Journal of Personality Disorders, 8(4), 307-319, 1994
©1994 The Guilford Press

NOTICE: This material may be protected
By copyright law (Title 17 U.S. Code)

VALIDATION OF THE WISCONSIN PERSONALITY DISORDERS INVENTORY WITH THE SCID-II AND PDE

Jacques P. Barber and Jennifer Q. Morse

One hundred and sixteen patients who were diagnosed using either the SCID-II (n = 64) or the PDE (n = 52) also filled out the Wisconsin Personality Disorders Inventory (WISPI). Adequate internal consistency was found for the WISPI subscales using raw scores. Three- to four-month test-retest reliability coefficients of the WISPI had a median correlation of .75. Five out of 11 SCID-II dimensional scores and 5 out of the 11 PDE dimensional scores for personality disorders demonstrated good convergent and discriminant validity with their respective WISPI scales. On both structured interviews, good convergent and discriminant validity between the structured interviews and the WISPI were found for avoidant, obsessive-compulsive, and borderline personality disorders.

Following the introduction of the multiaxial organization of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III—American Psychiatric Association, 1980), a surge occurred in the development of measures of personality disorders. Due to the effort and time involved in using any structured clinical interview, the possibility of assessing personality disorders by means of self-report inventories seems an attractive one. Regardless of whether self-report questionnaires could replace clinical interviews for the purpose of deriving a categorical diagnosis (Perry, 1992), self-report inventories are likely to have a role in measuring the severity of the disorders as well as providing researchers with instruments to record temporal fluctuations of those disorders.

From the Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania School of Medicine.

The authors want to thank Drs. Leigh McCullough and Karla Moras for training and supervising the diagnosticians and also to thank Dr. Jennifer Jones for organizing the data management for the Center for Psychotherapy Research. We also want to thank Dr. Crits-Christoph for overall support.

Part of this paper was presented at the International Meeting of the Society for Psychotherapy Research in Pittsburg, PA. The writing of this paper was supported in part by RO1 MH-49902 to Jacques P. Barber and by the NIMH Clinical Research Center grant MH 45178 to Paul Crits-Christoph. Address all correspondence to: Jacques P. Barber, Room 704, Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania, 3600 Market Street, Philadelphia, PA 19104-2648.

The agreed upon revised nosology of personality disorders for DSM-III-R is the result of empirical findings and of compromises among theorists of various orientations. The present nosology is not derived from any theoretical scheme. Many authors view the adherence to a strict empirical approach a compelling advantage of DSM-III. In Epstein's words (1987), "an empirical approach is clearly preferable to one that is based on poor theory" (p. 108). Others argue that without an organizing scheme, the classification of personality disorders is a doubtful enterprise (e.g., Cloninger, 1987; Costa & McCrae, 1992; McLemore & Brokaw, 1987).

Researchers' views concerning the importance of theory in guiding the organization of the nosology of personality disorders are also reflected in the way they develop self-report inventories to assess those disorders. Self-report inventories have been constructed by translating each DSM criteria into a self-report item, as in the Personality Diagnostic Questionnaire-Revised (PDQ-R—Hyler & Rieder, 1987), or by using items derived from a theory that comprehensively explains the current psychiatric nosology, as in the Millon Clinical Multiaxial Inventory II (MCMI-II—Millon, 1987). These two self-report questionnaires have already received much research attention. In this paper, we explore the validity of the Wisconsin Personality Disorders Inventory (WISPI—Klein, Benjamin, Rosenfeld, Treece, Husted, & Greist, 1993) that has been developed recently using a third approach combining some of the elements of both approaches.

Each WISPI item and scale content was derived directly from the definitions of the symptoms listed in the DSM and translated and reformulated according to a specific model of interpersonal theory—Benjamin's (1993) Structural Analysis of Social Behavior (SASB). Although the specific details vary, interpersonal approaches have generally included dimensions of affiliation and dominance as factors defining a circumplex space. Furthermore, as applied to personality disorders, these approaches have emphasized the importance of interpersonal relationships in the creation and maintenance of these disorders.

