

The factor structure of the DSM-III-R personality disorders: an evaluation of competing models

THOMAS L. RODEBAUGH,¹ DIANNE L. CHAMBLESS,² BABETTE RENNEBERG,³ THOMAS FYDRICH³

1. Department of Psychology, Temple University, USA

2. Department of Psychology, University of Pennsylvania, USA

3. Psychological Institute, University of Heidelberg, Germany

Abstract

Confirmatory factor analysis was used to test a series of competing models to determine the best way to characterize the factor structure of the DSM-III-R personality disorder scores. Data were collected from 301 clients with a primary diagnosis of anxiety disorder. The Structured Clinical Interview for DSM-III-R (SCID) was used to determine Axis I and Axis II diagnoses, and adequate to good reliability was obtained for the number of criteria met for a given personality disorder. Several factor models were tested, including a single-factor model, a three-factor model based on the DSM grouping of odd, dramatic, and anxious personality disorders, and a four-factor model based on normal personality theories. Only the DSM three-factor model received strong and unequivocal support.

Key words: personality disorders, factor analysis, outpatients

Personality disorders have a long and controversial history as diagnostic entities (see, for example, Tyrer, Casey, and Ferguson, 1991; Arntz, 1999). One aspect of this controversy involves the appropriate grouping of the personality disorders. Beginning with the DSM-III (APA, 1980), in which criteria for the personality disorders were first provided, the DSM system has grouped the personality disorders in three clusters (APA, 1987; APA, 1994). Often called the odd, dramatic, and anxious clusters, the clusters are presented as potentially being useful in educational and research settings; however, as noted in the DSM-IV (APA, 1994), the clusters themselves have not been uniformly supported by research findings. (We use the term *clusters* throughout, in order to be consistent with the DSM. However it is worth noting that our use of the term 'cluster' in this case is not related to the statistical technique of cluster analysis, which is not used in this paper.)

Most investigations into the validity of the clusters have conceptualized the clusters as three factors underlying the personality disorders. The following is a

brief review of attempts to determine the factor structure of the DSM personality disorders.

Attempts to determine the factor structure of the DSM-III personality disorder diagnoses or personality disorder criteria were, for the most part, marred by inconsistent solutions arrived at by questionable statistical means. Two studies used confirmatory factor analysis methods (at least in part), and neither of these studies supported the DSM-III three-factor approach (Bell and Jackson, 1992; Nestadt et al., 1994). Widiger et al. (1991) provided a review regarding six exploratory factor analyses, although these authors noted that there was little consistency across studies. The authors attempted to arrive at a final conclusion regarding these studies using a principal component analysis of the averaged correlation matrices of nine studies. Their results are subject to the same concerns about principal components analyses that are noted below. Three studies, one subsequent to Widiger et al.'s review, recovered a factor structure comparable to the DSM-III three-factor solution (Hyer and Lyons, 1988; Kass, Skodol, Charles, Spitzer and Williams,

1985; Hyler et al. 1992). However, all three studies used exploratory factor analysis methods that were unsuited to their data: Each used either (a) principal components analysis with communalities estimated to be 1, which assumes perfect and consistent reliability of their interview data, or (b) orthogonal rotation, which assumes factors to be uncorrelated – or both (a) and (b). The assumptions underlying both (a) and (b) clearly appear unwarranted.

With reference to point (a), above, it is worth recalling that, in factor analysis, each variable has a *communality*, which is the proportion of variance that the factors account for in the variable (Gorsuch, 1983). A communality of 1, which is assumed in principal components analysis, indicates that all of the variance associated with that variable is explained by the factors (that the measure is perfectly reliable and perfectly predicted by the factors). The studies using principal components analysis used clinician ratings with no known reliability. Although acceptable reliability has been achieved for diagnosis of personality disorders using structured clinical interviews, these findings do not generalize to clinician ratings, and the reliability estimates of structured interviews are themselves far from perfect (see, for example, First et al., 1995), indicating that the actual communalities should be lower than one. With a relatively small number of variables, and communalities that may be low, a principal components analysis is more likely to be misleading than are other factor methods (see Gorsuch, 1983: 121–5). In regard to point (b), inter-correlations among personality disorders in different clusters are often in the small to moderate range (for example as evident in the correlation matrix presented by Widiger et al., 1991, which was averaged across several studies), clearly indicating that, even if the three-factor solution is correct, those factors are likely to be correlated. Moreover, two of the studies found a four-factor solution in which obsessive-compulsive personality disorder was the sole variable displaying a large loading on the fourth factor (Kass et al., 1985; Hyler and Lyons, 1988). Such a result is a typical artefact of using principal components analysis with communalities estimated to be 1 (Lee and Comrey, 1979), as was the case in Hyler and Lyons's study (1988).

Subsequent to the introduction of the DSM-III-R, interest increased in reconciling personality disorders with a theory of normal personality. Overall, a variety

of studies suggested that four of the five factors in the Five Factor Model (FFM) were related to personality disorders, suggesting that a four-factor solution was more appropriate than the three-factor solution suggested by the DSM (see Austin and Deary, 2000 for a review). In summary, the personality disorders have been found to relate to the normal personality factors of conscientiousness, extraversion, agreeableness, and neuroticism. The fifth factor, Openness to Experience, has shown no consistent or strong relationship with any personality disorder (Austin and Deary). Despite general findings indicating a relationship between the FFM and personality disorders, no one solution was consistently obtained in every sample. Exploratory factor analysis of questionnaires based on the DSM-III-R criteria or of numbers of criteria met for each personality disorder sometimes found at least some support for a three-factor solution (Dowson and Berrios, 1991; Deary, Peter, Austin and Gibbon, 1998), but also sometimes found some or more support for a solution based on the FFM or a similar model based on normal personality (for example, Eysenck's three-factor model; Deary et al, 1998; Austin and Deary, 2000). Consistent with this general conclusion, O'Connor and Dyce (1998), using specific forms of exploratory factor analysis and data from nine previous studies (which used a variety of measures and criteria, some of which were based on DSM-III and some on the DSM-III-R), found support for both the DSM three-factor solution and the FFM four-factor solution.

