Person. individ. Diff. Vol. 11, No. 5, pp. 515–524, 1990
Printed in Great Britain. All rights reserved

0191-8869/90 $3.00 + 0.00
Copyright © 1990 Pergamon Press plc

COMPARING EXPLORATORY AND CONFIRMATORY FACTOR ANALYSIS: A STUDY ON THE 5-FACTOR MODEL OF PERSONALITY

PETER BORKENAU and FRITZ OSTENDORF
Universität Bielefeld, Postfach 8640, D-4800 Bielefeld, F.R.G.

(Received 3 July 1989)

Summary—Several authors claim that widespread support exists for a 5-factor model of personality ratings. In the present study, structural equation modeling was used to investigate this issue. The subjects (128 males and 128 females) were administered Costa and McCrae's NEO Personality Inventory, and they rated themselves, and were rated by three acquaintances, on the 20 adjective scales suggested by Norman as marker variables for the Big Five. Coefficients of factor comparability indicated that a 5-factor model accounted for the data better than any other model. Moreover, the five factors that were obtained matched conventional measures of the Big Five very well. A multitrait–multimethod analysis with five traits and three methods yielded acceptable convergent and discriminant validities, and a model with oblique trait as well as oblique method factors was supported by structural equation models. A confirmatory factor-analytic model, however, that predicted the correlations among 60 variables from five trait factors and three method factors, did not fit the data. It is concluded that this finding reflects a desirable heterogeneity of personality factors as higher-level constructs. The implications for the usefulness of confirmatory factor analysis as well as for the 5-factor model of personality are discussed.

Several authors claim that widespread support exists for a 5-factor model of personality trait ratings (Digman & Inouye, 1986; Digman & Takemoto-Chock, 1981; Fiske, 1949; Goldberg, 1989; McCrae & Costa, 1987; Norman, 1963; Tupes & Christal, 1961). According to numerous exploratory factor analyses, a 5-factor model meets the highly desirable criteria of robustness across targets (Norman, 1963; Tupes & Christal, 1961), across observers (Fiske, 1949; Tupes & Christal, 1961), and across methods of factoring and factor rotation (Goldberg, 1989). Moreover, several studies indicate that a 5-factor model accounts also for the correlations among unbiased samples of personality inventory scales (Amelang & Borkenau, 1982; Borkenau & Ostendorf, 1990; McCrae & Costa, 1987; Noller, Law & Comrey, 1987), and for the factor structure of the California Q-set (McCrae, Costa & Busch, 1986).

Given this convergence of the results of numerous exploratory factor analyses, it seems, at first glance, highly recommendable to apply the method of confirmatory factor analysis to data that support the 5-factor model in exploratory factor analyses. Mulaik (1987), for example, suggested that exploratory factor analysis should be regarded as a "hypothesis-generating method, providing information for the researcher to use in formulating hypotheses" (p. 302). This, however, "demands our finding a way of using experience ... going beyond the specific set of data stimulating the hypothesis. This can be done only by testing hypotheses with additional data. Hence, confirmatory common factor analysis is a logical sequel to exploratory common factor analysis" (Mulaik, 1987, p. 302).

Confirmatory factor analysis, however, is a quite restrictive tool to study the relations among concepts. First, whereas factor models that are established by exploratory factor analyses of personality trait ratings usually account for about 60% of the total variance in a set of rating data (Peabody & Goldberg, 1989), confirmatory factor analysis requires that the observed correlations do not significantly differ from those that are predicted by the model. Otherwise, the model has to be rejected. When quantitative fit indices are used instead of a χ^2 goodness of fit statistic, it is suggested that the indices exceed 0.90 to be recommended as acceptable (Bentler & Bonett, 1980). Thus, in order that a model is acceptable in terms of the standards that are used in confirmatory factor analysis, it has to account for a much higher proportion of the variance (Tucker & Lewis, 1973) than is usually accounted for by models that are established via exploratory factor analyses.