From a psychometric point of view, both the internal consistency and 2 week test-retest reliability of the WISPI were found appropriate. In terms of validity, Klein et al. (1993) showed that 8 of the 11 WISPI scales discriminate between normals and patients. The pattern of correlations with the PDQ indicated good convergent validity (average correlation of .69), but the correlations with the MCMI-I (Millon, 1982) were lower (average correlation of .39). The difference between the patterns of correlations with the PDQ and with the MCMI may have been due to the fact that the PDQ was developed to tap DSM-III constructs (as is the WISPI) whereas the MCMI was designed to tap Millon's more complex psychosocial theory (Klein et al., 1993). In addition the WISPI ratings were compared with clinician ratings on the Personality Assessment Form (PAF—Shea, Blass, Pilkonis, Watkins, & Docherty, 1987). The average uncorrected correlation between the WISPI and PAF scales was also moderate ($r = .34$; range from .15 to .42—Klein et al. 1993). Those results suggest that the WISPI may be an appropriate instrument in measuring personality disorders as defined by DSM-III. The present study therefore constitutes the next step in validating the WISPI,

that is, exploring the correspondence between specific personality disorder structured interviews and the WISPI.

Researchers interested in personality disorders differ not only in terms of their interests in theoretical guidance but also in how they conceptualize personality disorders. Soldz, Budman, Demby, and Merry (1993), for example, argue that the manner by which diagnostic agreement is assessed reflects how one conceptualizes personality disorders. If personality disorders are viewed as discrete entities, then one would use a categorical approach to assess agreement (e.g., kappa coefficients). If personality disorders are on a continuum with normality, then a dimensional approach to assess agreement is warranted (e.g., correlations). In the current paper, we follow this dimensional (quantitative) approach of measuring personality disorders that Widiger (1992) strongly recommended, and we examine dimensionally the relations between the WISPI and independent structured psychiatric interviews. Accordingly, we use the dimensional scores derived from the structured interviews. Those dimensional scores can be seen as a measure of the degree to which a person exhibits any one of the 11 personality disorders assessed in the study (e.g., Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991). Similarly, self-report questionnaires can easily be used to assess the degree of intensity of the presence of any personality disorder.

Because one of the assumptions of the construct of personality disorders is that they last over time, we also explore the test-retest reliability of the WISPI across a longer period than was previously done by Klein et al. (1993). In the present study we examine the stability of the WISPI scores over a 3- to 4-month period.

METHOD

Patients. Patients were recruited through a newspaper advertisement or had contacted the Center for Psychotherapy Research or the Center for Cognitive Therapy at the University of Pennsylvania. All callers were screened through a series of questions to examine if they were likely to be candidates for studies at the Center for Psychotherapy Research. If they were found likely to have a diagnosis of chronic depression (CD), general anxiety disorder (GAD), or obsessive-compulsive (OCPD) or avoidant personality disorder (AVPD), they were invited for an extensive intake interview including diagnostic studies for DSM-III-R Axes I and II. The interviews were conducted by experienced diagnosticians, mostly postdoctoral, clinical psychologists who had received extensive training with structured interviews. Following those interviews, 116 patients were accepted into different pilot studies, each diagnostic group receiving either cognitive or supportive expressive dynamic psychotherapy. The diagnoses reported in this study are based on the results from the structured interviews.

In the overall sample, the age of patients ranged from 21 to 62, with a mean of 37 (±8.9); 50% were women. The GAD group included 29 patients; 52% were women. Those patients met DSM-III-R criteria for GAD. Their age ranged from 21 to 50 with a mean of 35.4 (±8.6). The CD group included 34 patients; 41% were women. Those patients either met criteria for a 2-year-long major depressive episode or had a

diagnosis of dysthymia and current major depressive episode. Their age ranged from 22 to 56 with a mean of 40 (±8.9). The personality disorder (PD) group included 52 patients; 57% were women. Their ages ranged from 21 to 62, with a mean of 36.5 (±8.9). These patients received an Axis II diagnosis of either AVDP or OCPD based on either the SCID-II or the Personality Disorder Examination.

MEASURES

Wisconsin Personality Disorders Inventory (WISPI—Klein et al. 1993). Items of the WISPI are organized into 11 scales for each of the personality disorders. An additional 10 items make up the Marlowe Crowne Scale for social desirability response bias. Each item is rated on a 10 point scale (1 = never or not at all true of you; 10 = always or extremely true of you). Subjects are asked to rate their usual selves during the past 5 years or more. Summary scores for each scale are defined as the means of the ratings of the items for each scale, with missing items excluded. Scores are computed in three ways: raw scores, ipsatized scores, and *z*-scores. *Z*-scores are computed using the normals' data provided by Klein et al. (1993). All statistical analyses making use of WISPI subscale scores are based on *z*-scores.

