Journal of Personality Disorders, 24(6), 721–745, 2010

COMPARING PERSONALITY DISORDER MODELS: CROSS-METHOD ASSESSMENT OF THE FFM AND DSM-IV-TR

Douglas B. Samuel, PhD, and Thomas W. Widiger, PhD

The current edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association, 2000) defines personality disorders as categorical entities that are distinct from each other and from normal personality traits. However, many scientists now believe that personality disorders are best conceptualized using a dimensional model of traits that span normal and abnormal personality, such as the Five-Factor Model (FFM). However, if the FFM or any dimensional model is to be considered as a credible alternative to the current model, it must first demonstrate an increment in the validity of the assessment offered within a clinical setting. Thus, the current study extended previous research by comparing the convergent and discriminant validity of the current DSM-IV-TR model to the FFM across four assessment methodologies. Eighty-eight individuals receiving ongoing psychotherapy were assessed for the FFM and the DSM-IV-TR personality disorders using self-report, informant report, structured interview, and therapist ratings. The results indicated that the FFM had an appreciable advantage over the DSM-IV-TR in terms of discriminant validity and, at the domain level, convergent validity. Implications of the findings and directions for future research are discussed.

The current version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association [APA], 2000) represents "the categorical perspective that Personality Disorders are qualitatively distinct clinical syndromes" (p. 689). The DSM-IV-TR contains ten personality disorder (PD) categories arranged into three hierarchical clusters: Odd-Eccentric (paranoid, schizoid, and schizotypal), Dramatic-Emotional (antisocial, borderline, histrionic, and narcissistic), and Anxious-Fearful (avoidant, dependent, and obsessive-compulsive). Much of the theory and research is focused on the ten PDs, but a considerable body of research is also focused on the three clusters (e.g., Aluja, Cuevas, Garcia, & Garcia, 2007; Ehrensaft, Cohen, & Jonson, 2006; Lenzenweger & Willett, 2007).

From Yale School of Medicine (D. B. S.); and University of Kentucky (D. B. S., T. A. W.).

Address correspondence to Douglas B. Samuel, VA Connecticut Health Care—151D, 950 Campbell Avenue, Building 35, West Haven, CT 06516; E-mail: douglas.samuel@yale.edu

Many deficiencies of this categorical system have been noted, including: (a) Excessive diagnostic co-occurrence, (b) inadequate coverage, (c) excessive heterogeneity within diagnostic categories, (d) lack of a meaningful or well-validated boundary between normal and disordered personality, (e) questionable temporal stability, and (f) inadequate scientific foundation (Clark, 2007; Trull & Durrett, 2005; Widiger & Samuel, 2005a; Widiger & Trull, 2007). Excessive diagnostic co-occurrence is perhaps the most central of these problems since it impacts differential diagnosis. A fundamental purpose of the diagnostic manual is to help clinicians identify specific disorders and make treatment recommendations (Frances, First, & Pincus, 1995). It is evident that the DSM-IV-TR has not been particularly successful in this regard (Kupfer, First, & Regier, 2002) as the substantial diagnostic co-occurrence among the ten DSM-IV-TR PD diagnoses has been widely replicated (Bornstein, 1998; Clark, 2007; Livesley, 2003; Trull & Durrett, 2005). Suffice it to say, the maladaptive personality functioning of patients does not appear to be described adequately by just one DSM-IV-TR personality disorder and it is conceivable that this diagnostic co-occurrence also implies problematic discriminant validity.

AN ALTERNATIVE MODEL

It is acknowledged in the text of DSM-IV-TR that "an alternative to the categorical approach is the dimensional perspective that Personality Disorders represent maladaptive variants of personality traits that merge imperceptibly into normality and into one another" (American Psychiatric Association, 2000, p. 689). One of the more heavily researched dimensional alternatives is the five-factor model of personality (FFM; McCrae & Costa, 2003). The FFM was developed as a model of general personality functioning and consists of five bipolar dimensions (i.e., neuroticism vs. emotional stability, extraversion vs. introversion, openness vs. closedness to experience, agreeableness vs. antagonism, and conscientiousness vs. undependability). These five broad domains have been further differentiated into underlying facets (e.g., the facets of agreeableness include trust vs. mistrust, compliance vs. aggression, altruism vs. exploitation, tender-mindedness vs. tough-mindedness, straightforwardness vs. deception, and modesty vs. arrogance) by Costa and McCrae (1992, 1995) on the basis of their development of and research with the NEO Personality Inventory-Revised (NEO PI-R).

The FFM is a robust dimensional model that has succeeded in representing alternative models of personality and diverse collections of traits within an integrative, hierarchical structure (Ozer & Reise, 1994). The past two decades have seen a substantial body of research that has examined the relationship between the FFM and the DSM PDs (Widiger & Costa, 2002). Reviews of this research (Clark, 2007; Livesley, 2001), meta-analyses (Samuel & Widiger, 2008; Saulsman & Page, 2004), and an interbattery factor analysis of 20 previously published data sets that examined

relations between the FFM and the personality disorders (O'Connor, 2005), have all led to the conclusion that there are strong and robust links between the DSM-IV-TR personality disorders and the FFM dimensions of general personality.

However, much of this research has been confined to whether the DSM-IV-TR PDs can be understood from the perspective of the FFM, rather than whether the FFM can provide a more valid model of classification. Clark (2007) and Rounsaville and colleagues (2002) have suggested explicitly that it is useful and important to compare the validity of alternative dimensional models, including the FFM, with the existing diagnostic nomenclature. The current study uses a clinical sample to compare the DSM-IV-TR and the FFM with respect to convergent and discriminant validity across four assessment methodologies.

Blacker and Endicott (2000) suggested that an essential validity test for any psychiatric nomenclature is the agreement of descriptions across various assessment methodologies. There are four methods by which personality difficulties are typically assessed and the most common in general clinical practice is the unstructured clinical interview (Watkins, Campbell, Nieberding, & Hallmark, 1995). Clinicians and researchers also use self-report personality inventories and/or semi-structured interviews (Clark & Harrison, 2001; Widiger & Samuel, 2005b), with the latter method being the preferred approach in research (Rogers, 2001; Zimmerman, 2003). A final method recommended in the text of DSM-IV-TR (American Psychiatric Association, 2000, p. 686) is to have a knowledgeable informant provide a collateral report.

A number of studies have investigated the convergence of the DSM personality disorders across these methods. Researchers have commented on the inadequate convergent validity that is generally obtained between self-report and informant reports (Clark, 2007; Klein, 2003; Klonsky, Oltmanns, & Turkheimer, 2002; Oltmanns & Turkheimer, 2006) as well as between clinicians' unstructured assessments and semi-structured interview assessments (Garb, 2005; Segal & Coolidge, 2007; Widiger & Samuel, 2005b). There is also an existing literature concerning the convergence of the FFM, at least with respect to informant and self reports (McCrae et al., 2004; Ready & Clark, 2002). However, there is very little research on cross-method assessments involving a semi-structured interview of the FFM (Trull et al., 1998). Finally, we are aware of only one study (Piedmont & Ciarrocchi, 1999) that has compared FFM ratings obtained from clinicians with any other method of FFM assessment. However, even this study did not explicitly assess the FFM via clinician report, but instead relied on FFM markers (John, 1990) scored from the Adjective Checklist (Gough & Heilbrun, 1983) that had been completed by the counselor. There have been a few studies in which clinicians have provided explicit FFM ratings of patients (e.g., Blais, 1997; Samuel & Widiger, 2006; Sprock, 2002) but no data on the validity of these assessments have been published.

Only two studies have provided direct comparisons of the cross-method

convergent and discriminant validity of the FFM and the DSM within the same study (i.e., Ball, Rounsaville, Tennen, & Kranzler, 2001; Miller, Pilkonis, & Clifton, 2005). Miller, Pilkonis, and colleagues (2005) collected FFM and DSM PD ratings from a group of 69 psychiatric patients and a nominated informant. They found that the convergent correlations for the FFM domains rated ranged from .23 (agreeableness) to .71 (openness) with a median correlation of .43 across all the facets of the NEO PI (Costa & McCrae, 1992). The correlations between ratings of the eight DSM-IV-TR PDs rated were somewhat higher, ranging from .37 (avoidant) to .69 (antisocial) with a median convergent value of .51 (Miller, Pilkonis, et al., 2005). Although these results seem to indicate higher convergence for the DSM-IV-TR than for the FFM, they should be interpreted cautiously as the informants provided these DSM-IV-TR ratings using a mixed methodology. While the FFM data were collected exclusively via self-report, the DSM-IV-TR PD ratings were collected from informants using both a questionnaire and a structured interview procedure (Pilkonis et al., 1995).