Second, confirmatory factor analysis requires that the variables have literally zero-loadings on all factors, except the appropriate ones which are *a priori* specified in the model. This, however, is no reasonable assumption in case of the factor structure of personality trait ratings. Because factors are conceptualized as higher level constructs in most factor-analytic theories of personality (Cattell, 1965; Costa & McCrae, 1985; Eysenck & Eysenck, 1969; Guilford, 1975), they refer to a quite diverse domain of lower-level constructs such as primary factors, facets, habits, or specific behaviours. Eysenck and Eysenck (1969), for instance, emphasize the dual nature of Extraversion. They suggest that Sociability has a negative secondary loading on Neuroticism and that Impulsivity has a positive secondary loading, whereas both, Sociability and Impulsiveness, constitute the factor Extraversion. Cattell (1950) suggests that the marker variables for a factor may be heterogeneous, but that their center of gravity should coincide with the position of the factor. In contrast, confirmatory factor analysis allows the observed variables to differ with respect to their loadings on the appropriate factors only, whereas secondary loadings on other factors are treated as deviations from the model. Thus the model of confirmatory factor is much more restrictive than the model of exploratory common factor analysis.

It may be argued that the *methodology* of maximum likelihood factor analysis is not restrictive because models may be tested that are not restrictive at all. Thus one may go as far as to allow each variable to load on each factor, and the loadings of all variables on all factors may then be estimated. Rarely, however, are hypotheses available that allow reasonable predictions for such complex factor patterns. Thus, in this case, maximum likelihood factor analysis was used in an entirely exploratory way. Usually, it is only one or two factor loadings per variable that can be specified *a priori*. If only these factor loadings are allowed to deviate significantly from zero, the term *Confirmatory factor analysis* is appropriately applied.

Structural equation models, however, have the particular advantage to make the distinction between content factors and method factors possible. This is especially desirable for the 5-factor model of personality because appropraite marker variables are available from several varieties of data, namely, inventory scores, self-ratings on adjective scales, and peer ratings on adjective scales (Amelang & Borkenau, 1982; McCrae & Costa, 1987). But it is not reasonable to suppose that content is the only source of covariance among these measures of personality. Rather, the correlations among ratings of a given target are expected to be higher if the ratings stem from the same judges than if the ratings stem from different judges. Method components should therefore be considered in addition to content components, as the correlations among variables are expected to be highest if they share content as well as method variance.

There exists a method that uses this desirable property of structural equation models while simultaneously avoiding some of the disadvantages of confirmatory factor analysis. This is to evaluate multitrait-multimethod matrices by means of structural equation models. We use the term *multitrait–multimethod analysis* if the variables that are expected to load the same trait factor are homogeneous in content. In contrast, we speak of *confirmatory factor analysis* if the variables presumably loading the same trait factor are somewhat heterogeneous in content (as, for instance, the various primary factors of extraversion).

To our knowledge, measures of the five factors have not been evaluated before by means of structural equation models. We admit that multitrait–multimethod matrices have been established for the five factors (Amelang & Borkenau, 1982; McCrae & Costa, 1987), but they were evaluated in previous studies with the traditional Campbell and Fiske (1959) criteria. These criteria, however, may be criticized to the effect that: (a) testing the statistical significance of the overall pattern is questionable, due to the lack of independence of the correlations used; (b) precise estimates of trait-related and method-related variance for each measure are not obtainable, whereas such measures were very useful to indicate which measures should be refined; and (c) differences among variables in their level of reliability will distort the observed correlations among the measures. Structural equation models are most appropriate to overcome these ambiguities (Judd, Jessor & Donovan, 1986; Schmitt & Stults, 1986; Widaman, 1985).

Whereas the analysis of multitrait–multimethod matrices with structural equation models is useful to evaluate the equivalence of measures of the five factors across data-levels, this approach does not inform about the comparability of the factor patterns. A method that investigates the generalizability of factors across S samples, and that is less restrictive than confirmatory factor

analysis, has been suggested by Everett (1983). This procedure comprises three steps: (a) two or more independent factor scoring coefficient matrices are established from rotated factor solutions of the same set of variables for several *S* samples; (b) two or more factor scores are established for each factor for the combined *S* sample, each one relying on one of the several factor scoring coefficient matrices; and (c) the comparability of the factors is assessed by correlating the different factor scores for the same factor across *S*s (Everett, 1983; McCrae & Costa, 1987; Zuckerman, Kuhlman & Camac, 1988). This method also goes beyond the specific set of data stimulating the hypothesis. It does not imply the restrictive assumptions, however, that are implicit in confirmatory factor analysis.