Finally, there has been at least one attempt to model the factor structure of the DSM-IV personality disorders. Stanislow et al. (2002) used competing models to evaluate one-factor, three-factor, and four-factor solutions for the criteria for the avoidant, obsessive-compulsive, schizotypal, and borderline personality disorder. A structured clinical interview with acceptable reliability was used for this study. The authors found the four-factor solution to work the best, suggesting that the four disorders have more to distinguish themselves from each other than aspects in common. This result supports neither the DSM three-factor solution, nor a normal personality theory solution. However, despite this study's strengths, it did not evaluate the full range of personality disorders, which, by definition, means that it did not fully evaluate the DSM three-factor solution. In addition, the variables analysed appear to have used only a three-point scale. No information is given regarding whether

these variables met normality criteria necessary for the type of analysis used, but it appears extremely doubtful that they could have done so. It is therefore extremely difficult to interpret what bearing this study may have. Overall, then, the available literature is equivocal regarding the appropriate grouping of the personality disorders, with factor structures based on the DSM system, the FFM, and other normal personality theories all receiving some support in some studies.

Several caveats qualify the above conclusion. First, thus far in the current paper, no distinction has been made between studies investigating the factor structure of personality disorder *criteria* versus those investigating the factor structure of the personality *disorders*. Although the result mostly clearly in support of the DSM nomenclature would be a solution in which the criteria load on the appropriate diagnosis factors, and the diagnosis factors load, in turn, on three appropriate higher-order factors, the sample size of testing such a model is prohibitive. Thus, this is not our issue of interest. Rather, we are concerned with the following question: if personality disorders exist as defined in the DSM system, can they be reasonably characterized as having the three-factor structure the DSM implies? The question of whether the existing personality disorders can be characterized reasonably as being the ideal categories for abnormal personality traits is another issue altogether, and one that we cannot currently investigate adequately (but see Eussen, Arntz, Hoekstra and Hofman, 1992, cited in Arntz, 1999, for an attempt).

Second, most of the studies cited above use either (a) self-report questionnaires or (b) clinicians' judgments with unknown reliability. In contrast, Bell and Jackson (1992), Moldin, Rice, Erlenmeyer-Kimling and Squires-Wheeler (1994), and Nestadt et al. (1994) used a structured clinical interview for the DSM-III or DSM-III-R. Although self-report measures are understandably the norm in such studies, due to several inherent advantages (such as ease of use), they have a number of potential weaknesses, including bias on the part of respondents. Although clinical judgment alone is also not ideal, self-report in the realm of the personality disorders presents particular problems. By definition, people with some personality disorders are unlikely to be skilled at accurate self-observation. Of the three studies we have located using a structured clinical interview, none have supported the three-factor structure suggested by the DSM system or a

modified FFM solution. (Nestadt et al., 1994 present their solution as relating to the FFM, but closer inspection makes it clear that the five factors they found are significantly different from the standard FFM in a number of important ways. The equation of their 'animation' factor with extraversion and their 'warmth' factor with openness seems particularly problematic.) Thus, the literature lacks clear support for either of the more favoured models when data are collected using the method that appears most likely to yield reliable (and, by implication, at least potentially valid) results.

Third, most of the studies cited above either report either on (a) samples of normal undergraduates or volunteers from the community, or (b) psychiatric inpatients who have relatively severe Axis I psychopathology. Regardless of how one conceptualizes personality and disorders of personality, it would seem desirable to investigate the factor structure of the personality disorders in a variety of samples. Thus, similar studies investigating the factor structure of the personality disorders in outpatient samples are needed.

Finally, a surprisingly large number of studies in the literature have used exploratory factor analytic methods. Exploratory methods are useful when there is exploring to be done. However, several viable factor structure models have been proposed, and the most advantageous strategy for testing them is confirmatory factor analysis of competing models (Bollen, 1989; Floyd and Widaman, 1995). Although confirmatory factor analysis was used, to some degree, in a subsample of the studies we located, only one appears to have involved the appropriate approach of testing competing models (Moldin et al., 1994). However, as the authors of that study note, their approach cannot be called confirmatory factor analysis in the true sense, because they used preliminary results in deciding how to fix factor loadings to permit estimation.

To help address the above concerns, we present our current study, which tests the factor structure of the DSM-III-R personality disorders. Using a combination of self-report questionnaire and structured clinical interview, we collected personality disorder information from 301 outpatients who were diagnosed with one or more anxiety disorders. Employing a confirmatory factor analysis framework, we tested a number of competing models, including the three-factor solution suggested by the DSM system and the four-factor solution suggested by the FFM.

Method

Measures

The *Structured Clinical Interview for DSM-III-R* (Spitzer et al., 1990) was used to determine both Axis I and II disorders. Across the current samples, the SCID was administered by trained staff ranging from BA-level research assistants to doctoral level clinical psychologists. Training for raters included review of the SCID manual (Spitzer, Williams, Gibbon and First, 1989), review and rating of previously administered SCID audiotapes, observation and co-rating of SCID interviews conducted by experienced members of the research staff, roleplay SCIDs, and experienced interviewer supervision for initial SCIDs administered by trainees. Categorical diagnoses were assigned in the standard manner according to DSM-III-R criteria (American Psychiatric Association, 1987). For Axis II personality diagnoses, most ($n = 161$) participants completed the SCID-II questionnaire form, and interviewers then inquired further about those diagnoses for which participants reported some traits (see below).