Ball and colleagues (2001) provided the only direct comparison of the convergence for the FFM and DSM PDs using identical assessment methodologies. In a sample of substance abuse outpatients they reported a mean convergent correlation between a self-report and an informant report that was similar for both the FFM ($r = .31$) and the DSM ($r = .29$). The authors concluded that the results "did not provide strong evidence for the superiority of the NEO-FFI trait domains over SCID-II PQ personality disorder symptom severity with regard to internal consistency, temporal stability, or self-informant agreement" (Ball et al., 2001, p. 351). However, even this effort was limited by the fact that the FFM self-report measure was the NEO Five Factor Inventory (NEO-FFI), an abbreviated form that only measures the five broad domains of the FFM (Costa & McCrae, 1992). Considering research indicating that the facet-level of the FFM is necessary to adequately differentiate PD pathology (Axelrod, Widiger, Trull, & Corbitt, 1997; Reynolds & Clark, 2001), further research on the convergent validity of informant ratings should be conducted with a more detailed, facet-level measure of the FFM.

In sum, it is important that any alternative dimensional model proposed to replace the current nomenclature demonstrate superior convergent and discriminant validity across methods of assessment. Nevertheless, the only studies to date that have provided a direct comparison have suggested that the FFM has convergent validity that is either no better than (Ball et al., 2001) or perhaps even worse (Miller, Pilkonis, et al., 2005) than the current DSM system. However, in one of these two studies the assessment of the DSM-IV-TR nomenclature was itself aggregated across two methods (Miller, Pilkonis, et al., 2005) and the other study was confined to an assessment of the FFM domains (Ball et al., 2001). Additionally, neither study compared the two models with respect to discriminant validity and both were confined simply to self-report and informant report. The current study extends this previous research by (1) directly comparing the

discriminant as well as the convergent validities of the FFM and the DSM-IV systems within a clinical sample, (2) considering the FFM at both the domain and facet levels as well as the DSM-IV-TR at both the cluster and individual disorder levels, and (3) assessing the models not only with self and informant reports but also with respect to semi-structured interviews and ratings provided by their clinicians.

METHOD

Eighty-eight females receiving ongoing psychotherapy were recruited from the Lexington, Kentucky area. The women ranged in age from 19 to 60, with a mean age of 34.8 years ($SD = 8.8$; 3 participants failed to provide demographic information). They were primarily Caucasian (72.9%), with 23.5% indicating African-American, and 3 providing the response of other. Forty percent were single, 43.5% were divorced or widowed, and 16.5% were married or cohabitating. Their level of education ranged from junior high (20.2%) to a graduate degree (6.0%), with the modal participant indicating some college.

A total of 14 clinicians, who served as the primary therapists, provided ratings for 79 of the participants. The number of patients assessed by each clinician ranged from a low of one to a high of 18, with two being the median number. These clinicians were all female and predominantly Caucasian (78.6%), but two were Asian-American (14.3%) and one was African-American (7.1%). Level of training and experience varied considerably. Three had doctoral degrees (21.4%), eight had master's degrees (57.2%), and three were enrolled in graduate programs. Their experience ranged from a low of one year to a high of 21, with a mean of 4.2 years. The percentage of working time they spent providing clinical services ranged from 20% to 100%, with a mean of 53.2% of their work hours. All clinicians identified their theoretical orientation as cognitive, while 78.6% also listed behavioral, 57.1% interpersonal, 28.6% humanistic, and 21.4% psychodynamic. The mean length of treatment (all patients were engaged in individual therapy) at the time the ratings were provided was 4.9 months ($SD = 5.4$).

MATERIALS

Structured Interview for the Five-Factor Model (SIFFM; Trull & Widiger, 1997). The SIFFM assesses the 5 domains and 30 facets of the FFM using a series of guided questions. The SIFFM is the only existing interview measure of the FFM and has shown strong convergence with other measures of the FFM (Trull et al., 1998). Internal consistency in the current sample was good, with alphas ranging from a low of .74 (openness) to a high of .87 (conscientiousness), with a median value of .86.

Personality Disorder Interview—IV (PDI-IV; Widiger, Mangine, Corbitt, Ellis, & Thomas, 1995). The PDI-IV is a semi-structured interview that yields a dimensional rating for each of the 10 DSM-IV-TR PDs. A set of

three to four open-ended questions are used to assess the individual diagnostic criteria for each PD, for which the interviewer assigns a score of 2 (prototypic), 1 (threshold), or 0 (absent). For the purposes of the current study, we collapsed the scores to indicate whether the symptom was either present or absent. This was both to be more consistent with the scoring system used in other semi-structured interviews and to allow for a more straightforward interpretation of the findings as a presence versus absence of each criterion. In the current sample alphas ranged from .41 (schizoid) to a high of .72 (borderline), with a median of .58. While these values are somewhat low, they are consistent with the findings from studies employing other structured interviews for the PDs (e.g., Miller, Bagby, Pilkonis, Reynolds, & Lynam, 2005).

NEO Personality Inventory—Revised (NEO PI-R; Costa & McCrae, 1992). The NEO PI-R contains 240 statements to which the individual responds Strongly Disagree, Disagree, Neutral, Agree, or Strongly Agree (0–4 Likert-type scale). In the current sample alphas ranged from a low of .82 (extraversion) to a high of .89 (conscientiousness), with a median of .86. The informant version of the NEO PI-R is identical to the self-report version, except that the items are written in the 3rd person, rather than the first person (Form R; Costa & McCrae, 1992). The alphas for the informant version ranged from .78 (openness) to .94 (conscientiousness), with a median of .90.

Schedule of Nonadaptive and Adaptive Personality (SNAP; Clark, 1993). The SNAP consists of 375 statements to which the patient answers True or False and provides a self-report assessment of 3 primary temperaments (e.g., positive temperament), 12 maladaptive trait scales (e.g., exhibitionism), as well as the 10 personality disorders within DSM-IV. These scales provide a dimensional assessment of the PDs and range in length from 16 (schizoid) to 34 (antisocial) items. In the current sample, the SNAP PD scales obtained reasonable internal consistency, with exception of the obsessive-compulsive scale which had an alpha of .36. The other PDs ranged from .62 (schizoid) to .81 (paranoid), with an overall median of .70.

Clinician and Informant Rating Forms. Descriptions of the patient in terms of the DSM-IV-TR and FFM constructs were also obtained from the clinicians and informants. In order to be consistent with their routine unstructured clinical assessments (i.e., not informed by a supplementary or validated instrument) the clinicians' descriptions of the patients were confined to straightforward recordings of their current clinical understanding. For the DSM-IV-TR, the clinicians used a one-page rating form to rate the extent to which the patient exhibited characteristics for each of the ten PDs on a 1–5-point Likert scale (hereafter referred to as the DSMRF). The clinician version listed each of the ten DSM-IV PDs as well as a brief description of their central features drawn from the diagnostic manual (e.g., pattern of disregard for and violation of the rights of others, for antisocial PD). The informant version did not include the diagnostic labels (as these terms were not likely to be understood) and the descriptive features of each

disorder were altered to avoid jargon when necessary (e.g., they feel strong, fleeting emotions and seek to be the center of attention, for histrionic), consistent with the informant methodology of Oltmanns and Turkheimer (2006). Because each PD was assessed by only a single-item on the DSMRF, internal consistency statistics could not be computed. However, in a previous study, reliability of clinicians' ratings using the same rating form was adequate with a mean intraclass correlation of DSMRF profiles across clinicians of .61 (Samuel & Widiger, 2006).

For the FFM, clinicians and informants were provided a comparable one-page rating form that asked the rater to describe the patient on the 30 facets of the FFM using a 1–5 Likert-type scale (hereafter referred to as the Five Factor Model Rating Form [FFMRF]). Two adjective descriptors were included at both poles of each facet. The internal consistencies for the therapists' FFMRF ratings had a median of .78, but ranged from a low of .61 (neuroticism) to a high of .83 (conscientiousness). The informant ratings showed better overall internal consistency with four of the domains over .74 and a median of .80. However, the alpha value for openness was low, with a value of .39.

In total, there were nine measures describing each participant. All of these methods of description were blind to each other (e.g., the clinicians and interviewers were not aware of the self-report or informant ratings). The DSM-IV-TR measures included the PDI-IV (interview), SNAP (self-report), and DSMRF (clinician and informant). The FFM measures included the SIFFM (interview), NEO PI-R (self and informant), and FFMRF (clinician and informant).

PROCEDURE

Eighty-one of the 88 patients and ten of the therapists were recruited from a residential substance-abuse treatment facility for women. An additional seven patients and four clinicians were obtained from outpatient clinics within the community. None of the participating clinicians were affiliated with the research laboratory conducting the study. After providing written informed consent, these clinicians were asked to distribute solicitation flyers to all patients without a history of psychosis. Interested patients then contacted the experimenter to schedule an initial appointment, wherein the interviewer further explained the procedures and obtained written informed consent. Participants completed a packet, including a short demographic questionnaire as well as two self-report inventories (i.e., NEO-PI-R, SNAP). Following the completion of these inventories, participants completed the two semi-structured interviews administered by the primary investigator or other trained graduate students. Patients were compensated with $20 for their participation in the study.