The present study evaluated the 5-factor model of personality by means of several methods. Sets of presumed marker variables for the five factors, including inventory scales, self-ratings on adjective scales, and peer ratings on adjective scales, were administered to 256 *S*s. The factor solutions were then evaluated using three criteria. First, coefficients of factor comparability across male and female *S* samples (Everett, 1983) were separately determined for the personality inventory scales, the self-ratings on adjective scales, and the peer ratings on adjective scales. This approach established the generalizability of factor solutions across sex. Second, structural equation models were used to evaluate multitrait–multimethod matrices with five trait factors and three method factors. This procedure estimated the equivalence, across data-levels, of comprehensive measures for the five factors. Finally, a confirmatory factor analysis with 60 variables was carried out.

METHOD

Subjects

*S*s were 256 German adults (128 females, 128 males). Their mean age was 26.4 yr and their age standard deviation was 9.84 yr. The *S*s were recruited via an article in the local newspapers, pursued various professions, and were paid for their participation. A precondition for their participation was that each *S* had to be accompanied by three relatives and/or acquaintances who had to provide peer ratings under the experimenter's supervision. In this way, the independence of self-reports and peer reports was secured.

Measures

All measures that were used were presumed markers for one of the five major factors of personality. Otherwise, the assignment of variables to trait factors had been arbitrary in the confirmatory factor analyses as well as in the multitrait–multimethod analyses. Thus, to obtain personality inventory data, we translated and used Costa and McCrae's (1985) NEO Personality Inventory. It measures Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Furthermore, Costa and McCrae (1985) distinguish among six facets of each, N, E, and O, whereas A and C are treated as unitary constructs. To obtain self-ratings and peer ratings on adjective scales, we translated and administered the 20 scales that were suggested by Norman (1963). These scales are usually regarded as marker variables for the Big Five. Four of the scales refer to Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Culture, respectively. We considered Emotional Stability to be the opposite of Neuroticism, whereas the terms 'Openness to Experience' and 'Culture' denote highly similar constructs. Thus it is only for reasons of convenience, that the terms suggested by Costa and McCrae (1985) are used in the present article.

RESULTS

The comparability of factors

We used Everett's (1983) procedure to investigate: (a) what the number of factors was for which factor scores were best replicated across *S*s, and (b) whether the factors obtained in the 5-factor solutions corresponded to other well-established measures of the Big Five. Concerning the first issue, Everett suggested that the number of factors to be retained and interpreted should be determined by obtaining rotated solutions from different subject samples, and adopting the solution

Table 1. Comparabilities of varimax-rotated factors in the combined sample

Data level	Factors rotated	Factor comparabilities after varimax rotation: Factor 1st	2nd	3rd	4th	5th	6th	7th	8th
Inventory scales (NEO-PI)	8	0.96	0.96	0.95	0.79	0.72	0.40	0.31	0.15
	7	0.97	0.95	0.95	0.76	0.70	0.31	0.09	
	6	0.97	0.96	0.96	0.78	0.75	0.34		
	5	0.98	0.96	0.95	0.93	0.87			
	4	0.94	0.92	0.82	0.81				
	3	0.93	0.88	0.74					
Self-ratings on Norman's adjectives	8	0.94	0.98	0.90	0.81	0.80	0.69	0.65	0.45
	7	0.98	0.96	0.85	0.76	0.70	0.61	0.27	
	6	0.98	0.95	0.90	0.89	0.70	0.67		
	5	0.98	0.97	0.96	0.90	0.78			
	4	0.97	0.89	0.70	0.68				
	3	0.95	0.91	0.65					
Peer ratings on Norman's adjectives	8	0.99	0.99	0.97	0.93	0.87	0.65	0.22	0.19
	7	0.99	0.99	0.97	0.95	0.92	0.58	0.56	
	6	0.99	0.99	0.96	0.96	0.94	0.81		
	5	0.99	0.98	0.97	0.96	0.94			
	4	0.97	0.82	0.75	0.68				
	3	0.94	0.79	0.67					

$n = 256$.

that can be replicated. Everett (1983) suggested that, if his method was used, comparability coefficients above 0.90 indicated a factor match.