Samples

Data were obtained from three archival data sets. Although each has been used for other studies, none of those studies concerned the factor structure of the DSM-III-R personality disorders. The first data set ($n = 104$) was from a treatment study focusing on expressed emotion's relationship to treatment outcome in people with panic disorder with agoraphobia and obsessive-compulsive disorder (Chambless and Sketetee, 1999). The second data set ($n = 56$) was from a study of generalized social phobia versus avoidant personality disorder (Fydrich and Chambless, unpublished data). The third data set ($n = 141$) was drawn from clients entering treatment at the Agoraphobia and Anxiety Treatment Center (AATC) of suburban Philadelphia, as part of routine intakes or for research purposes. All participants available in the dataset were used. It is possible that a number of participants in the original studies refused to complete the assessment process, but this number is unknown.

Inter-rater reliability was provided on each subsample. Table 1 displays reliability coefficients per subsample and personality disorder. Average reliability for each data set, as well as the average reliability for each personality disorder (weighted by size of reliability sample) is also displayed in Table 1. In all cases, the

reliability coefficient is the one-way random intraclass correlation coefficient for the single measure. For the first two samples, reliability ratings were obtained, by random selection, for about 25% of the sample. For the subsample drawn from the clinic population, reliability ratings were performed when possible, depending on pragmatic issues (for example, personnel availability). The reliability ratings here are taken from the clinic during the same time period that the data used in the current study were collected.

Overall, in each subsample, and in most personality disorders overall, reliability was acceptable to good, with average reliability ratings ranging from 0.58 to 0.88, with most ratings at 0.70 or above. Reliability was not as good for the APD data set as for the other two; however, the APD sample was smaller than the other two data sets, suggesting that these reliability estimates may themselves be less reliable. Across the data sets, histrionic and narcissistic personality disorders showed the smallest reliability coefficients, although even these coefficients were sufficient for research purposes (defined by Fleiss, 1986, as an intraclass correlation coefficient of > 0.40).

Participants

Across the entire sample used here, most participants were women ($n = 208$), with a smaller number of men ($n = 93$). The average age of participants across samples was 35 years ($SD = 9.56$). All participants had a primary diagnosis of an anxiety disorder. Overall, participants met criteria for panic disorder with agoraphobia ($n = 168$), social phobia ($n = 96$), obsessive-compulsive disorder ($n = 76$), generalized anxiety disorder ($n = 56$), dysthymic disorder ($n = 48$), major depressive disorder ($n = 42$), simple phobia ($n = 33$), post-traumatic stress disorder ($n = 10$), hypochondriasis ($n = 6$), undifferentiated somatoform disorder ($n = 2$), bulimia nervosa ($n = 2$), and adjustment disorder ($n = 2$). As reflected in the above frequency counts, many participants had more than one diagnosis.

Procedure

Most participants ($n = 161$) completed the SCID-II Personality Disorder Questionnaire. Clinicians used the SCID-P to interview participants about any personality disorder to which the participant had responded positively in regard to a minimum number of criteria defined as no fewer than two criteria less

Table 1. Inter-rater reliability by subsample and personality disorder for number of personality disorder criteria met

Personality disorder	EE	APD	AATC	Weighted average
Antisocial	0.62 (23)	NA	0.99 (23)	0.81
Avoidant	0.80 (26)	0.89 (25)	0.84 (26)	0.84
Borderline	0.80 (26)	0.19 (15)	0.78 (25)	0.65
Dependent	0.69 (28)	0.80 (16)	0.85 (25)	0.77
Histrionic	0.78 (27)	0.84 (12)	0.56 (25)	0.62
Narcissistic	0.64 (27)	0.40 (20)	0.67 (25)	0.58
Obsessive-compulsive	0.66 (30)	0.69 (22)	0.90 (25)	0.75
Paranoid	0.63 (26)	0.77 (18)	0.83 (25)	0.74
Passive-aggressive	0.79 (26)	0.79 (13)	0.70 (24)	0.76
Schizoid	0.86 (24)	0.86 (13)	0.91 (24)	0.88
Schizotypal	0.63 (24)	0.71 (9)	0.72 (24)	0.68
Average	0.72	0.69	0.80	0.73

Note. All reliability estimates are one-way random ICCs for absolute agreement of the single measure. Numbers in parentheses are the sample sizes available for the individual analyses. EE = expressed emotion dataset. APD = avoidant personality disorder dataset. AATC = agoraphobia and anxiety treatment centre dataset. The cell marked NA could not be computed because there was no variability in the reliability sample.

than the requirement for diagnosis of the personality disorder. For example, if six criteria had to be met for diagnosis, the module was administered only if the participant endorsed at least four of the criteria on the questionnaire. To reduce participant burden, if the minimum was not met, the interview module for that personality disorder was not administered, and the criteria score was set to zero (see Farmer and Chapman, 2002 for a summary of research and arguments in support of this procedure). The rest of the participants ($n = 140$), who were all from the AATC sample, completed the entire SCID-P interview.

In order to test whether the differing procedures affected the results (below), we fit all acceptable models to each subsample. Thus, we fit the DSM three-factor model to each subsample, both with and without passive aggressive personality disorder criteria in the model. Substantive results remained the same, and it was thus concluded that there were minimal effects of procedure on the fit of the models.

Results

Data analytic strategy

Confirmatory factor analysis procedures were used in both Amos Version 4 (Arbuckle, 1999) and LISREL. Each program provides unique advantages, but statistics from LISREL will be reported (unless otherwise noted), in order to provide consistency and conform to

the guidelines for reporting fit indices proposed by Hu and Bentler (1999). In all cases, the programs provided identical output when each reported the same statistic.