Patients were also asked to designate an informant who knew them well to complete the other-report version of the NEO-PI-R, as well as informant versions of the FFMRF and *DSMRF*. Sixty-seven nominated informants

provided ratings on the participating patients (76.1% of the sample). These informants received $5 for their participation. Finally, the therapist for each participant completed a demographic questionnaire, a brief treatment history questionnaire (including all five axes of the *DSM* and a medication history), and provided ratings using the DSMRF and FFMRF. Therapists provided written, informed consent and received $50 for their effort.

The semi-structured interviews were audiotaped and selected sessions were coded by other interviewers to calculate interrater reliability. A total of five different interviewers read both interview manuals before the study began and received extensive training from an author of both instruments (i.e., T. A. Widiger). Throughout data collection, interviewers met weekly to discuss coding issues, ensure uniformity, and prevent rater drift. Sixteen of the SIFFM sessions were coded and interrater reliability was excellent at the domain level with Pearson correlations ranging from a low of .90 (openness) to a high of .99 (agreeableness and conscientiousness), with a median of .97. Agreement was also strong at the facet level with a median correlation of .94. The interrater reliability was lower, but still acceptable for the 18 PDI-IV sessions coded, ranging from a low of .57 (narcissistic) to a high of .92 (dependent) with a median of .83.

RESULTS

The prevalence rates of the DSM-IV-TR personality disorders, according to the PDI-IV, were examined to determine the level and type of pathology present in the sample. Although the disorders from the dramatic-emotional cluster predominated the sample, there was at least one individual that met criteria for each of the ten DSM-IV-TR PD diagnoses. However, it should be pointed out that for some of the disorders, only a single individual met criteria (i.e., narcissistic, schizoid). Overall, 57 individuals (67.9%) met criteria for any PD. Antisocial was the most prevalent PD with 30 individuals (35.7%) meeting criteria. Additionally, 19 individuals (22.6%) met criteria for borderline PD and 18 (21.4%) met criteria for avoidant. These findings are generally consistent with expectations for a population of female substance-abusers. Table 1 presents the overall distributions of the Axis II diagnoses based on the PDI-IV interview. Axis I disorders were not assessed via structured interview, but clinicians did provide chart diagnoses. The prevalence rates of Axis I disorders as provided by the clinicians are also presented within Table 1. It is clear from this table that substance-related disorders were prevalent. Seventy-three patients had a diagnosis and the modal participant had two substance-related diagnoses. The most prevalent were cocaine dependence (59.5%), alcohol dependence (43.0%), and opioid dependence (30.4%). Additional Axis I disorders included major depressive disorder (11.4%), bipolar disorder (10.1%, secondary diagnosis, by history), and posttraumatic stress disorder (7.6%).

TABLE 1. Axis I and II Diagnostic Information on Patient Sample

Personality Disorders	*N*	%	Mean # of criteria	Range	Substance-Related Disorders	Dependence *N*	Dependence %	Abuse *N*	Abuse %	Other Axis I Conditions	*N*	%
Paranoid	7	8.3%	1.6	6	Cocaine	47	59.5%	3	3.8%	Drug Induced Mood Disorder	1	1.3%
Schizoid	1	1.2%	0.6	5	Alcohol	34	43.0%	8	10.1%	Major Depressive Disorder	9	11.4%
Schizotypal	2	2.4%	1.2	6	Opioid	24	30.4%	2	2.5%	Bipolar Disorder	8	10.1%
Antisocial	30	35.7%	4.5	8	Cannabis	12	15.2%	10	12.7%	Dysthymic Disorder	2	2.5%
Borderline	19	22.6%	3.0	9	Sedative/Hypnotic/					Depressive Disorder NOS	2	2.5%
Histrionic	5	6.0%	1.9	7	Anxiolytic	16	20.3%	2	2.5%	Anorexia—Restricting	1	1.3%
Narcissistic	1	1.2%	1.0	6	Amphetamine	4	5.1%	1	1.3%	Posttraumatic Stress Disorder	6	7.6%
Avoidant	18	21.4%	2.0	7						Generalized Anxiety Disorder	2	2.5%
Dependent	8	9.5%	2.1	8						Obsessive-Compulsive Disorder	1	1.3%
Obsessive-Compulsive	8	9.5%	1.4	5						Kleptomania	1	1.3%
										Pica	1	1.3%

Notes. Axis II diagnoses are drawn from the PDI-IV Interview (Widiger et al., 1995), while the Axis I diagnoses are from the chart diagnoses provided by the clinicians. *N* = the number of patients with the diagnosis. % = the percentage of the sample with the diagnosis. NOS = Not otherwise specified.

CONVERGENT AND DISCRIMINANT VALIDITY

The FFM and DSM-IV-TR both provide hierarchical models of personality description. At the highest levels are the five domains of the FFM (i.e., neuroticism, extraversion, openness, agreeableness, and conscientiousness) and the three clusters of the DSM-IV-TR (i.e., odd-eccentric, dramatic-emotional, and anxious-fearful). Beneath this broad level of description are the 30 facets of the FFM (e.g., anxiousness and mistrust) and the ten PDs (e.g., avoidant and paranoid). Comparisons are perhaps most appropriate at similar levels of the hierarchy (e.g., domains of the FFM versus clusters of the DSM-IV-TR) but comparisons are also provided in the current study across levels (e.g., domains of the FFM versus the PDs of the DSM-IV-TR).

In order to provide an overall comparison of the convergent and discriminant validity of the FFM and DSM-IV-TR assessments the validity correlations were first averaged across all four methodologies. For example, the convergent validity correlation for the NEO PI-R self-report assessment of neuroticism was first obtained by averaging its correlations with the SIFFM, therapist FFMRF, and informant FFMRF assessments of neuroticism (the informant NEO PI-R assessments were excluded from this combined analysis). The overall FFM convergent validity was obtained by averaging these values across all five domains of the FFM. The overall discriminant validity coefficient was similarly obtained by averaging the discriminant validity coefficient of the NEO PI-R self-report assessment of neuroticism with all of the other domains of the FFM as assessed by all methods of assessment, and then averaging this value across the domains.

Table 2 provides these omnibus results across the DSM-IV-TR clusters and FFM domains as well as across the DSM-IV-TR PDs and FFM facets, for each of the four methods of assessment. This table presents mean convergent and discriminant values for each model that are method specific (averaged across variables). For example, the first column presents the mean correlation between the self-report instrument and each of the other

TABLE 2. Comparison of Convergent and Discriminant Validity Coefficients Across Methods and Models

	Self-Report		Interview		Therapist		Informant	
	Converg.	**Discrim.**	**Converg.**	**Discrim.**	**Converg.**	**Discrim.**	**Converg.**	**Discrim.**
DSM Clusters	0.21	0.16	0.23	0.16	0.14	0.16	0.16	0.21
FFM Domains	0.38	0.00	0.37	–0.03	0.23	–0.02	0.24	0.01
DSM Pds	0.22	0.10	0.23	0.09	0.14	0.10	0.15	0.15
FFM Facets	0.24	0.00	0.25	–0.01	0.15	–0.01	0.14	0.01

Notes. Converg. = Average convergent validity correlation of all variables from the indicated methodology with variables from all other methodologies. For instance, the convergent value for the interview method for the FFM facets is the average of 3 values (i.e., interview with self; interview with informant; interview with clinicians) for each facet, and then averaged again across the 30 facets. Discrim. = The average discriminant validity correlation of all variables from the indicated methodology with all other variables both within and across methodologies. DSM = Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000); FFM = Five Factor Model.

four methods. Specifically, for the FFM domains, the first column value (.38) is the average correlation between the NEO PI-R domains and those from the SIFFM (interview), FFMRF (clinician), and FFMRF (informant). It is evident that convergent validity was lower for the therapist and informant assessments than for the self-report and interview methods of assessment, for both the DSM-IV-TR and FFM. More importantly, convergent validity was higher for the FFM domain assessments than for the DSM-IV-TR cluster assessments, across each of the four methods. This finding still holds if one compares the assessments of the FFM domains with the assessment of the DSM-IV-TR personality disorders. However, there was no appreciable difference in convergent validity between the DSM-IV-TR and FFM assessments when the FFM was assessed at the level of the 30 facets. On the other hand, the FFM consistently obtained better discriminant validity at both the domain and facet levels across all four methods. In fact, the overall discriminant validity coefficients for the DSM-IV-TR clusters were actually higher than the overall convergent validity for both the therapist and informant ratings. This suggests that a given cluster rating by an informant or therapist is likely to relate just as strongly to a measure of a different cluster as it is to another measure of the same cluster.