The sample was divided according to sex, and separate factor analyses were conducted for the 128 male and the 128 female subjects. Factor scoring coefficients were then obtained from both independent analyses. These factor scoring coefficients were used to estimate two factor scores per factor for the combined sample of 256 *S*s. The two parallel factor scores were then correlated among each other across the 256 *S*s. This procedure was repeated for 3-, 4-, 5-, 6-, 7-, and 8-factor solutions as well as for personality inventory scores, self-ratings on adjective scales, and peer ratings on adjective scales.

In the factor analyses of inventory scores, we used the NEO-PI but not its usual scoring key. Rather, we randomly divided each of the scales A and C into three subscales, to ensure that the factors A and C were each represented by several variables and had thus a chance to turn out as separate factors. Accordingly, 24 variables were entered, i.e. the six facets of N, E, and O, and three NEO-A and NEO-C subscales. The 20 scales suggested by Norman (1963) as markers for the Big Five were used in the factor analyses of self-ratings and peer ratings on adjective scales. In all analyses of peer ratings, the ratings by three peers were first averaged in order to obtain more reliable raw data. Iterative procedures were used to estimate communalities, and the factors were submitted to a Varimax rotation. The coefficients of factor comparability are reported in Table 1.

It was only for the peer ratings that a solution was obtained that met Everett's $r = 0.90$-criterion for each factor, and this was the 5-factor solution. For the inventory scales and the self-ratings, the 5-factor solution yielded higher indices of factor-match than any other solution, although the match for the fifth factor did not come up to the 0.90-criterion (which was, however, somewhat arbitrary). Thus, for all three data-levels, a 5-factor solution, more than any other solution, yielded factors that were replicated across *S*s. The next question then was whether these five factors could be interpreted as the conventional Big Five. The two factor scores per factor (their correlations were reported in Table 1) were therefore correlated with available conventional measures of the five factors. As conventional inventory measures of the Big Five, the N, E, O, A, and C domain scores were used that were obtained by applying the usual NEO scoring key (Costa & McCrae, 1985). As comprehensive self-rating and peer rating measures, being based on adjective scales, the unweighted sums of the ratings on the four marker scales for each of Norman's factors were used. Accordingly, the self-rating on adjective scales for Extraversion was the sum of the self-ratings on the four scales *talkative–silent*, *frank*, *open–secretive*, *adventurous–cautious*, and *sociable–reclusive*. Correspondingly, the peer rating for Extraversion was the sum of the ratings by the three peers on these four scales. The correlations between these measures and the factor scores are reported in Table 2, indicating that the 5-factor solutions were not only replicated across *S*s, but that the corresponding factor scores also matched conventional measures of the Big Five very well.

Table 2. Correlations between conventional measures of the five factors and the factor scores obtained in the exploratory factor analyses

Factor	Factor scoring coefficients obtained from Male *Ss*	Female *Ss*
Inventory scales:		
Neuroticism	0.92	0.93
Extraversion	0.88	0.97
Openness to Experience	0.95	0.93
Agreeableness	0.93	0.93
Conscientiousness	0.95	0.95
Adjective self-ratings:		
Neuroticism	0.93	0.95
Extraversion	0.96	0.96
Culture	0.92	0.85
Agreeableness	0.96	0.91
Conscientiousness	0.93	0.96
Adjective peer ratings:		
Neuroticism	0.92	0.94
Extraversion	0.98	0.98
Culture	0.87	0.96
Agreeableness	0.94	0.95
Conscientiousness	0.93	0.90

$n = 256$.

The exploratory factor analyses therefore demonstrated the superiority of the 5-factor model. For the present set of variables, the 5-factor solution yielded factors that were best replicated across sex and that were highly correlated with conventional measures of the Big Five.

Multitrait–multimethod analyses

Next, a multitrait–multimethod matrix was established for the 15 conventional measures of the five factors mentioned above. The correlations among these measures were separately calculated for the 128 male and the 128 female *Ss*. Table 3 reports these correlations.