Three fit indices were considered to assess model fit: (a) Bentler's (1990) comparative fit index, CFI, (b) the standardized root mean square residual, SRMR (Jöreskog and Sörbom, 1981), and (c) the root mean square error of approximation, RMSEA (Steiger and Lind, 1980, as cited in Hu and Bentler, 1998), and the 90% confidence interval for the RMSEA. The magnitude of these indices was evaluated based on recommendations by Hu and Bentler (1999). For the CFI, 0.96 or above was considered very good fit, whereas for the RMSEA, less than 0.06, and for the SRMR less than 0.08 were considered very good fit. All three of these indices were considered of equal weight in evaluating fit.

Descriptive statistics

Table 2 presents correlations, means, and standard deviations for each set of personality disorder criteria. No set of criteria was distributed in a completely normal manner, but loglinear transformations were used successfully in rendering sufficiently normal (skewness of an absolute value of less than 1; kurtosis of an absolute value of less than 1.3) all but the set of criteria for antisocial personality disorder. As would be expected in a sample of people who primarily have difficulty with anxiety, the modal number of criteria met

for antisocial personality disorder was 0. In order to account for the effects of this non-normal distribution on the models, attempts were made to estimate all models both with and without antisocial criteria. Unless otherwise noted below, in each case there were no substantive differences between the models with and without antisocial personality disorder. Each primary model was also estimated using both the male ($n = 93$) and female ($n = 208$) subsamples. Unless noted below, no substantive differences were found in model fit or parameter estimates across gender. The models presented below used the entire sample of 301 participants.

Single-factor model

The most parsimonious model for the personality disorders is of one underlying latent factor, which could be called personality pathology. A model with a single latent factor was constructed using the DSM-III-R sets of criteria for the 10 personality disorders retained in the DSM-IV. This model had a fair fit, although fit indices suggested that considerable improvement in fit was possible ($\chi^2(30) = 62.55$, $p < 0.002$; CFI = 0.95; SRMR = 0.06; RMSEA = 0.06, CI = 0.04, 0.09).

DSM three-factor model

A three-factor model, based on the DSM clusters, was

constructed. Error terms were allowed to be intercorrelated when the SCID-P used the same item or question for more than one personality disorder. This was the case for avoidant and schizoid; schizoid and schizotypal; schizotypal and avoidant; schizotypal and paranoid; and avoidant and dependent criteria. The model included three latent factors, each corresponding to a DSM cluster: odd, dramatic, and anxious. These factors were allowed to intercorrelate.

Primary three-factor model.

The primary three-factor model included the 10 primary personality disorders retained in the DSM-IV: paranoid, schizoid, schizotypal, histrionic, narcissistic, antisocial, borderline, avoidant, dependent, and obsessive-compulsive. This model fit the data well ($\chi^2(27) = 44.81$, $p = 0.02$; CFI = 0.97; SRMR = 0.04; RMSEA = 0.05, CI = 0.02, 0.07). In addition, this model was a significant improvement over the one-factor model ($\chi^2(3) = 17.74$; $p < 0.001$). All variables had significant loadings on their latent factor (all p s < 0.05), except in the model for men only, in which the loading for the antisocial variable did not reach significance.

Modification indices did not suggest any theoretically consistent changes in variables' factor loading pattern. The overall model is presented in Figure 1. Factor loadings and squared multiple correlations for

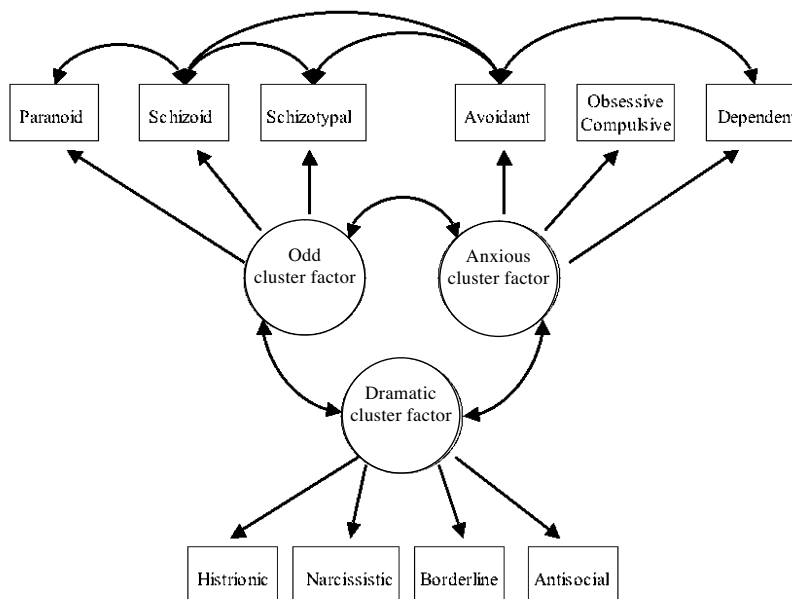


Figure 1. Primary DSM three-factor model with personality disorders retained in the DSM-IV. Single-headed arrows represent factor loadings, and double-headed arrows represent correlations. Note that correlations shown between personality disorders represent correlated error terms.