Table 3 provides the averaged convergent validity coefficients and discriminant validity coefficients for the assessment of the three DSM-IV-TR clusters and the FFM domains for each pair-wise method comparison. It is evident from Table 3 that the weakest convergent validity across all methods was obtained by the therapist and informant assessments (for both the DSM-IV-TR and FFM), and the highest convergent validity was obtained between the self-report and semi-structured interview assessments. Whereas Table 2 presented method specific (and variable neutral) results, the final columns of Table 3 present mean convergent and discriminant correlations that are method neutral and variable specific. In other words, they include all possible method combinations and thus weight each method evenly, but provide specific values for each of the variables that make up the given model (e.g., the 3 clusters of the DSM). From these columns it is evident that each FFM domain obtained a higher convergent and lower discriminant validity coefficient than each DSM-IV-TR cluster. In fact, the averaged convergent validity values for the odd-eccentric and anxious-fearful clusters were no higher than their discriminant validity coefficients. However, it should be noted that convergent validity of the dramatic-emotional cluster was statistically significant (and comparable to the FFM domain assessments) for three of the cross-method comparisons. On the other hand, the FFM obtained significant convergent validity coefficients for all five domains between the self-report and informant assessments and four of the five domains for the agreement between the interview and therapist assessments. The most consistent convergence was obtained for the assessment of FFM extraversion (including even a significant relationship of the self-report with the therapist assessments), with a mean convergent validity coefficient of .39.

TABLE 3. Convergent and Discriminant Validity of the DSM-IV PD Clusters and Five Factor Model (FFM) Domains Across Methodology

	Self-Report With Interview n = **86**	**Self-Report With Therapist** n = **79**	**Self-Report With Informant** n = **67**	**Interview With Therapist** n = **77**	**Interview With Informant** n = **65**	**Therapist With Informant** n = **61**	**Mean Convergent**	**Mean Discriminant**
DSM-IV-TR Clusters								
Odd-Eccentric	**0.33**	–0.05	0.25	0.02	0.10	0.13	0.13	0.17
Dramatic-Emotional	**0.40**	0.05	**0.46**	**0.42**	0.27	0.24	0.31	0.17
Anxious-Fearful	**0.26**	0.24	0.01	**0.28**	0.02	–0.03	0.13	0.16
Mean	0.33	0.08	0.24	0.24	0.13	0.11	0.19	0.17
Median	0.33	0.05	0.25	0.28	0.10	0.13	0.13	0.17
FFM Domains								
Neuroticism	**0.64**	0.08	**0.40**	**0.28**	0.19	0.13	0.29	–0.09
Extraversion	**0.62**	**0.26**	**0.34**	**0.43**	**0.41**	**0.30**	0.39	0.01
Openness	**0.53**	**0.35**	0.13	**0.42**	0.14	0.23	0.30	0.03
Agreeableness	**0.54**	0.13	**0.55**	**0.35**	0.20	0.13	0.31	0.01
Conscientiousness	**0.55**	0.20	**0.34**	0.17	0.13	0.02	0.23	0.01
Mean	0.58	0.20	0.35	0.33	0.21	0.16	0.31	–0.01
Median	0.55	0.20	0.34	0.35	0.19	0.13	0.30	0.01

Notes. DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision (American Psychiatric Association, 2000); FFM = Five Factor Model. Self-report assessed by the Schedule for Nonadaptive and Adaptive Personality (Clark, 1993) for the DSM and NEO Personality Inventory—Revised (Costa & McCrae, 1992) for the FFM; Interview assessed by the Personality Disorder Interview-IV (Widiger et al., 1995) for the DSM and Structured Interview for the Five Factor Model (Trull & Widiger, 1997) for the FFM; Therapist and Informant assessed by the DSM Rating Form and the FFM Rating Form. The first six columns present all possible convergent correlations across each possible comparison. Mean Convergent = the mean convergent correlation across all possible combinations (i.e., the average across the first six columns). Mean Discriminant = The mean correlation of each DSM Cluster or FFM Domain with all same-model discriminant variables within and across methodologies. Those values marked in boldface type indicate correlations significant at $p < .05$.

Table 4 provides the convergent and averaged discriminant validity coefficients for the assessment of the 10 DSM-IV-TR personality disorders. It is evident from Table 4 that the higher convergent validity obtained for the dramatic-emotional PDs was due largely to the assessments of the antisocial, borderline, and histrionic PDs (the assessment of the histrionic PD even obtained a significant convergence between self-report and therapist). The weakest was obtained for the assessment of the schizoid and schizotypal PDs, for which the convergent validity coefficients averaged across all four methods did not exceed the averaged discriminant validity coefficients.

Table 5 provides the convergent and averaged discriminant validity coefficients for the assessment of the 30 facets of the FFM. Statistically significant convergent validity across self-report and interview assessments was obtained for 29 of the 30 FFM facets; the one exception occurred for the FFM assessment of openness to feelings. The mean convergent validity across all methods for the assessment of the 30 FFM facets was comparable to the convergent validity for the assessment of the 10 DSM-IV-TR personality disorders (see Table 4). Discriminant validity was consistently good across all 30 facets and lower than was obtained for the 10 DSM-IV-TR personality disorders.

As is evident from Tables 2–5, convergent validity was highest between the self-report and semi-structured interview assessments for both the DSM-IV-TR and FFM. Conversely, the assessments provided by the infor-

TABLE 4. Convergent and Discriminant Validity of DSM-IV-TR Personality Disorders Across Methodology

	Self-Report With Interview *n* = 86	Self-Report With Therapist *n* = 79	Self-Report With Informant *n* = 67	Interview With Theraist *n* = 77	Interview With Informant *n* = 65	Therapist With Informant *n* = 61	Mean Convergent	Mean Discriminant
Paranoid	**0.39**	–0.03	0.25	0.13	0.25	0.13	0.19	0.15
Schizoid	0.19	–0.09	0.13	–0.05	–0.10	0.20	0.05	0.06
Schizotypal	**0.28**	–0.11	0.09	0.22	–0.02	–0.15	0.05	0.09
Antisocial	**0.43**	0.05	**0.40**	**0.24**	**0.28**	0.12	0.25	0.11
Borderline	**0.37**	0.11	**0.48**	**0.47**	0.21	**0.33**	0.33	0.19
Histrionic	**0.50**	**0.35**	**0.33**	**0.41**	**0.29**	0.19	0.34	0.13
Narcissistic	**0.30**	0.14	**0.26**	0.23	–0.02	0.19	0.18	0.11
Avoidant	**0.46**	0.02	0.13	0.06	0.01	–0.10	0.10	0.09
Dependent	**0.41**	0.20	**0.28**	**0.29**	**0.29**	–0.10	0.23	0.10
Obsessive	**0.23**	**0.24**	–0.12	**0.26**	–0.16	**0.34**	0.13	0.07
Mean	0.35	0.09	0.22	0.23	0.10	0.12	0.18	0.11
Median	0.38	0.08	0.26	0.24	0.11	0.16	0.18	0.10

Notes. DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision (American Psychiatric Association, 2000); Self-Report assessed by the Schedule for Nonadaptive and Adaptive Personality (Clark, 1993); Interview assessed by the Personality Disorder Interview-IV (Widiger et al., 1995); Therapist and Informant by the DSM Rating Form. The first six columns present all possible convergent correlations for each personality disorder across each possible comparison. Mean Convergent = the mean convergent correlation across all possible combinations (i.e., the average across the first six columns). Mean Discriminant = The mean correlation of each personality disorder variable with all discriminant variables within and across methodologies. Those values marked in boldface type indicate correlations significant at $p < .05$.

mants and therapists obtained the weakest cross-method convergent validity. This could be attributable, in part, to the single item ratings that were obtained from the informants and therapists. The informant and therapist assessments of the DSM-IV-TR PDs and the FFM facets were single items, whereas the NEO PI-R, SNAP, SIFFM, and PDI-IV used multiple items for each construct.

Table 6 provides the convergent and averaged discriminant validity coefficients for the assessment of the FFM domains and facets when the informants used Form R of the NEO PI-R (rather than the FFMRF). It is evident from Table 6 that when the more psychometrically robust NEO PI-R was administered the convergent validity increased appreciably. For example, the mean agreement between the self-report and informant assessments of the FFM domains increased from .35 to .52, approaching that obtained between the semi-structured interview and self-report inventory assessments. Mean convergent validity for the interview and informant assessments increased from .21 to .33. Convergent validity for the informant and therapist assessments remained low and in fact decreased from .16 to .11.