The correlations were similar for male and female *Ss*, in that: (a) the convergent validities were higher than the discriminant validities, and they were highest for the factors Extraversion and Conscientiousness; (b) the correlations between inventory data and self-ratings were higher than those between self-ratings and peer ratings, and (c) some substantial heterotrait–monomethod coefficient were obtained for both sexes (e.g. negative correlations between Neuroticism and Conscientiousness).

Structural equation modelling. We used maximum-likelihood estimation of latent variable structural models with LISREL VI (Jöreskog & Sörbom, 1984). The χ^2 provided by this procedure indicates whether the correlation matrix reproduced by a given model is significantly different from the observed correlation matrix. Statistical nonsignificance implies that a model cannot be rejected on statistical grounds and may provide an adequate representation of the data. Unfortunately, the likelihood of rejecting a model based on χ^2 depends strongly on sample size. On the one side, models that explain essentially all of the relevant information in a set of data based on a large sample may still be rejected on statistical grounds. On the other hand, too many models may appear to provide adequate fit to data from small samples if only statistical significance is relied upon. Thus, to complement the χ^2 statistic, two quantitative measures of fit, ρ (Tucker & Lewis, 1973) and Δ (Bentler & Bonett, 1980), that have both been suggested as being independent of sample size, were applied.* Whereas ρ is not normed to necessarily lie between zero and one, Δ is a normed index. Both indices represent the increment in fit obtained by using k common factors rather than none; they compare a null model of independence among all variables with a model with k common factors. According to Bentler and Bonett (1980), these indices of overall fit should exceed 0.90 before a model is accepted.

*It has recently been reported by Marsh, Balla and McDonald (1988) that, in their study, Bentler and Bonett's index Δ was substantially affected by sample size, whereas Tucker and Lewis' index ρ was by and large independent of sample size. For this reason, the Tucker–Lewis index may be regarded as more appropriate than the Bentler–Bonett index for the evaluation of the structural equation models.

Table 3. Correlations in the multitrait–multimethod matrix

	Inventory scales					Self-ratings					Peer ratings				
	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C
N		0.00	0.25	−0.08	−0.32	0.59	−0.13	−0.10	−0.24	−0.32	0.27	0.01	−0.03	−0.14	−0.20
E	0.05		0.35	−0.07	−0.12	−0.15	0.64	0.12	0.10	−0.22	0.05	0.46	−0.15	−0.09	−0.38
O	0.28	0.31		−0.04	−0.23	0.06	0.18	0.39	−0.04	−0.13	0.07	0.10	0.17	−0.05	−0.22
A	−0.09	−0.06	−0.04		0.18	−0.18	0.02	−0.04	0.56	0.27	−0.02	−0.10	−0.15	0.29	0.17
C	−0.38	−0.06	−0.29	0.23		−0.09	−0.13	−0.01	0.01	0.64	0.06	−0.15	−0.10	−0.03	0.43
N	0.59	−0.12	0.10	−0.18	−0.15		−0.23	−0.19	−0.41	−0.30	0.31	−0.03	0.07	−0.08	−0.15
E	−0.10	0.66	0.17	0.07	−0.06	−0.23		0.27	0.34	−0.08	0.04	0.52	−0.11	0.10	−0.31
O	−0.05	0.11	0.39	−0.08	−0.03	−0.14	0.24		0.28	0.14	−0.03	−0.02	0.21	−0.10	−0.05
A	−0.23	0.08	−0.04	0.54	0.06	−0.38	0.32	0.30		0.27	−0.09	0.10	−0.07	0.24	0.02
C	−0.39	−0.17	−0.20	0.29	0.70	−0.33	−0.05	0.09	0.28		−0.13	−0.21	−0.08	0.11	0.51
N	0.25	0.12	0.14	−0.06	−0.04	0.27	0.06	−0.02	−0.08	−0.17		−0.07	−0.24	−0.53	−0.45
E	0.00	0.40	0.10	−0.07	−0.12	0.00	0.45	−0.02	0.04	−0.17	−0.04		0.13	0.22	−0.26
O	−0.01	−0.17	0.20	−0.12	−0.11	0.09	−0.13	0.20	−0.08	−0.09	−0.21	0.14		0.30	0.31
A	−0.10	−0.14	−0.08	0.25	−0.01	−0.03	0.04	−0.09	0.20	0.07	−0.55	0.17	0.26		0.41
C	−0.26	−0.32	−0.27	0.20	0.50	−0.18	−0.25	−0.06	0.05	0.57	−0.47	−0.21	0.28	0.40	

Correlations for females are reported above, and correlations for males are reported below the diagonal.
Correlations beyond $r = 0.18$ are significant ($P < 0.05$).