Table 2. Correlations, means, and standard deviations of personality disorder criteria

	Antisocial	Avoidant	Borderline	Dependent	Histrionic	Narcissistic	OC	PA	Paranoid	Schizoid	ST
Antisocial	(0.81)	0.01	0.15*	0.05	0.10	0.23***	0.05	0.16**	0.07	-0.07	0.10
Avoidant	0.01	(0.84)	0.27***	0.38***	0.16**	0.22***	0.36***	0.32***	0.30***	0.40***	0.53***
Borderline	0.13*	0.30***	(0.65)	0.39***	0.45***	0.48***	0.23***	0.37***	0.38***	0.02	0.25***
Dependent	0.05	0.45***	0.41***	(0.77)	0.35***	0.33***	0.29***	0.39***	0.31***	0.07	0.31***
Histrionic	0.09	0.19**	0.44***	0.36***	(0.62)	0.49***	0.18**	0.32**	0.35***	0.03	0.16**
Narcissistic	0.14*	0.26***	0.42***	0.40***	0.46***	(0.58)	0.24***	0.36***	0.39***	0.07	0.25***
OC	0.12*	0.40***	0.23***	0.31***	0.31***	0.19**	(0.75)	0.32***	0.37***	0.27***	0.31***
PA	0.13*	0.34***	0.37***	0.43***	0.30***	0.37***	0.34***	(0.74)	0.40***	0.08	0.30***
Paranoid	0.07	0.29***	0.37***	0.34***	0.35***	0.40***	0.36***	0.39***	(0.76)	0.23***	0.47***
Schizoid	-0.07	0.39***	0.07	0.12*	0.08	0.10	0.25***	0.14*	0.22***	(0.88)	0.44***
ST	0.08	0.52***	0.25***	0.32***	0.18**	0.24***	0.34***	0.34***	0.46***	0.45***	(0.68)
Mean	0.16	2.44	1.36	1.86	0.84	1.04	1.99	1.23	1.42	0.57	0.97
SD	(.71)	(2.00)	(1.61)	(1.82)	(1.29)	(1.26)	(1.72)	(1.50)	(1.60)	(1.00)	(1.11)

Note. Correlations in the upper right and descriptives use variables prior to linear transformations. OC = obsessive-compulsive, PA = passive aggressive, ST = schizotypal. Diagonal values are weighted mean interrater reliability for personality disorder criteria (cf. Table 1). Lower left values are correlations between the loglinearly transformed variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Factor loadings, standard errors, and squared multiple correlations for DSM three-factor model.

Personality disorder	Odd Factor loading	Dramatic Factor loading	Anxious Factor loading	SMC
Antisocial	–	0.16* (0.06)	–	0.03
Avoidant	–	–	1.00	0.30
Borderline	–	1.00	–	0.44
Dependent	–	–	1.18*** (0.15)	0.44
Histrionic	–	0.85*** (0.10)	–	0.42
Narcissistic	–	0.91*** (0.10)	–	0.48
Obsessive-Compulsive	–	–	0.88*** (0.15)	0.28
Paranoid	1.00	–	–	0.68
Schizoid	0.35** (0.12)	–	–	0.08
Schizotypal	0.64*** (0.09)	–	–	0.39

Note. Factor loadings are maximum likelihood estimates (standard error). Loadings fixed at zero are indicated by dashes. Loadings fixed at unity were constrained to permit estimation, and have no associated standard error. Thus, no direct statistical significance can be computed. SMC = squared multiple correlations.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.00$

this model are presented in Table 3. Correlations among the three factors were large (anxious and dramatic, $r = 0.79$; anxious and odd, $r = 0.71$; dramatic and odd, $r = 0.65$).

Model with the addition of passive-aggressive criteria

A second model, including the criteria for passive-aggressive personality disorder, was constructed. Because of suggestions that passive-aggressive personality disorder should be considered a variant of narcissistic personality disorder (for example, Fossati et al., 2000), the set of criteria for passive-aggressive personality disorder was allowed to load on both the anxious and dramatic cluster. Model fit was excellent ($\chi^2(36) = 47.44$, $p = 0.08$; CFI = 0.98; SRMR = 0.04; RMSEA = 0.03, CI = 0, 0.06). Factor loadings indicated that the set of passive aggressive traits loaded significantly on the anxious factor, but not on the dramatic cluster ($p < 0.01$ and $p > 0.20$, respectively).

Full FFM four-factor model

Two sources were used to create a four-factor model, using agreeableness, conscientiousness, extraversion, and neuroticism from the FFM of normal personality. Austin and Deary (2000) present a summary of relationships between personality disorders and the FFM in their Table 1. We used the relationships marked as 'strong' in this table as our primary source. We supplemented this information with the target loading matrices presented by O'Connor and Dyce (1998).

Following O'Connor and Dyce, we added an association between obsessive-compulsive traits and conscientiousness and between narcissistic traits and agreeableness. Both of these associations are consistent with the criteria for these disorders (APA, 1994). In summary, the factors were defined as follows, with the relevant personality disorders following each factor (expected negative loadings in italics): (a) agreeableness: *narcissistic, antisocial, passive-aggressive, schizotypal, borderline, paranoid*, and *obsessive-compulsive*; (b) extraversion: *schizoid, avoidant*, histrionic, and *schizotypal*; (c) conscientiousness: *antisocial*, obsessive-compulsive, and *passive-aggressive*; (d) neuroticism: avoidant, dependent, passive-aggressive, schizotypal, borderline, and paranoid. Passive aggressive traits were included in this analysis because (a) they were included in both reference sources, and (b) they supplied a third indicator for the factor of conscientiousness, allowing the factor to be adequately represented by the data. Covariances among error variances due to expected shared method variance described in the three-factor model were retained. Factors were allowed to be intercorrelated.