It should also be noted that convergent validity for the FFMRF ratings with the NEO PI-R was good when the evaluation was confined to the same method of assessment. Convergent validity of the informant FFMRF with the informant NEO PI-R for the domains of neuroticism, extraversion, agreeableness, and conscientiousness were .62, .56, .57, and .63, respectively ($p < .001$ in each case). Convergent validity was also statistically significant for 21 of the 24 facets of these domains (ranging from .27 for anxiousness to .51 for warmth). However, the assessments of the domain of openness did not converge well (correlating only .15) and only one facet score correlated significantly (i.e., fantasy at .34).

DISCUSSION

The current study compared the convergent and discriminant validity of the DSM-IV-TR and the FFM constructs across (a) clinicians' descriptions of a patient, (b) self-report inventory, (c) semi-structured interview, and (d) informant report. This builds upon previous cross-methods assessment research (Clark, 2007; Garb, 2005; Oltmanns & Turkheimer, 2006; Segal & Coolidge, 2007; Widiger & Samuel, 2005b) and provides the first same-sample convergent and discriminant validity data across four assessment methods and for both the FFM and for the DSM-IV-TR.

Only two previous studies have directly compared the DSM-IV-TR and the FFM across methodologies, and these only compared self-report with informant report. Contrary to these two prior studies, the self to informant agreement in the current study were consistently more favorable for the FFM, at least at the domain level. The differences in findings are perhaps attributable to Ball et al. (2001) relying only on an abbreviated measure to assess the FFM and the Miller, Pilkonis, et al. (2005) utilizing DSM-IV-TR assessment aggregated across both a questionnaire and a semi-structured interview.

TABLE 5. Convergent and Discriminant Validity of the Five Factor Model Facets Across Methodology

	Self-Report with Interview $n = 86$	Self-Report with Therapist $n = 79$	Self-Report with Informant $n = 67$	Interview with Therapist $n = 77$	Interview with Informant $n = 65$	Therapist with Informant $n = 61$	Mean Convergent	Mean Discriminant
Anxiousness(n1)	**0.33**	**0.24**	0.16	**0.35**	0.22	0.08	0.23	–0.03
Angry Hostility(n2)	**0.47**	0.16	0.25	**0.37**	0.10	0.16	0.25	–0.06
Depressiveness(n3)	**0.60**	0.14	**0.37**	0.16	0.25	0.08	0.27	–0.06
Self-Consciousness(n4)	**0.34**	–0.04	0.20	0.12	0.18	0.06	0.14	–0.04
Impulsivity(n5)	**0.45**	**0.25**	0.19	**0.29**	0.23	0.21	0.27	–0.03
Vulnerability(n6)	**0.44**	–0.02	**0.29**	0.21	0.17	0.23	0.22	–0.05
Warmth(e1)	**0.36**	0.19	0.16	0.18	0.17	–0.18	0.14	0.05
Gregariousness(e2)	**0.49**	0.04	0.04	0.15	**0.27**	**0.29**	0.21	–0.01
Assertiveness(e3)	**0.38**	0.15	**0.38**	**0.36**	**0.38**	**0.47**	0.35	–0.03
Activity(e4)	**0.40**	0.17	0.28	0.18	**0.43**	0.21	0.28	0.01
Excitement Seeking(e5)	**0.41**	**0.32**	0.18	**0.33**	0.23	0.26	0.29	–0.02
Positive Emotions(e6)	**0.39**	0.08	0.21	0.12	0.20	–0.12	0.15	0.02
Fantasy(o1)	**0.37**	0.19	0.13	0.01	0.13	0.03	0.14	–0.01
Aesthetics(o2)	**0.34**	**0.28**	0.09	0.19	0.01	**0.32**	0.20	0.03
Feelings(o3)	0.13	0.03	0.10	0.16	–0.12	0.12	0.07	0.05
Actions(o4)	**0.37**	–0.04	–0.08	0.00	0.13	–0.05	0.06	–0.01
Ideas(o5)	**0.39**	0.20	0.21	**0.39**	0.25	0.10	0.26	0.01
Values(o6)	**0.27**	**0.35**	0.18	**0.45**	0.08	**0.29**	0.27	0.02
Trust(a1)	**0.43**	**0.26**	0.10	0.19	0.02	0.19	0.20	0.02
Straightforwardness(a2)	**0.39**	0.03	0.25	**0.29**	0.03	0.01	0.17	0.00
Altruism(a3)	**0.39**	0.01	**0.29**	0.10	0.16	0.04	0.17	0.02
Compliance(a4)	**0.38**	0.18	**0.44**	**0.31**	0.08	0.20	0.27	–0.01
Modesty(a5)	**0.44**	0.03	0.08	0.20	0.10	–0.06	0.13	–0.03
Tender-Mindedness(a6)	**0.27**	–0.04	**0.40**	0.21	0.14	–0.03	0.16	0.02
Competence(c1)	**0.36**	0.22	0.15	0.09	–0.07	0.01	0.13	0.03
Order(c2)	**0.58**	0.08	**0.29**	0.17	0.18	–0.10	0.20	0.02
Dutifulness(c3)	**0.31**	0.15	0.24	0.08	0.13	–0.15	0.13	–0.01
Achievement(c4)	**0.45**	**0.27**	0.24	**0.34**	0.09	0.05	0.24	0.00
Self-Discipline(c5)	**0.40**	0.04	0.18	0.15	0.00	–0.11	0.11	0.00
Deliberation(c6)	**0.37**	0.17	0.05	0.15	0.03	0.11	0.15	–0.01
Mean	0.39	0.14	0.20	0.21	0.14	0.09	0.19	0.00
Median	0.39	0.16	0.19	0.19	0.14	0.08	0.20	–0.01

Notes. Self-report assessed by the NEO Personality Inventory—Revised (Costa & McCrae, 1992); Interview assessed by the Structured Interview for the Five Factor Model (Trull & Widiger, 1997); Therapist and Informant collected using the Five Factor Model Rating Form. The first six columns present all possible convergent correlations for each Five Factor Model facets across each possible comparison. Mean Convergent = the mean convergent correlation across all possible combinations (i.e., the average across the first six columns). Mean Discriminant = The mean correlation of each Five Factor Model facet with all discriminant variables outside its domain, both within and across methodologies. Those values marked in boldface type indicate correlations significant at $p < .05$.

TABLE 6. Convergent and Discriminant Validity of the Five Factor Model Across Methodology Using Informant NEO PI-R

	Self-Report with Interview n = **86**	**Self-Report with Therapist** n = **79**	**Self-Report with Informant** n = **67**	**Interview with Therapist** n = **77**	**Interview with Informant** n = **65**	**Therapist with Informant** n = **61**	**Mean Convergent**	**Mean Discriminant**
FFM Domains								
Neuroticism	**0.64**	0.08	**0.46**	**0.28**	**0.31**	0.09	0.31	–0.08
Extraversion	**0.62**	**0.26**	**0.54**	**0.43**	**0.48**	**0.29**	0.44	0.03
Openness	**0.53**	**0.35**	**0.67**	**0.42**	**0.40**	0.23	0.43	0.03
Agreeableness	**0.54**	0.13	**0.61**	**0.35**	**0.30**	0.11	0.34	0.02
Conscientiousness	**0.55**	0.20	**0.33**	0.17	0.15	–0.14	0.21	–0.01
Mean	0.58	0.20	0.52	0.33	0.33	0.11	0.35	0.00
Median	0.55	0.20	0.54	0.35	0.31	0.11	0.34	0.02
FFM Facets								
Anxiousness(n1)	**0.33**	**0.24**	0.17	**0.35**	0.23	0.16	0.24	–0.01
Angry Hostility(n2)	**0.47**	0.16	**0.44**	**0.37**	**0.34**	**0.43**	0.37	–0.07
Depressiveness(n3)	**0.60**	0.14	**0.45**	0.16	0.25	–0.02	0.26	–0.03
Impulsivity(n5)	**0.45**	**0.25**	**0.32**	**0.29**	0.44	0.17	0.29	0.01
Self-Consciousness(n4)	**0.34**	**0.25**	**0.32**	**0.29**	0.24	0.17	0.29	0.01
Vulnerability(n6)	**0.44**	–0.02	**0.40**	0.21	0.24	–0.02	0.21	–0.06
Warmth(e1)	**0.36**	0.19	**0.49**	0.18	**0.40**	0.07	0.28	0.06
Gregariousness(e2)	**0.49**	0.04	**0.31**	0.15	**0.48**	0.19	0.28	0.00
Assertiveness(e3)	**0.38**	0.15	**0.29**	**0.36**	0.21	**0.44**	0.31	0.00
Activity(e4)	**0.40**	0.17	**0.54**	0.18	0.18	**0.33**	0.30	0.01
Excitement Seeking(e5)	**0.41**	**0.32**	**0.49**	**0.33**	**0.36**	**0.30**	0.37	–0.01
Positive Emotions(e6)	**0.39**	0.08	**0.37**	0.12	**0.40**	0.03	0.23	0.04
Fantasy(o1)	**0.37**	0.19	**0.50**	0.01	0.23	0.23	0.25	0.02
Aesthetics(o2)	**0.34**	**0.28**	**0.53**	0.19	0.21	0.10	0.28	0.04