Table 4. Goodness of fit indices for the multitrait–multimethod analysis with structural equation models

Model		Female *Ss* ($n = 128$)				Male *Ss* ($n = 128$)			
		χ^2	d.f.	ρ	Δ	χ^2	d.f.	ρ	Δ
1A	Null Model	820.73	105	—	—	754.62	105	—	—
1B′	Three orthogonal methods only	621.18	90	0.13	0.24	576.16	90	0.13	0.24
2′A	Five orthogonal traits only	395.47	90	0.50	0.52	—	—	—	—
2′B′	Five orthogonal traits and three orthogonal methods	156.27	75	0.84	0.81	149.88	75	0.84	0.80
2′C	Five orthogonal traits and three oblique methods	150.93	72	0.84	0.82	113.21	72	0.91	0.85
3B′	Five oblique traits and three orthogonal methods	107.62	65	0.90	0.87	104.43	65	0.90	0.86
3C	Five oblique traits and three oblique methods	103.97	62	0.90	0.87	79.92	62	0.95	0.89

In addition to assessing the statistical significance and quantitative fit of one specified model, the difference in fit among hierarchically nested models may also be tested for significance. Hierarchically nested models are more restrictive special cases of a more general model with a higher number of estimated parameters. For instance, factor models with orthogonal trait factors are nested within factor models with oblique trait factors, because the correlations among trait factors are fixed to zero in the orthogonal model whereas they are estimated from the data in the oblique model. The difference in the χ^2 values of two nested models is itself distributed as χ^2 with degrees of freedom equal to the difference in degrees of freedom for the two models. Widaman (1985) suggested a taxonomy of hierarchically nested structural models for the analysis of multitrait–multimethod data. Models without trait factors are nested within models with orthogonal trait factors which are nested within models with oblique trait factors. Correspondingly, models without method factors are nested within models with orthogonal method factors which are nested within models with oblique method factors. Accordingly, the so-called null model without any trait and method factors is most restrictive, whereas models with oblique trait as well as oblique method factors are least restrictive. Crossing the three degrees of restrictiveness for trait and method factors, nine hierarchically nested models may be compared (see Widaman, 1985).

We applied Widaman's procedure to evaluate the two multitrait–multimethod matrices in Table 3. Thus either there were no trait factors specified, or five orthogonal trait factors were specified, or five oblique trait factors were specified. Correspondingly, either there were no method factors specified, or three orthogonal method factors were specified, or three oblique method factors were specified. The nine MTMM-models were separately evaluated for male and female *Ss*. In Table 4, the results are reported for those models that could be identified for at least one of the sexes.

Reasonable indices of statistical and quantitative fit were only obtained for those models that considered trait and method factors. However, it was Model 3C with correlated trait factors and with correlated method factors only that yielded a nonsignificant χ^2 value. Moreover, the differences in overall fit of model 3C (that had five oblique trait factors) and model 2′C (that had five orthogonal trait factors) was also statistically significant, $\chi^2(10, n = 128) = 33.29$, $P < 0.01$ for the male *Ss*, and $\chi^2(10, n = 128) = 46.96$, $P < 0.01$ for the female *Ss*. The best fitting model 3C and the parameter estimates for the male *Ss* are depicted in Fig. 1.

Substantial negative correlations were found between the traits Neuroticism and Conscientiousness, and between Conscientiousness and Openness to Experience. The values for the male *Ss* are reported in Fig. 1. The correlations for the female *Ss* were $r = -0.43$ between Neuroticism and Conscientiousness and $r = -0.21$ between Conscientiousness and Openness to Experience.

