New York: Wiley.