The model as described above proved to be unidentified, according to the output of the AMOS-IV program. Three additional constraints were imposed on the model, according to *a priori* hypotheses, in the following order: (a) covariances among the error variances for schizoid, schizotypal, and avoidant variables were constrained to equality, because shared method variance should be equivalent across these variables;

(b) the loadings of the schizoid and avoidant variables on the extraversion factor were constrained to equality, as suggested by the target loading matrices reported by O'Connor and Dyce (1998); (c) also as suggested by O'Connor and Dyce's target loading matrices, the loadings of the antisocial, paranoid, and narcissistic variables on the agreeableness factor were constrained to equality. With these three additional constraints, the model converged. However, an improper solution, in the form of a non-positive definite covariance matrix among the latent factors, was reached. The same improper solution was found both in AMOS 4 and LISREL and resulted from correlations between the conscientiousness, agreeableness, and extraversion factors that were greater than 1.

Examination of the factor loadings and the correlation matrix of the personality disorder criteria showed that many of the personality disorders displayed a stronger relationship with disorders in a factor they were not permitted to load on than with the disorders in the factor(s) they were permitted to load on. The indefiniteness problem noted above can result from such relationships among variables and is, in itself, a strong indication of poor model fit (Wothke, 1993). Moreover, patterns of factor loadings in the improper solution differed from those that were hypothesized. Most notably, the schizoid and avoidant variables loaded positively on the extraversion factor, when they were expected to load negatively. This unexpected pattern held true in the male subsample, and was slightly different among women: Not only were loadings of the avoidant and schizoid variables on the extraversion factor positive, but the expected positive loading for histrionic traits on this factor was not obtained. Although modification indexes were examined, they suggested no feasible changes. It seemed inadvisable to examine this model without the antisocial variable, because deleting the variable would lead to further identification problems.

Discussion

Previous attempts to determine the factor structure of the personality disorders have been marked by methodological and statistical problems, resulting in ambiguous findings that are not consistent across studies. To our knowledge, we present the first attempt to use confirmatory factor analysis in the proper sense to test competing models of the factor structure of the personality disorders. The use of a structured clinical

interview with known reliability is another feature that has been sorely lacking in the available literature.

Our findings provide strong support for the DSM three-factor model, with no support for a model based on normal personality. In addition to estimation problems indicating extremely poor fit, the model based on normal personality was also undermined by unexpected factor loadings in the obtained solutions. Overall, the current results suggest that the DSM three-factor model has considerable validity in the current context. It should be noted that exploratory factor analyses supporting a four-factor model or different three-factor models are not necessarily in direct contradiction to our findings. Exploratory and confirmatory factor analyses are conceptually distinct procedures, and, when theories regarding factor structures are available, a confirmatory approach allows a more rigorous test of model viability (Bollen, 1989; Floyd and Widaman, 1995).

Our findings certainly do not rule out the possibility that models based on normal personality may offer an important way to understand disordered personality. The FFM, for example, seems somewhat at a disadvantage to explain the currently accepted set of personality disorders. Besides the relative lack of support for a relationship between openness and personality disorders, a large number of the personality disorders involve high neuroticism, either high or low extraversion, and low agreeableness. Thus, reliance on the five broad factors provides relatively little distinction between the personality disorders. In contrast, the use of facets of the five factors, as suggested by Lynam and Widiger (2001) provides greater specificity, although it could not be appropriately evaluated in this paper. Similarly, other studies finding strong support for the FFM in disordered personality do not use the full set of disorders or criteria in the DSM (for example, Livesley, Lang and Vernon, 1998). Overall, we find the FFM and other models of normal personality to be highly valuable to the understanding of normal personality and its extremes and we do not wish to imply that our current results are a strong mark against these models as broadly applied. However, in reference to the question of how the personality disorders of the DSM may be grouped most appropriately, we find no evidence in the current study that a model based on the FFM provides any advantage over the DSM three-factor system.

We included passive-aggressive criteria in some analyses, although this personality disorder is only

included as a research diagnosis in the DSM-IV. Our results suggest that passive-aggressive personality disorder, as defined by the DSM-III-R, belongs in the anxious cluster, as, indeed, it was placed in that version of the DSM. These results are in contrast to those of Fossati et al. (2000), who found that passive-aggressive personality had a specific relationship to narcissistic personality disorder, which would imply that it should be on the dramatic cluster. However, in reviewing nine studies, Widiger et al. (1991) noted that passive-aggressive personality disorder appears to correlate moderately with many personality disorders, suggesting it may be more of a general maladaptive style than a specific personality configuration. If that conclusion is correct, one would expect passive-aggressive personality disorder to load on whatever factor is most highly represented in a given sample; in our sample, disorders in the anxious cluster were over-represented. Thus, it may be premature to conclude, based on our data, that passive-aggressive personality disorder truly belongs in the anxious cluster.

Limitations of this study include reliance on DSM-III-R criteria, the characteristics of the sample, and the method of data collection. In regard to the first, practical restraints prevented us from using DSM-IV criteria. The data collection effort represented in the current study was considerable, and spanned 8 years; we have essentially opted out of the use of up-to-date diagnostic criteria in exchange for a sample size that is appropriately large for confirmatory factor analysis. In addition, because there appear to be no similar extant studies using DSM-IV criteria, our use of DSM-III-R criteria actually makes our study more comparable to existing studies. Finally, existing data on a variety of participants suggest that most of the personality disorders included in the DSM-IV show strong relationships with their DSM-III-R versions, with the exception of dependent and histrionic personality disorders, which repeatedly show lower reliability coefficients (for both: Poling et al., 1999; Sunday et al., 2001; for histrionic: Blais, Hilsenroth and Castlebury, 1997).

In the individual studies, some other personality disorders had low reliability coefficients, but these were probably due either to too few participants meeting either DSM-III-R or DSM-IV criteria (for example, Sunday et al., 2001 found a low coefficient for schizotypal, but only four participants met criteria under either DSM system), or were inconsistent across studies.