Confirmatory factor analyses

The correlations among: (a) the 20 facet scores that are obtained if the NEO scoring key is used, (b) the 20 self-ratings on Norman's adjective scales, and (c) the 20 peer ratings on these scales, were simultaneously analyzed. A structural equation model with five trait factors and three method factors was used. The trait factors were N, E, O, A, and C, and the method factors were Inventory Scales, Self-Ratings on Adjective Scales, and Peer Ratings on Adjective Scales. All loadings of the 60 variables on the eight factors were assumed to be zero, except for the appropriate trait and method factor for which the loadings were estimated. Thus, for instance, the six facets of N in the NEO Personality Inventory were allowed to have nonzero loadings on the trait factor Neuroticism

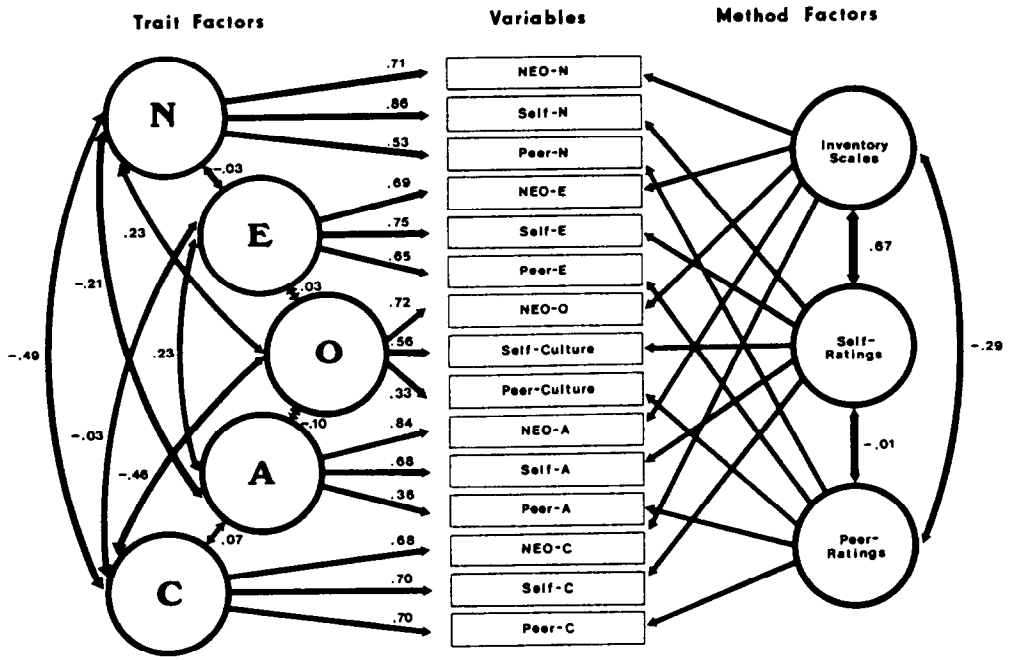

Fig. 1. Multitrait–multimethod model for 15 variables with five oblique trait and three oblique method factors (male *Ss*).

and on the method factor Inventory Scales only. The nine hierarchically nested models were separately evaluated for the female and the male *Ss* to check the replicability of the results and to unconfound personality differences and systematic gender differences.

It was found that even the least restrictive model with three oblique method factors and five oblique trait factors did not appropriately account for the data. The χ^2 statistics suggested rejection of this model for the female *Ss*, χ^2 (1637, $n = 128$) = 3323.08, $P < 0.001$ as well as for the male *Ss*, χ^2 (1637, $n = 128$) = 3140.41, $P < 0.001$. Furthermore, the indices of quantitative fit were low, $\rho = 0.55$, and $\Delta = 0.44$ for the female *Ss*, and $\rho = 0.58$ and $\Delta = 0.46$ for the male *Ss*. They did not even approach the 0.90 criterion suggested by Bentler and Bonett (1980). Consequently, these models had to be rejected.

DISCUSSION

In the present study, the five-factor model is successfully replicated across sexes, across instruments, and across observers. The procedure suggested by Everett demonstrates that, if more or less factors are retained, the factors tend to be less comparable across sexes. Moreover, the five factors that are obtained in the present study correlate highly with conventional measures of the Big Five, such as Costa and McCrae's (1985) NEO-PI and composites of Norman's (1963) marker scales.