Feelings(o3)	0.13	0.03	**0.32**	0.16	0.11	0.04	0.13	0.04
Actions(o4)	**0.37**	–0.04	0.22	0.00	**0.29**	0.01	0.14	–0.01
Ideas(o5)	**0.39**	0.20	**0.42**	**0.39**	0.24	0.22	0.31	0.02
Values(o6)	**0.27**	**0.35**	0.23	**0.45**	**0.33**	0.24	0.31	0.03
Trust(a1)	**0.43**	**0.26**	**0.48**	0.19	**0.33**	**0.43**	0.35	0.04
Straightforwardness(a2)	**0.39**	0.03	**0.41**	**0.29**	0.24	0.00	0.23	0.00
Altruism(a3)	**0.39**	0.01	**0.48**	0.10	0.11	–0.13	0.16	0.04
Compliance(a4)	**0.38**	0.18	**0.57**	**0.31**	**0.31**	0.16	0.32	–0.02
Modesty(a5)	**0.44**	0.03	**0.35**	0.20	**0.31**	–0.02	0.22	–0.02
Tender-Mindedness(a6)	**0.27**	–0.04	**0.32**	0.21	0.21	0.05	0.17	0.03
Competence(c1)	**0.36**	0.22	**0.35**	0.09	0.12	–0.05	0.18	0.02
Order(c2)	**0.58**	0.08	**0.31**	0.17	0.04	–0.21	0.16	0.00
Dutifulness(c3)	**0.31**	0.15	**0.37**	0.08	0.12	–0.16	0.15	–0.02
Achievement(c4)	**0.45**	**0.27**	0.26	**0.34**	0.23	0.02	0.26	0.00
Self-Discipline(c5)	**0.40**	0.04	**0.34**	0.15	0.21	0.09	0.20	0.01
Deliberation(c6)	**0.37**	0.17	**0.39**	0.15	0.19	0.21	0.25	–0.01
Mean	0.39	0.14	0.38	0.21	0.24	0.12	0.25	0.00
Median	0.39	0.16	0.37	0.19	0.24	0.10	0.25	0.00

Notes. Self-report assessed by the NEO Personality Inventory—Revised (Costa & McCrae, 1992); Interview assessed by the Structured Interview for the Five Factor Model (Trull & Widiger, 1997); Therapist collected using the Five Factor Model Rating Form; Informant assessed by the NEO Personality Inventory—Revised (Costa & McCrae, 1992). The first six columns present all possible convergent correlations for each Five Factor Model variable across each possible comparison. Mean Convergent = the mean convergent correlation across all possible combinations (i.e., the average across the first six columns). Mean Discriminant = The mean correlation of each Five Factor Model domain or facet with all discriminant variables within and across methodologies. Those values marked in boldface type indicate correlations significant at $p < .05$.

One might question, though, whether the FFM cross-method convergent validity coefficients were inordinately high in the current study or, conversely, whether the DSM-IV-TR results were inordinately low. However, neither appears to be the case when the findings are compared to prior studies confined to either the FFM or the DSM-IV-TR. For example, Widiger and Boyd (2009) summarized the convergent validity from 25 studies and found the median correlation between dimensional assessments of the DSM-IV-TR PDs across self-report and semi-structured interviews was .38. The median value for this relationship within the current study was, in fact, also .38.

A few studies have reported on the agreement between self-report and clinician ratings of the DSM PDs (Bronisch, Flett, Garcia-Borreguero, & Wolf, 1993; Chick, Sheaffer, Goggin, & Sison, 1993; Hyler et al., 1989; Rossi, Van Den Brande, Tobac, Sloore, & Hauben, 2003). Median convergence was only .08 in Hyler et al. (1989) and .12 in Bronisch et al. (1993) using the kappa statistic. Convergence was not much higher in Chick et al. (1993) and Rossi et al. (2003), which reported a median correlation of only .05 and .20, respectively, for dimensional ratings. The current study was consistent with these results, with a median correlation of .08 between the self-report and clinician ratings of the DSM-IV-TR PDs.

The current study, however, did obtain somewhat higher convergence between the self and informant NEO PI-R assessments than is typical. McCrae and colleagues (2004) reported that the median convergence for the domains between self and informant NEO PI-R reports of the FFM, across 29 samples, was .43. The current study obtained a median correlation of .54 between the self and informant versions of the NEO PI-R. Klonsky and colleagues (2002) summarized 17 studies that provided the correlation between self-report and informant reports methods for the DSM PDs and found a median value of .36. This value is somewhat higher than the .26 obtained within the current study.

The FFM domains higher convergent validity than the DSM-IV-TR PDs could reflect that the FFM constructs are more straightforward, understandable, and coherent (i.e., unidimensional) than the clinically complex and heterogeneous DSM-IV-TR diagnostic constructs (Shedler & Westen, 2004; Widiger & Trull, 2007). The convergent validity of the particular PDs, averaged across all methods, was typically quite low and, in three instances, was no higher than the discriminant coefficients (schizoid, schizotypal, and avoidant). In contrast, the averaged convergent coefficient was higher than the discriminant coefficient for all of the 30 FFM facets. Additionally, several of the DSM-IV-TR scales (e.g., schizoid) obtained rather low internal consistency values, which limit their convergent correlations. The heterogeneity within these PDs highlights an inherent assessment weakness of the DSM-IV-TR system relative to the more homogenous constructs within the FFM (Clark, 2007; Widiger & Trull, 2007).

The current study went beyond prior FFM and DSM-IV-TR cross-method assessment comparisons to consider discriminant validity, as well as con-

vergent. It is evident from these comparisons that the FFM obtained considerably better discriminant validity than the DSM-IV-TR. For example, the median discriminant validity for the five FFM domains was .01, versus .17 for the DSM-IV-TR clusters. These findings are in harmony with a recent systematic review of the discriminant validity of the three DSM-IV-TR clusters provided by Sheets and Craighead (2007). They suggested that the three cluster organization of the personality disorders in DSM-IV-TR is inconsistent with the covariation among the disorders and noted that some of the PDs covary across the clusters as much as they covary within a cluster. Given this finding and the lack of a strong theoretical or empirical basis for their original construction (Millon, 1981), one might question whether it is fair to consider the convergent and discriminant validity of the DSM-IV-TR clusters. Discriminant validity did, in fact, improve for the 10 PDs relative to the three clusters, dropping from .17 to .11. However, quite a few studies continue to be conducted with respect to the three DSM-IV-TR clusters (e.g., Aluja et al., 2007; Ehrensaft et al., 2006; Lenzenweger & Willet, 2007), researchers have argued for their validity (e.g., Fossati et al., 2000; Rodebaugh, Chambless, Renneberg, & Fydrich, 2005). In any case, the improved discriminant validity at the level of the individual PDs still did not compare to the median discriminant validity of the FFM at either the domain (.01) or the facets (.00) levels.

The stronger discriminant validity obtained for the assessment of the FFM domains and facets is consistent with the fact that this model was derived, in large part, through factor analyses of personality traits and trait terms that place considerably more emphasis on identifying empirically (and conceptually) distinct constructs (McCrae & Costa, 2003). In contrast, inadequate discriminant validity has been a longstanding problem for the psychiatric nomenclature (Bornstein, 1998; Clark, 2007; Trull & Durrett, 2005; Widiger & Trull, 2007) despite the fact that a considerable amount of time and effort are spent on differential diagnosis in the development and application of the diagnostic manual (Frances et al., 1995). In fact, much of the diagnostic co-occurrence among the DSM-IV-TR personality disorders can be explained by different diagnoses sharing the same FFM facets (Lynam & Widiger, 2001).

The results of the current study also contribute to the more general literature on cross-method assessments of psychological variables (Meyer et al., 2001). An additional notable finding in the current study was the relatively weak cross-method convergent validity for the therapist ratings. This finding is compatible with prior research on the correlation of therapist ratings with self-report instruments (e.g., Bronisch et al., 1993; Chick et al., 1993; Hyler et al., 1989; Rossi et al., 2003). These previous results have generally been interpreted as indicating that clinicians perceive and describe their patients' personality traits and disorders differently than the patients describe themselves (Klein, 2003; Klonsky et al., 2002; Ready & Clark, 2002). This contention is supported by a literature suggesting that clinicians' judgments and diagnoses are less valid than are those based on

more systematic assessments (Garb, 2005; Grove, Zald, Lebow, Snitz, & Nelson, 2000).