Personality Disorder Examination (PDE—Loranger, 1988). The PDE, a structured clinical interview for Axis II disorders, was developed from 1983 to 1988. Its 126 questions are arranged under 6 domains or headings: work, self, interpersonal relationships, affects, reality testing, and impulse control. All positive answers are followed by requests for examples. Each item is scored as: 0, absent or normal; 1, exaggerated, accentuated, of uncertain clinical significance or limited in duration; 2, criterion level or pathological. All items are scored with the help of a detailed manual that defines the scope and meaning of each criterion, in addition to providing anchor points. From the ratings of each item, one derives, using a complex algorithm, probable and definite diagnoses for all DSM-III-R Axis II disorders including the ones in the appendix. Dimensional scores are calculated by summing the answers to all questions from each disorder. Interrater reliability of the earlier version was excellent, with intraclass correlations above .84 (Loranger, Susman, Oldham, & Russakoff, 1987). Diagnoses of sadistic and self-defeating personality disorders are not included in this report.

Structured Clinical Interview for DSM-III-R—Personality Disorders (SCID-II—Spitzer, Williams, Gibbon, & First, 1990). The SCID-II is a 113 item structured clinical interview for Axis II disorders. The items are organized on a disorder by disorder basis. Individual items or criteria are scored as: 0, absent; 1, subthreshold; or 2, threshold. All positive answers are followed by requests for examples. For each Axis II disorder, a specified number of items have to receive a threshold rating in order to derive a probable or definite diagnosis. As in the PDE, dimensional scores can be calculated by summing the ratings from each disorder scale, that is, adding together all subthreshold and threshold criteria scores from each individual item for each separate personality disorder. For this study, whether or not the SCID-II was preceded by the self-report questionnaire, all SCID-II items were inquired by the diagnostician.

Procedure. All patients accepted into psychotherapy pilot studies conducted at the Center for Psychotherapy are used for the present report. All patients received the Structured Clinical Interview for DSM-III-R Axis I (SCID-I) and a structured interview for deriving the major Axis I DSM-III-R diagnoses, and filled out the WISPI. The first 52 patients entered into the studies were interviewed using the PDE. Later on, the PDE was replaced with the SCID-II, which was administered to 64 additional patients at intake. Of these 116 patients, 63 completed the WISPI again 3 to 4 months later.

Table 1. Axis I Diagnoses Among the Patients with a Diagnosis of Personality Disorder (n = 52)

Number of patients	Diagnoses
12	Major depression recurrent
8	Major depression single
8	Depression NOS
14	Dysthymia
20	Social phobia
1	Panic disorder with agorophobia
3	Panic disorder without agorophobia
6	Simple phobia
6	Generalized anxiety disorder
2	Obesseive-compulsive disorder
4	Anxiety NOS
1	Substance dependence
1	Alcohol abuse
3	Other Axis I diagnoses

RESULTS

Description of the patient samples. The diagnoses on Axis I received by the patients in the PD group are presented in Table 1. Among those 52 PD patients, 32 received a primary diagnosis of avoidant personality disorder; 20, obsessive compulsive personality disorder. Thirty-one (60%) had an additional Axis II disorder, 12 had 2 additional Axis II disorders, 6 had 3 additional Axis II disorders, and 1 had 4 additional Axis II disorders.

Thirty-five percent of the CD patients were diagnosed as having at least one personality disorder by either structured diagnostic interview for personality disorders; only 20% of the GAD group received a diagnosis of at least one personality disorder (see Table 2). The most frequent Axis II diagnosis in both groups was avoidant personality disorder.

We also examined potential differences between patients who received the SCID-II versus those who received the PDE. We compared the two groups in terms of presence vs. absence of any depressive disorder, of any anxiety disorder, and of any other Axis I diagnoses and found no difference between the two groups (χ^2 = 0.433, df = 1, ns; χ^2 = 2.91, df = 1, ns; and, χ^2 = .32, df = 1, ns; respectively). Furthermore, the mean GAF score of the two groups

Table 2. Axis II Diagnoses Among Patients with a Diagnosis of Chronic Depression or Generalized Anxiety Disorder

	Chronic Depression (n = 34)	Generalized anxiety (n = 29)
Any PD Diagnoses	12	6
Avoidant PD	7	3
Dependent PD	3	1
Obsessive Compulsive PD	2	1
Paranoid PD	1	0
Self-Defeating PD	3	0
Passive Aggressive PD	1	1

Note. Axis II diagnoses that are not mentioned were not found in this sample.