Our use of criteria totals, rather than diagnoses or the criteria themselves, improves the probability that our results would apply to the DSM-IV personality disorders, in that the DSM-IV version of histrionic personality disorder was found to have an acceptable relationship with its previous version when the corresponding change in diagnostic threshold was ignored (Poling et al.). That is, the poor agreement was primarily related to whether participants met the diagnosis (for which the threshold had been altered), not to the number of criteria met. Poling et al. speculate that the changes made to the DSM criteria for dependent personality disorder inherently make reliability more difficult to achieve, suggesting that the weaker relationship between the DSM-III-R and DSM-IV versions of the disorder may be a result of low reliability of the DSM-IV version. Thus, all available evidence we are aware of suggests that our results are generalizable to the DSM-IV system, with the possible exception of dependent personality disorder, although the evidence for even that exception is equivocal.

Although the use of an outpatient sample is a strength of the present study, the relative infrequency of some personality disorders represents a limitation. These results require replication in an outpatient sample with a more evenly distributed range of personality pathology. In essence, the concern is for generalizability because the relative lack of criteria met for some personality disorders appears to be tied to the use of a sample of people with anxiety disorders. Nonetheless, anxiety disorders are among the most common complaints encountered in clinical practice, thus making the current sample largely representative of many types of outpatient clinical samples. The latter point is especially clear when the high rate of mood disorders in the current sample (primarily major depressive disorder and dysthymia) is also taken into account.

On initial evaluation, it may seem plausible that characteristics of the sample may have produced the unusual loadings in the FFM model. That is, perhaps in a sample more representative of the range of personality disorders the FFM model would have shown the expected properties. (We thank an anonymous reviewer for raising this concern based on a previous version of this paper.)

We cannot absolutely rule this argument out given our current data, but its assumptions are not logically

consistent upon further examination. First, the results of the FFM model are, by definition, not interpretable, because the model failed to converge on a proper solution. We present the factor loadings not for interpretation but as evidence of the problems with the model. Second, the majority of the personality disorder criteria were represented well enough to obtain approximately normality (albeit using transformations in some cases) in all but one set of criteria. Third, because the FFM is primarily based on *normal* personality, one would logically expect it to best fit in a sample of people with a normal range of personality characteristics. A sample selected particularly for the presence of personality disorders is unlikely to present a normal range of personality. Arguably, then, our sample of participants with anxiety disorders should present less trouble for the FFM than a sample with more overall personality pathology. We wish to iterate, however, that we do not believe our current results should be taken as an indication that the broader application of the FFM to normal personality is invalid. Nor would we necessarily be averse to an account of disordered personality using a dimensional format.

Finally, our method of data collection, a structured clinical interview, is a strength in terms of reliability and accuracy, but it also presents some interpretation problems. First, although reliability was acceptable to good for most personality disorders, there were some instances of low agreement between raters. However, low agreement primarily limits the extent to which a given construct can relate to another. Therefore, if reliability were an issue in the current sample, it would have primarily related to poor model fit across all models, which did not occur. Second, the questionnaire and interview method used for the majority of the sample, while time-efficient, did not involve the interviewer's asking about and evaluating all criteria in all participants. It is unclear how the interview procedure would have affected a confirmatory factor analysis but it would obviously have been desirable for all criteria to have been evaluated in all participants. However, all criteria were evaluated for a large minority of the sample, and our results remained substantively unchanged in that subsample. In addition, our interviewers were at least generally aware of the DSM-III-R, and it remains possible that their expectations about personality disorders influenced the current data. It appears safe to assume, however,

that most of our participants were not well acquainted with the DSM-III-R, and their input was also important in the data collection process. Furthermore, the assessment materials themselves make no reference to the DSM grouping of personality disorders, and neither the participants nor the interviewers were aware that a study such as this one might be conducted using archival data from the original studies. It is unclear what methodology would allow reliable measurement of personality disorder criteria without raising the possibility of expectations on the part of the people giving ratings. However, it is also unclear how the methodology could have produced such systematic bias that it alone could account for the current results.

We believe that the above limitations are considerably outweighed by the strengths of the study. Indeed, given the controversy over the personality disorders, a careful review of the literature reveals a striking lack of methodologically and statistically rigorous evaluations of how best to group the personality disorders. Although by no means conclusive, the results of this study suggest that the DSM grouping has considerable conceptual utility and statistical merit in that it appears to reflect adequately the underlying latent variables reflected in the personality disorders.

Acknowledgements

Collection of these data was supported by NIH grants R01-MH44190 to Dianne Chambless and Gail Steketee and R03-MH47968 to Dianne Chambless and Thomas Fydrich; Drs Fydrich and Renneberg were supported by the Advanced Council for Learned Societies and the German Academic Exchange Service; and Thomas Rodebaugh by NIMH grant 1 F31 MH65005-01. We thank Dr Steketee for access to her data and James Arbuckle for statistical consultation.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3 edn. Washington DC: APA, 1980.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3 edn, revised. Washington DC: APA, 1987.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4 edn. Washington DC: APA, 1994.
- Arbuckle JL. Amos 4.0 User's Guide. Chicago IL: Small Waters Corp., 1999.
- Arntz A. Do personality disorders exist? On the validity of the concept and its cognitive-behavioral formulation and treatment. Behaviour Research and Therapy 1999; 37: 97-137.