However, Westen and Weinberger (2004) echoed Meehl (1954) in separating the source of the ratings (clinical judgment) from the method of assessment. This is particularly relevant as the therapist ratings for both the DSM-IV-TR and the FFM were collected in the current study using one-page rating forms, consistent with all prior therapist cross-method assessment studies of the DSM PDs (Bronisch et al., 1993; Chick et al., 1993; Hyler et al., 1989; Rossi et al., 2003). Only a single study had ever examined the agreement between self-report and clinician ratings of the FFM and in that study the FFM scores were generated from a lengthier instrument completed by the clinician (Piedmont & Ciarrocchi, 1999). Perhaps not surprisingly, that study reported a mean convergent value of .32 for the FFM domains, which was higher than the .20 obtained in the current study. The single-item assessment of FFM facets do have empirical support when the comparison is within-method, as indicated in Mullins-Sweatt, Jamerson, Samuel, Olson, and Widiger (2006) and by the convergence of the informant FFMRF ratings and the Form R NEO PI-R assessment in the current study. However, it may be unrealistic to expect single item assessments to produce strong cross-method validity. In any case, before concluding that therapists have markedly different perceptions of their patients' personalities or personality disorders than the patients themselves, it would be important to obtain the therapists' assessments with comparably thorough methods of assessment. Just as the cross-method convergence for the informants improved when the informant NEO PI-R was used, the correlation between clinicians' reports and other methods might also improve if they complete a full SNAP or NEO PI-R on their patients.

We did not ask therapists to complete a SNAP and/or NEO PI-R in the current study as the methodology was already quite labor intensive. In fact, no published study has ever asked therapists to complete an informant version of a PD measure. Clinicians have provided extensive PD descriptions of their patients using the Shedler-Westen Assessment Procedure and research indicates these can be quite valid (Westen & Shedler, 2007). Nevertheless, there has been very limited research on the cross-method validity of SWAP-200 assessments (Clark, 2007; Widiger & Boyd, 2009; Wood, Garb, Nezworski, & Koren, 2007). In sum, given the increased convergent validity of the informant description when they completed Form R of the NEO PI-R, it would be informative for future research to obtain comparable assessments by clinicians. It does seem likely that therapist ratings using more extensive instruments would evidence greater convergence with other assessment methods.

Such a finding, though, would not necessarily suggest that the previous literature regarding problematic clinical descriptions is in error (Garb, 2005). The FFMRF and DSMRF ratings obtained in the current study are

more externally valid and mirror the assessments that clinicians routinely make on the basis of unstructured clinical interviews. Clinicians do not typically complete a SNAP (for instance) when reaching their conclusions regarding a patient's DSM-IV-TR PD diagnoses. Nevertheless, future research demonstrating that convergent validity improves significantly if clinicians complete a SNAP would reaffirm that clinicians can provide valid descriptions of their patients if they take the time to complete a more thorough and systematic assessment.

LIMITATIONS

A limitation of the current study was that the sample did not obtain an even distribution of PD symptomatology due in part, perhaps, to being confined to females. The participants were characterized primarily by antisocial and borderline PD symptomatology, which is consistent with the sampling of a female substance-abusing population, but is not typical of a general clinical sample of female patients. Nonetheless, there was a more restrictive range of schizoid symptomatology, and perhaps schizotypal, narcissistic, and obsessive-compulsive symptomatology (see Table 1). Although there is no reason to expect significant differences in convergent or discriminant validity of the FFM relative to the DSM-IV-TR across gender (Costa & McCrae, 1992; Morey, Warner, & Boggs, 2002) the relatively weaker convergent validity obtained for the schizoid and schizotypal PDs could be due in part to this restriction in range. In sum, it would be useful for future studies to sample a broader and more representative range of PD symptomatology.

The current study compared the DSM-IV-TR exclusively to the FFM. There are, of course, several other dimensional models that have been proposed as alternatives to the current model (Clark, 2007; Widiger & Simonsen, 2005). While it is likely that other dimensional models would also show improved discriminant validity, relative to the current DSM-IV-TR model, this is an empirical question that should be tested. Future research that extends the current study to compare the DSM-IV-TR to one or more of these alternative models would be quite informative. Additionally, research that evaluates the ability of these alternative models, as well as the assessment methods, to predict important behavioral and clinical outcomes would clarify the advantages and disadvantages of shifting to a dimensional system.

Finally, the patients within the current study were self-referred from selected treatment clinics. While this allowed for increased external validity it also resulted in an uneven distribution in the number of patients described by the participating clinicians. Thus, the overall validity of the clinicians' ratings would be impacted by the degree to which an individual clinician's ratings were idiosyncratic. Future research that collected only a single rating from a clinician would negate such a potential confound.

CONCLUSIONS

An essential test of the validity of any clinical nomenclature is the agreement of descriptions across various assessment methodologies (Blacker & Endicott, 2000). In the only two prior studies to directly compare the convergent validity of the FFM and DSM-IV-TR, the results indicated either no appreciable difference (Ball et al., 2001) or favored the DSM-IV-TR (Miller, Pilkonis, et al., 2005). The current study examined convergent validity across four methods of assessment (i.e., self-report, informant-report, clinician rating, and semi-structured interview) and also considered discriminant validity. The results indicate that the FFM has an appreciable advantage over the DSM-IV-TR in terms of discriminant validity and, at the domain level, convergent validity. Finally, the convergent validity for therapist descriptions was notably low for both models due, perhaps, to the absence of systematic assessments of DSM-IV-TR or FFM constructs.

REFERENCES

Aluja, A., Cuevas, L., Garcia, L. F., & Garcia, O. (2007). Zuckerman's personality model predicts MCMI-III personality disorders. *Personality and Individual Differences, 42*, 1311–1321.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.) Washington, DC: Author.

Axelrod, S. R., Widiger, T. A., Trull, T. J., & Corbitt, E. M. (1997). Relation of five-factor model antagonism facets with personality disorder symptomatology. *Journal of Personality Assessment, 69*, 297–313.

Ball, S. A., Rounsaville, B. J., Tennen, H., & Kranzler, H. R. (2001). Reliability of personality disorder symptoms and personality traits in substance-dependent inpatients. *Journal of Abnormal Psychology, 110*, 341–352.

Blacker, D., & Endicott, J. (2000). Psychometric properties: Concepts of reliability and validity. In A. J. Rush et al. (eds.), *Handbook of psychiatric measures* (pp. 7–14) Washington, DC: American Psychiatric Association.

Blais, M. A. (1997). Clinical ratings of the five-factor model of personality and the DSM-IV personality disorders. *Journal of Nervous and Mental Disease, 185,* 388–393.

Bornstein, R. F. (1998). Reconceptualizing personality disorder diagnosis in the DSM-V: The discriminant validity challenge. *Clinical Psychology: Science and Practice, 5*, 333–343.

Bronisch, T., Flett, S., Garcia-Borreguero, D., & Wolf, R. (1993). Comparison of self-rating questionnaire with a diagnostic checklist for the assessment of DSM-III-R personality disorders. *Psychopathology, 26*, 102–107.

Chick, D., Sheaffer, C. I., Goggin, W. C., & Sison, G. F. (1993). The relationship between MCMI personality scales and clinician-generated DSM-III-R personality disorder diagnoses. *Journal of Personality Assessment, 61*, 264–276.

Clark, L. A. (1993). *Manual for the schedule for nonadaptive and adaptive personality*. Minneapolis, MN: University of Minnesota Press.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58*, 227–257.

Clark, L. A., & Harrison, J. A. (2001). Assessment instruments. In W. J. Livesley (Ed.), *Handbook of personality disorders. Theory, research and treatment* (pp. 277–306). New York: Guilford.

Costa, Jr., P. T., & McCrae, R. R. (1992). *Professional manual: Revised NEO personality inventory* (NEO-PI-R) *and NEO*

five-factor inventory (NEO-FFI).Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64,* 21–50.

Ehrensaft, M. K., Cohen, P., & Johnson, J. G. (2006). Development of personality disorder symptoms and the risk for partner violence. *Journal of Abnormal Personality, 115*, 474–483.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J.B.W., & Benjamin, L. S. (1997). *Structured clinical interview for DSM-IV Axis-II personality disorders*. Washington, DC: American Psychiatric Association.

Fossati, A., Maffei, C., Bagnato, M., Battaglia, M., Donati, D., Donini, M., et al. (2000). Patterns of covariation of DSM-IV personality disorders in a mixed psychiatric sample. *Comprehensive Psychiatry, 41,* 206–215.

Frances, A. J., First, M. B., & Pincus, H. A. (1995). *DSM–IV guidebook.* Washington, DC: American Psychiatric Press.

Garb, H. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89.