Table 3. Cronbach Alpha Values for the WISPI Subscales Using Raw and Ipsatized Scores

Scale	Raw	Ipsatized
Avoidant	.94	.85
Dependent	.88	.74
Obsessive	.86	.71
Passive	.85	.71
Paranoid	.88	.50
Schizoid	.84	.69
Schizotypal	.81	.40
Histrionic	.85	.69
Borderline	.83	.55
Narcissistic	.83	.44
Antisocial	.88	.81

did not differ ($t = .08$, $df = 112$, ns) indicating that the two groups were fairly similar in overall level of psychiatric severity.

Reliability of the WISPI. (1) Internal Consistency. The internal consistency (Cronbach alpha) of the subscales of the WISPI ranged from .81 (schizotypal PD) to .94 (avoidant PD) when raw item scores were used (see Table 3). These coefficients are certainly adequate, although they are slightly lower than the ones reported by Klein et al. (1993). When ipsatized item scores were used, the range of alphas was somewhat lower: .40 (schizotypal PD) to .86 (avoidant PD).

(2) Test-retest Reliability. For 63 patients we had the data at intake and either 3 or 4 months later. For those patients, we calculated test-retest reliability coefficients (see Table 4). The mean and median stability coefficient was .75 (range from .69 to .80), indicating good stability across a relatively long period of time. Although the correlations between the two time points were large and significant, patients' level of "personality disor-

Table 4. WISPI *z*-Scores at Two Different Times, Test-Retest Coefficients and t-Tests Values

	z scores					
	Time 1 ($n = 63$)		Time 2		Test-retest ($n = 63$)	
Scale	M	SD	M	SD	r	T-test
Paranoid	-0.04	1.24	-0.46	1.20	.72****	3.647****
Schizoid	-0.08	1.32	-0.40	1.21	.80****	3.035***
Schizotypal	-0.31	0.76	-0.74	0.53	.69****	6.190****
Histrionic	-0.86	1.00	-1.05	0.94	.78****	2.261**
Narcissistic	-0.45	0.94	-0.83	0.95	.79****	4.917****
Antisocial	-0.55	0.67	-0.66	0.55	.75****	1.987
Borderline	0.05	1.18	-0.43	1.06	.71****	4.361****
Avoidant	0.94	1.32	0.41	1.40	.73****	4.219****
Dependent	0.14	1.27	-0.43	1.23	.74****	4.997****
Compulsive	0.72	1.25	0.09	1.36	.78****	5.763****
Passive	0.02	1.18	-0.39	1.17	.77****	4.046****
Mean and median r					.75	

**$p \leq .01$; ***$p \leq .001$; ****$p \leq .0001$.

derness" decreased significantly over time for all the dimensional scores except for antisocial personality disorder (see Table 4).

Validity. From the structured interviews, we derived three different scores for each personality disorder: (1) "categorical diagnosis"—a score reflecting the presence, subthreshold, or absence of the categorical diagnosis of personality disorder; (2) "number of items true"—the number of criteria met for each diagnosis, that is, the number of items that were rated as true or present for each personality disorder category; and (3) "dimensional score"—the total of all ratings for a specific personality disorder. Although both items true and dimensional scores have been used in the literature as a dimensional representation of personality disorders, we thought that dimensional scores may be more sensitive as an index of severity (Skodol et al., 1991).

As an index of correspondence, each WISPI subscale *z*-score was correlated separately with each of those three different scores from the structured interviews (Table 5). Not surprisingly, the correlation coefficients between the WISPI and the number of items true or between the WISPI and the dimensional scores did not differ much (columns 2 and 5 from sections B and C of Table 5, respectively); thus we emphasize the dimensional scores in the rest of the paper. The percentage columns (%) in Table 5 present the base rates for each of the disorders based on the structured interview. More specifically, in the percentage columns in section A of the table, we present the percentage of patients who had either a subthreshold or threshold diagnosis; in section B, the percentage of patients who had at least one item true and in section C, the percentage of patients who had dimensional scores higher than zero. We now summarize the results from Table 5.