- Austin EJ, Deary IJ. The 'four As': a common framework for normal and abnormal personality? *Personality and Individual Differences* 2000; 28: 977–95.
- Bell RC, Jackson HJ. The structure of personality disorders in DSM-III. *Acta Psychiatrica Scandinavia* 1992; 85: 279–87.
- Bentler PM. Comparative fit indices in structural models. *Psychological Bulletin* 1990; 107: 238–46.
- Blaise MA, Hilsenroth MJ, Castlebury FD. Psychometric characteristics of the cluster B personality disorders under DSM-III-R and DSM-IV. *Journal of Personality Disorders* 1997; 11: 270–81.
- Bollen KA. *Structural Equations with Latent Variables*. New York: Wiley, 1989.
- Chambless DL, Steketee G. Expressed emotion and the prediction of outcome of behavior therapy: a prospective study with agoraphobic and obsessive-compulsive outpatients. *Journal of Consulting and Clinical Psychology* 1999; 67: 658–65.
- Dowson JH, Berrios GE. Factor structure of DSM-III-R personality disorders shown by self-report questionnaire: implications for classifying and assessing personality disorders. *Acta Psychiatrica Scandinavia* 1991; 84: 555–60.
- Deary IJ, Peter A, Austin E, Gibbo G. Personality traits and personality disorders. *British Journal of Psychology* 1998; 89: 647–61.
- Farmer RF, Chapman AL. Evaluation of DSM-IV personality disorder criteria as assessed by the structured clinical interview for DSM-IV personality disorders. *Comprehensive Psychiatry* 2002; 43: 285–300.
- First MB, Spitzer RL, Gibbon M, Williams JBW, Davies M, Borus J, Howes MJ, Kane J, Pope HG Jr., Rounsaville B. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II), Part II: Multi-site test-retest reliability study. *Journal of Personality Disorders* 1995; 9: 92–104.
- Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: John Wiley, 1986.
- Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 1995; 7: 286–99.
- Fossati A, Maffei C, Bagnato M, Donati D, Donini M, Fiorilli M, Novella, L. A psychometric study of DSM-IV passive-aggressive (negativistic) personality disorder criteria. *Journal of Personality Disorders* 2000; 14: 72–83.
- Gorsuch RL. *Factor Analysis*. 2 edn. New Jersey: Lawrence Erlbaum, 1983.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 1999; 6: 1–55.
- Hyder SE, Lyons M. Factor analysis of the DSM-III personality disorder clusters: a replication. *Comprehensive Psychiatry* 1988; 29: 304–8.
- Hyder SE, Lyons M, Rieder RO, Young L, Williams JBW, Spitzer RL. The factor structure of self-report DSM-III Axis II symptoms and their relationship to clinicians' ratings. *American Journal of Psychiatry* 1990; 147: 751–7.
- Jöreskog KG, Sörbom D, Lisrel V. *Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources, 1981.
- Kass F, Skodol AE, Charles E, Spitzer RL, Williams JBW. Scaled ratings of DSM-III personality disorders. *American Psychiatry* 1985; 142: 627–30.
- Lee HB, Comrey AL. Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research* 1979; 14: 301–21.
- Livesley WJ, Lang KL, Vernon PA (1998) Phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry* 1998; 55: 941–8.
- Lynam DR, Widiger TA. Using the five-factor model to represent the DSM-IV personality disorders: an expert consensus approach. *Journal of Abnormal Psychology* 2001; 110: 401–12.
- Moldin SO, Rice JP, Erlenmeyer-Kimling L, Squires-Wheeler E. Latent structure of DSM-III-R Axis II psychopathology in a normal sample. *Journal of Abnormal Psychology* 1994; 2: 259–66.
- Nestadt G, Eaton WW, Romanoski AJ, Garrison R, Folstein MF, McHugh PR. Assessment of DSM-III personality structure in a general-population survey. *Comprehensive Psychiatry* 1994; 35: 54–63.
- O'Connor BP, Dyce JA. A test of models of personality disorder configuration. *Journal of Abnormal Psychology* 1998; 107: 3–16.
- Poling J, Rounsaville BJ, Ball S, Tennen H, Kranzler HR, Triffleman E. Rates of personality disorders in substance abusers: A comparison between DSM-III-R and DSM-IV. *Journal of Personality Disorders* 1999; 13: 375–84.
- Sanislow CA, Morey LC, Grilo CM, Gunderson JG, Tracie Shea M, Skodol AE, Stout RL, Zanarini MC, McGlashan TH. Confirmatory factor analysis of DSM-IV borderline, schizotypal, avoidant and obsessive-compulsive personality disorders: findings from the Collaborative Longitudinal Personality Disorders Study. *Acta Psychiatrica Scandinavica* 2002; 105: 28–36.
- Spitzer RL, Williams JBW, Gibbon M, First MB. *Structured Clinical Interview for DSM-III-R Patient Version*. New York: Biometrics Research Department, New York State Psychiatric Institute, 1989.
- Spitzer RL, Williams JBW, Gibbon M, First MB. *Structured Clinical Interview for DSM-III-R Patient Edition (SCID-I)*. Washington, DC: American Psychiatric Press, 1990.
- Sunday SR, Peterson CB, Adreyka KA, Crow SJ, Mitchell JE, Halmi KA. Differences in DSM-III-R and DSM-IV

- diagnoses in eating disorder patients. *Comprehensive Psychiatry* 2001; 42: 448–55.
- Tyrer P, Casey P, Ferguson B. Personality disorder in perspective. *British Journal of Psychiatry*, 1991; 159: 462–71.
- Widiger TA, Frances A, Harris J, Jacobsberg LB, Fyer M, Manning D. Comorbidities among Axis II disorders. In J Oldham (ed.) *Personality Disorders: New Perspectives on Diagnostic Validity*. Washington, DC: American Psychiatric Press, 1991, pp. 163–94.
- Wothke W. Nonpositive definite matrices in structural equation modeling. In K Bollen, S Long (eds) *Testing Structural Equation Models*. Newbury Park: Sage, 1993, pp. 257–93.
- Correspondence: DL Chambless, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104-6241, USA.
Telephone +01 215 8985030.
Fax +01 215 8987301.
Email chambless@psych.upenn.edu.*