Gough, H. G., & Heilbrun, A. B. (1983). *The adjective check list manual.* Palo Alto, CA: Consulting Psychologists Press.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Hyler, S. E., Reider, R. O., Williams, J.B.W., Spitzer, R. L., Lyons, M., & Hendler, J. (1989). A comparison of clinical and self-report diagnoses of DSM-III personality disorders in 552 patients. *Comprehensive Psychiatry, 30*, 170–178.

John, O. P. (1990). The "big five" taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. Pervin (Ed.), *Handbook of personality theory and research* (pp. 66–100). New York: Guilford Press.

Klein, D. N. (2003). Patients' versus informants' reports of personality disorders in predicting 7.5 year outcome in outpatients with depressive disorders. *Psychological Assessment, 15*, 216–222.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9*, 399–311.

Kupfer, D. J., First, M. B., & Regier, D. E. (2002). Introduction. In D. J. Kupfer, M. B. First, & D. E. Regier (Eds.), *A research agenda for DSM-V* (pp. xv–xxiii). Washington, DC: American Psychiatric Association.

Lenzenweger, M. F., & Willett, J. B. (2007). Predicting individual change in personality disorder features by simultaneous individual change in personality dimensions linked to neurobehavioral systems: The longitudinal study of personality disorders. *Journal of Abnormal Psychology, 116*, 684–700.

Livesley, W. J. (2001). Conceptual and taxonomic issues. In W. J. Livesley (Ed.), *Handbook of personality disorders: Theory, research, and treatment* (pp. 3–38). New York: Guilford.

Livesley, W. J. (2003). Diagnostic dilemmas in classifying personality disorder. In K. A. Phillips, M. B. First, & H. A. Pincus (Eds.), *Advancing DSM. Dilemmas in psychiatric diagnosis* (pp. 153–190). Washington, DC: American Psychiatric Association.

Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology, 110*, 401–412.

McCrae, R. R., & Costa P. T., Jr. (2003). *Personality in adulthood. A five-factor theory perspective* (2nd ed.). New York: Guilford.

McCrae, R. R., Costa, P. T., Jr., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., Hrebickova, M., & Urbanek, T. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38*, 179–201.

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota Press.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A

review of evidence and issues. *American Psychologist, 56*, 128–165.

Miller, J. D., Pilkonis, P. A., & Clifton, A. (2005). Self- and other-reports of traits from the five-factor model: Relations to personality disorders. *Journal of Personality Disorders, 19*, 400–419.

Miller, J. D., Bagby, R. M., Pilkonis, P. A., Reynolds, S. K., & Lynam, D. R. (2005). A simplified scoring technique for the DSM-IV personality disorders with the five-factor model. *Assessment, 12*, 404–415.

Millon, T. (1981). *Disorders of personality: DSM-Ill, Axis 11*. New York: Wiley.

Morey, L., Warner, M., & Boggs, C. (2002). Gender bias in the personality disorders criteria: An investigation of five bias indicators. *Journal of Psychopathology and Behavioral Assessment, 24*, 55–65.

Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated measure of the five-factor model. *Assessment, 13*, 119–137.

O'Connor, B. P. (2005). A search for consensus on the dimensional structure of personality disorders. *Journal of Clinical Psychology, 61*, 323–645.

Oltmanns, T. F., & Turkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In R. F. Krueger and J. L. Tackett (Eds.), *Personality and psychopathology* (pp. 71–111). New York: Guilford.

Ozer, D. J., & Reise, S. P. (1994). Personality assessment. *Annual Review of Psychology, 45*, 357–388.

Piedmont, R. L., & Ciarrocchi, J. W. (1999). The utility of the revised NEO personality inventory in an outpatient, drug rehabilitation context. *Psychology of Addictive Behaviors, 13*, 213–226.

Pilkonis, P. A., Heape, C. L., Proietti, J. M., Clark, S. W., McDavid, J. D., & Pitts, T. E. (1995). The reliability and validity of two structured diagnostic interviews for personality disorders. *Archives of General Psychiatry, 52*, 46–54.

Ready, R. E., & Clark, L. A. (2002). Correspondence of psychiatric patient and informant ratings of personality traits, temperament, and interpersonal problems. *Psychological Assessment, 14*, 39–49.

Reynolds, S. K., & Clark, L. A. (2001). Predicting dimensions of personality disorder from domains and facets of the five-factor model. *Journal of Personality, 69*, 199–222.

Rodebaugh, T. L., Chambless, D. L., Renneberg, B., & Fydrich, T. (2005). The factor structure of the DSM-III-R personality disorders: An evaluation of competing models. *International Journal of Methods of Psychiatric Research, 14*, 43–55.

Rogers, R. (2001). *Diagnostic and structured interviewing. A handbook for psychologists*. New York: Guilford.

Rossi, G., Van den Brande, I., Tobac, A., Sloore, H., & Hauben, C. (2003). Convergent validity of the MCMI-III personality disorder scales and the MMPI-2 scales. *Journal of Personality Disorders, 17*, 330–340.

Rounsaville, B. J., Alarcon, R. D., Andrews, G., Jackson, J. S., Kendell, R. E., & Kendler, K. (2002). Basic nomenclature issues for DSM-V. In D. J. Kupfer, M. B. First, & D. E. Regier (Eds.), *A research agenda for DSM-V* (pp. 1–29). Washington, DC: American Psychiatric Association.

Samuel, D. B., & Widiger, T. A. (2006). Clinicians' judgments of clinical utility: A comparison of the DSM-IV and five-factor models. *Journal of Abnormal Psychology, 115*, 298–308.

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review, 28*, 1326–1342.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review, 23*, 1055–1085.

Segal, D. L., & Coolidge, F. L. (2007). Structured and semistructured interviews for differential diagnosis: Issues and applications. In M. Hersen, S. M. Turner, & D. C. Beidel (Eds.), *Adult psychopathology and diagnosis* (5th ed., pp. 78–100). New York: Wiley.

Shedler, J., & Westen, D. (2004). Dimensions of personality pathology: An alternative to the five-factor model. *American Journal of Psychiatry, 161*, 1743–1754.

Sheets, E., & Craighead, W. E. (2007). Toward an empirically based classification of personality pathology. *Clinical*

Psychology: Science and Practice, 14, 77–93.

Sprock, J. (2002). A comparative study of the dimensions and facets of the five-factor model in the diagnosis of cases of personality disorder. *Journal of Personality Disorders, 16*, 402–423.

Trull, T. J., & Durrett, C. A. (2005). Categorical and dimensional models of personality disorder. *Annual Review of Clinical Psychology, Vol. 1*, 355–380.

Trull, T. J., & Widiger, T. A. (1997). *Structured Interview for the Five-Factor Model of Personality.* Odessa, FL: Psychological Assessment Resources.

Trull, T. J., Widiger, T. A., & Burr, R. (2001). A structured interview for the assessment of the five-factor model of personality: Facet-level relations to Axis II personality disorders. *Journal of Personality, 69*, 1175–1198.

Trull, T. J., Widiger, T. A., Useda, J. D., Holcomb, J., Doan, B.-T., Axelrod, S. R., Stern, B. L., & Gershuny, B. S. (1998). A structured interview for the assessment of the five-factor model of personality. *Psychological Assessment, 10*, 229–240.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54–60.

Westen, D., & Shedler, J. (2007). Personality diagnosis with the Shedler-Westen assessment procedure (SWAP): Integrating clinical and statistical measurement and prediction. *Journal of Abnormal Psychology, 116*, 810–822.

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*, 595–613.

Widiger, T. A., & Boyd, S. (2009). Assessing personality disorders. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (3rd ed., pp. 336–363). New York: Oxford University Press.

Widiger, T. A., & Costa, P. T., Jr. (2002). Five-factor model personality disorder research. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (2nd ed., pp. 59–87). Washington, DC: American Association.

Widiger, T. A., Mangine, S., Corbitt, E. M., Ellis, C. G., & Thomas, G. V. (1995). *Personality disorder interview-IV. A semi-structured interview for the assessment of personality disorders. Professional manual.* Odessa, FL: Psychological Assessment Resources.

Widiger, T. A., & Samuel, D. B. (2005a). Diagnostic categories or dimensions? A question for the Diagnostic and statistical manual of mental disorders—fifth edition. *Journal of Abnormal Psychology, 114*, 494–504.

Widiger, T. A., & Samuel, D. B. (2005b). Evidence based assessment of personality disorders. *Psychological Assessment, 17*, 278–287.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*, 110–130.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*, 71–83.

Wood, J. M., Garb, H. N., Nezworski, M. T., & Koren, D. (2007). The Shedler-Westen assessment procedure as a basis for modifying DSM personality disorder diagnostic categories. *Journal of Abnormal Psychology, 116*, 823–836.

Zimmerman, M. (2003). What should the standard of care for psychiatric diagnostic evaluations be? *Journal of Nervous and Mental Disease, 191*, 281–286.