Categorical Diagnoses (Section A). Correlations between the WISPI subscales and the variable representing the full or subthreshold presence or the absence of the respective diagnoses yielded the following results:

1. On the SCID-II the correlations between the diagnostic variable and the WISPI scores averaged .28 (after being translated into *z* scores and back into *r*, and excluding the five correlations that could not be computed because of lack of variance due to the fact that patients did not receive those diagnoses). The correlations ranged from -.02 (paranoid PD) to .50 (avoidant PD). Four of the six correlations were significant (see Table 5, section A).
2. The correlations between the WISPI and the PDE ranged from -.11 (paranoid PD) to .49 (avoidant PD) and their average was .19. Out of the nine correlations computed with the PDE, only two were significant (see Table 5).

Dimensional Scores (Table 5, Section C). Using the dimensional scores from the semistructured interviews, better correlations with the WISPI subscale scores were obtained:

1. SCID-II: Convergent validity coefficients between the WISPI subscales and the dimensional scores on the SCID-II averaged .44 and ranged from

Table 5. Correlations of WISPI Subscales with the Corresponding Subscales of the SCID-II and PDE at the Level of the Categorical Diagnoses, of the Number of Items True and of Dimensional Scores

	SCID-II (N = 64)			PDE (N = 52)		
	Correlation	%[a]	p	Correlation	%[a]	p
A. Categorical Diagnoses[b]						
Avoidant	.50	41	.000	.49	31	.000
Dependent	.47	6	.000	.24	2	.084
Obsessive	.33	19	.008	.18	33	NS
Passive	.27	6	.032	.43	2	.001
Paranoid	-.02	2	NS	-.11	2	NS
Schizoid	—	—	—	.18	2	NS
Schizotypal	—	—	—	—	—	—
Histrionic	—	—	—	.09	2	NS
Borderline	—	—	—	—	—	—
Narcissistic	.06	3	NS	.13	4	NS
Antisocial	—	—	—	—	—	—
B. Numbers of items true						
Avoidant	.58	83	.000	.54	87	.000
Dependent	.51	98	.000	.45	79	.001
Obsessive	.58	77	.000	.39	87	.005
Passive	.37	59	.003	.54	58	.000
Paranoid	.39	28	.002	.00	56	NS
Schizoid	.26	31	.038	.29	29	.038
Schizotypal	.49	45	.000	.21	48	NS
Histrionic	.30	31	.016	.21	31	NS
Borderline	.55	56	.000	.50	60	.000
Narcissistic	.21	47	.093	.31	42	.028
Antisocial	.18	14	NS	.51	20	.001
C. Dimensional Score						
Avoidant	.65	95	.000	.58	90	.000
Dependent	.49	91	.000	.51	85	.000
Obsessive	.59	95	.000	.40	96	.003
Passive	.46	83	.000	.54	75	.000
Paranoid	.43	58	.000	.11	79	NS
Schizoid	.40	61	.001	.36	52	.010
Schizotypal	.51	58	.000	.18	62	NS
Histrionic	.29	66	.020	.22	60	NS
Borderline	.61	81	.000	.47	79	.001
Narcissistic	.15	73	NS	.38	60	.006
Antisocial	.24	33	NS	.39	43	.014

[a] The base rate or the percentage of patients who had nonzero values for the variable presented. For the categorical diagnoses which are categorical variables, the percentage of patients who had nonthreshold or threshold diagnoses. For the number of items true, the percentage of patients who had at least one item true on the SCID-II or PDE specific diagnostic category. For the dimensional scores, the percentage of patients with nonzero dimensional scores on the structured interviews.
[b] Diagnosis is either absent, subthreshold, or threshold.

.15 (narcissistic PD) to .65 (avoidant PD). Nine out of the 11 correlations were significant (see Table 5, section C).

2. PDE: The correlations with the PDE averaged .39 and ranged from .11 (paranoid PD) to .58 (avoidant PD). Eight out of the 11 correlations between the WISPI subscales and respective scales of the PDE were significant, indicating good convergent validity.

Table 6. Correlations Between the WISPI subscales and the Dimensional Scores of the SCID-II at Time 1

	SCID-II										
WISPI	AVO	DEP	OCP	PAS	PAR	STY	SCI	HIS	NAR	BOR	ASP
AVO	.65	.25	.08	.13	.27	.42	.34	-.04	.18	.29	.05
DEP	.25	.49	.09	.07	.14	.13	.01	.21	.10	.28	.00
OCP	.18	.35	.59	.23	.19	.20	.25	.03	-.07	.02	-.16
PAS	.20	.48	.36	.46	.32	.21	.22	.08	.08	.30	.04
PAR	.40	.38	.29	.28	.43	.34	.19	.06	.16	.41	.06
STY	.34	.48	.30	.20	.29	.51	.32	.02	.06	.32	.12
SCI	.41	.20	.15	.13	.26	.38	.40	.01	.14	.36	.00
HIS	.18	.47	.33	.26	.10	.12	-.02	.29	.09	.37	.14
NAR	.40	.48	.26	.33	.41	.34	.15	.11	.15	.38	.07
BOR	.24	.57	.25	.24	.28	.26	.22	.20	.18	.61	.10
ASP	.21	.40	.24	.27	.10	.19	.09	.30	.08	.42	.24

Note:, AVO = avoidant PD, DEP= dependent PD; OCP = obsessive-compulsive PD; PAS = passive-aggressive PD, PAR = paranoid PD, STY = schizotypal PD, SCI = schizoid PD, HIS = histrionic PD, NAR = narcissistic PD, BOR = borderline PD, ASP = Antisocial PD. Coefficients on the diagonal that are underlined are the largest coefficients for both the WISPI and SCID-II.

Divergent and Convergent Validity. To examine the degree of convergent and divergent validity, the pattern of correlations among the similar subscales of the different instruments and between those subscales and other subscales was examined (e.g., Trull, 1993). Nine out of 11 SCID-II scales had their highest correlations with their respective WISPI scales (except for dependent and narcissistic; see Table 6). The dependent scale of the SCID-II correlated highly with many WISPI scales. Six out of 11 WISPI scales had their highest correlations with their respective SCID-II scales (except for passive aggressive, schizoid, histrionic, narcissistic, and antisocial). On the SCID-II, therefore, good convergent and discriminant validity were obtained for the avoidant, obsessive-compulsive, paranoid, schizotypal, and borderline personality disorders (those coefficients are underlined in Table 6).

In regard to the PDE, 6 out of the 11 PDE scales had their highest correlations with their respective WISPI scales (avoidant, dependent, obsessive, passive-aggressive, borderline, and narcissistic; see Table 7). In contrast to the SCID-II, the PDE dependent PD scale did not correlate highly with other WISPI scales. Six out of the 11 WISPI scales had their highest correlations with their respective PDE scales (avoidant, dependent, obsessive, passive-aggressive, borderline, and antisocial). Thus, the PDE demonstrated good convergent and discriminant validity for the avoidant, dependent, obsessive-compulsive, passive-aggressive, and borderline personality disorders (those coefficients are underlined in Table 7).

Profile Correlations. Another way of looking at the data is to examine whether the self-report and interview methods detect similar people. Similar individuals have scores that have a similar shape on the 11 dimensions of "personality disorderness" on both the WISPI and the structured interview administered to them. In other words, we assessed whether patients had similar profiles on the two instruments. One advantage of examining the

Table 7. Correlations Between the WISPI Subscales and the Dimensional Scores of the PDE

	PDE										
WISPI	AVO	DEP	OCP	PAS	PAR	STY	SCI	HIS	NAR	BOR	ASP
AVO	.58	.33	.13	-.03	.26	.38	.45	.07	.16	.41	.15
DEP	.33	.51	.06	.18	.03	.14	.13	.23	.13	.18	.40
OCP	-.11	-.13	.40	.36	.03	.08	.00	.29	.33	.30	-.03
PAS	-.12	-.06	.30	.54	.03	.15	-.06	-.01	.34	.18	.32
PAR	.22	.07	.16	.13	.11	.20	.17	.11	.22	.40	.35
STY	.17	.03	.22	.31	.08	.18	.20	-.03	.31	.31	.35
SCI	.16	-.21	.20	.21	.09	.20	.36	-.01	.27	.37	.16
HIS	-.07	.02	.25	.17	-.06	-.11	-.19	.22	.23	.19	.32
NAR	.16	.05	.34	.18	.03	.15	.08	.36	.38	.39	.24
BOR	.22	.06	.24	.22	.06	.11	-.02	.23	.28	.47	.34
ASP	.09	.01	.07	.33	.09	.17	.03	-.08	.21	.26	.39

Note:, AVO = avoidant PD, DEP= dependent PD; OCP = obsessive-compulsive PD; PAS = passive-aggressive PD, PAR = paranoid PD, STY = schizotypal PD, SCI = schizoid PD, HIS = histrionic PD, NAR = narcissistic PD, BOR = borderline PD, ASP = Antisocial PD. Coefficients on the diagonal that are underlined are the largest coefficients for both the WISPI and PDE.

correspondence of profiles is that they enable one to consider simultaneously all 11 dimensions. Pearson moment correlations were computed as a measure of profile similarity (see Nunally, 1978) between each participant's scores on the self-report inventory and the dimensional scores of the structured clinical interview. After computing the correlation for each patient, those coefficients were translated into *z*-scores and averaged across subjects using Fisher's *r* to *z* transformation, resulting in an estimate of the average correlation of subjects' similarity of profile.

The individual profile similarity between the SCID-II and the WISPI ranged from -.55 to .95, with a mean correlation of .53 (20 out of 64 patients had a correlation of less than .40). Similar results were obtained when correlating PDE and WISPI profiles; those correlations ranged from -.02 to .89, with a mean of .57. (14 out of 51 patients had a correlation of less than .40). These correlations indicate that the overall shapes of SCID-II and WISPI, on one hand, and of PDE and WISPI, on the other hand, were fairly similar.

DISCUSSION

As in Klein et al.'s (1993) original sample, the WISPI scales demonstrated good internal consistency. The internal consistency of the subscales was reduced when we used ipsatized scores instead of raw scores. One reason for lower coefficients among ipsatized scores might be their reduced variance. It remains to be determined, however, whether the reduction in internal consistency due to the use of ipsatized scores is specific to our sample, to the WISPI, or to the use of ipsatized scores in general.

The coefficients for test-retest reliability over the 3- to 4-month period were adequate and compared well with Trull's (1993) findings on the stability of the MMPI-Personality Disorder scales over a 3-month period. On average, the coefficients obtained in the present study were somewhat better than the ones on the PDQ-R (Trull, 1993). The fact that the dimen-

sional levels decreased with elapsed time is consistent with the fact that those patients received treatment and were more likely, on average, to be less symptomatic after 3 or 4 months.

Results of the study indicate that the conceptualization of personality disorders as measured by the WISPI and by structured interviews, such as the SCID-II or PDE, converge in this sample of outpatients. With the dimensional scores of the SCID-II, we found significant convergent validity coeffficient for 9 of the 11 personality disorders scales. With the dimensional scores from the PDE, we found significant convergent validity coeffficients for 8 of the 11 scales.

Furthermore, the WISPI scales demonstrated convergent and divergent validity with 5 personality disorders measured dimensionally on the SCID-II scales, and with 5 dimensional scores from the PDE. Good convergent and divergent validity was found for avoidant, obsessive compulsive and borderline personality disorders on both structured interviews. The present results with the WISPI are more encouraging than those reported by Hunt and Andrews (1992), who found poor convergence between the PDQ-R and the PDE in a group of Australian anxiety patients and controls. Soldz et al. (1993), however, reported convergent coefficients between the MCMI-II and the PDE that are similar in magnitude to the ones obtained in the present study. Nevertheless, Soldz et al. (1993) found relatively weak divergent validity. Applying the criteria used in the present study to Soldz et al.'s results, only 2 out of the 11 same scales (borderline and avoidant) showed good convergent and divergent validity between the PDE and the MCMI-II.

In summary, the WISPI has shown good convergent and divergent validity in regard to structured clinical interviews. It remains to be seen whether those relatively good coefficients will be replicated in future studies using a more heterogeneous sample of patients. Future studies are also required to address the development of sophisticated algorithms and criteria for deriving categorical diagnoses from the dimensional scores of the WISPI. Using a large data set one could develop algorithms that could be partly derived from empirical data. For example, one could determine which level on each WISPI subscale yields an acceptable hit rate using one of the structured interviews as a gold standard. Once those criteria are available, one would be able to cross-validate the degree of categorical concordance between the structured interviews and the WISPI. As Loranger (1992) pointed out, if such studies show that self-report questionnaires are associated with a "tolerable level of false-negative cases" (p. 323), they could be used as economical screening instruments. If researchers are interested in using self-report for screening purposes, our results are a first step toward encouraging the development of specific and psychometrically sound criteria for deriving categorical diagnoses from the WISPI dimensional scores.

Widiger (1992), however, summarized research evidence indicating that categorical diagnoses of personality disorders have many limitations such as arbitrary thresholds for defining a diagnosis, and that these limitations may be the cause for poor interjudge reliability. Structured interviews do not seem to have resolved this difficulty. For example, Hyler, Skodol, Kellman, Oldham, and Rosnick (1990) found low levels of agreement

between the SCID-II and the PDE when computed at the categorical level but found higher correlations when used dimensionally (Skodol et al., 1991). The correlations between the dimensional scores of the two structured interviews were somewhat higher than the ones obtained in the current study between a self-report and a structured interview most likely because in the present study we had a heteromethod design whereas Hyler et al. (1990) used the same method of collecting data (i.e., structured interviews). Together, these findings provide additional support for using dimensional scores from both structured interviews and self-report questionnaires to examine concordance between those two methods of collecting data.

Because of the lack of a gold standard in the diagnosis of personality disorders (e.g., Perry, 1992), we are unable to determine whether the WISPI or the diagnostic interviews are more accurate in detecting personality disorders. In that regard, it is interesting that dimensional scores may show more concordance than categorical scores.

REFERENCES

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

Benjamin, L. S. (1993). *Interpersonal diagnosis and treatment of personality disorders*. New York: The Guilford Press

Cloninger, C. R. (1987). A systematic method for clinical description and classification of personality variants: A proposal. *Archives of General Psychiatry, 44*, 573-588.

Costa, P. T., & McCrae, R. R. (1992). The five factor model of personality and its relevance ot personality disorders. *Journal of Personality Disorders, 6*, 343-359.

Epstein, S. (1987). The relative value of theoretical and empirical approaches for establishing a psychological diagnostic system. *Journal of Personality Disorders, 1*, 100-109.

Hunt, C., & Andrews, G. (1992). Measuring personality disorder: The use of self-report questionnaires. *Journal of Personality Disorders, 6*, 125-133.

Hyler, S., & Rieder, R. (1987). *PDQ-R personality questionnaire*. New York: New-York State Psychiatric Institute.

Hyler, S., Skodol, A., Kellman, D., Oldham, J., & Rosnick, L. (1990). Validity of the personality diagnostic questionnaire: Development and preliminary results. *American Journal of Psychiatry, 147*, 1043-1048.

Klein, M. H., Benjamin, L. S., Rosenfeld, R., Treece, C., Husted, J. & Greist, J. H. (1993). The Wisconsin personality disorders inventory: Development, reliability and validity. *Journal of Personality Disorders, 7*, 285-303.

Loranger, A. W. (1988). *Personality disorder examination (PDE) manual*. Yonkers, NY: DV Communications.

Loranger, A. W. (1992). Are current self-report and interview measures adequate for epidemiological studies of personality disorders? *Journal of Personality Disorders, 6*, 313-325.

Loranger, A. W., Susman, V. L., Oldham, J. M., & Russakoff, L. M. (1987). The personality disorder examination: A preliminary report. *Journal of Personality Disorders, 1*, 1-13.

McLemore, C. W., & Brokaw, D. W. (1987). Personality disorders as dysfunctional interpersonal behavior. *Journal of Personality Disorders, 1*, 270-285.

Millon, T. (1982). *Manual for the Millon clinical multiaxial inventory (MCMI)*. Minneapolis, MN: National Computer Systems.

Millon, T. (1987). *Manual for the Millon clinical multiaxial inventory II (MCMI-II)*. Minneapolis, MN: National Computer Systems.

Nunally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill Book.

Perry, J. C. (1992). Problems and considerations in the valid assessment of person-

ality disorders. *American Journal of Psychiatry, 149*, 1645-1652.

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

Soldz, S., Budman, S., Demby, A. & Merry, J. (1993). Diagnostic agreement between the personality disorder examination and the MCMI-II. *Journal of Personality Assessment, 60*, 486-499.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. (1990). *Structured clinical interview for DSM-III-R—Personality disorders*. Washington, DC: American Psychiatric Press.

Trull, T. J. (1993). Temporal stability and validity of two personality disorder inventories. *Psychological Assessment, 5*, 11-18.

Widiger, T. A. (1992). Categorical versus dimensional classification: Implications from and for research. *Journal of Personality Disorders, 6*, 287-300.