It may be argued that this is not surprising because a set of well-established marker variables was used. Indeed, the replicability of the present findings across variables is not demonstrated. However, the 5-factor model and these marker variables were established in numerous earlier exploratory factor analyses of representative sets of variables (see John, Angleitner & Ostendorf, 1988). It was the purpose of the present study to use confirmatory strategies to investigate the robustness of the 5-factor model for this set of marker variables.

Furthermore, multitrait–multimethod analyses of measures of the five factors reveal similar trait factors in personality inventory scales, in self-ratings on adjective scales, and in peer ratings on adjective scales. Thus exploratory factor analyses as well as multitrait–multimethod analyses with structural equation models support a model with five trait factors. Moreover, some agreement between self-reports and observer reports on the position of individuals on the five dimensions is found. This agreement is most pronounced for extraversion and conscientiousness. The present findings have implications for confirmatory factor analysis as well as for the 5-factor model of personality.

Implications for confirmatory factor analysis

We found that a well-established factor model with highly replicable factors and a high convergent and discriminant validity was not supported by a confirmatory factor analysis via structural equation models. The most reasonable explanation of this finding is that confirmatory factor analysis is not so much a logical sequel to exploratory factor analysis, but rather a method that implies different assumptions. First, with the notable exception of Spearman's (1927) general factor model of intelligence, factor-analytic models assume that the correlation between two variables can be accounted for by their respective loadings on *several* common factors. In contrast, models that are submitted to confirmatory factor analysis are usually specified such that each observed variable loads on only one trait factor. This reflects the circumstance that no detailed hypotheses on secondary loadings of variables are available. But the hypothesis of only one non-zero loading is not at all plausible in factor-analytic personality research. The NEO Personality Inventory, for instance, distinguishes among six facets of each, Neuroticism, Extraversion, and Openness to Experience. Norman's (1963) marker scales for the Big Five are no less heterogeneous; the factor Conscientiousness, for instance, subsumes adjectives as diverse as *fussy* and *persevering*. It is not reasonable to suppose that these variables differ in their loadings on the factor Conscientiousness only. Rather, the variables are likely to have somewhat different secondary loadings on other factors in addition to their common loading on the factor Conscientiousness. These secondary loadings, however, are difficult to specify. Accordingly, the simple structure of factor solutions is only approximate in personality research, and it can only be improved at the cost of using near synonyms as marker variables for each factor. But if simple structure is arranged this way, the results of exploratory factor analysis get trivial.

Second, confirmatory factor analysis requires that the bulk of the common variance in a matrix is accounted for by the factor model. In the exploratory factor analyses of the present study, however, the percentages of variance accounted for by the 5-factor model are, for male and female *S*s respectively, 53.2 and 54.8% for the NEO-PI, 51.5 and 46.3% for the self-ratings on adjective scales, and 63.4 and 62.6% for the peer ratings on adjective scales. Obviously, it had been possible to retain a higher number of factors that by definition account for a higher proportion of the variance. But Table 1 demonstrates nicely that, if additional factors are retained, they account mainly for highly specific variance in the data and are not replicated across *S* samples. Thus the 5-factor model of personality limits the extraction of factors to the most robust and meaningful dimensions, instead of attempting to account for as much of the variance as possible within one dataset.

However, even if confirmatory factor analysis turns out to be too restrictive to make it a useful tool for personality research, structural equation models are nevertheless useful for multitrait–multimethod analyses. It is here that the present study sheds some new light on the 5-factor model of personality.

Implications for the 5-factor model

The multitrait–multimethod analyses of comprehensive measures of the Big Five demonstrate that: (a) only models with correlated trait factors appropriately account for the data, and (b) structural equation models can be fitted to the data such that measures of Openness to Experience and measures of Culture constitute the same trait factor. The first finding shows that orthogonality of factors may be an artifact in case of the 5-factor model that is usually imposed on the data by means of Varimax-rotation. Two conclusions may be drawn from this finding. Either one may abandon the orthogonality assumption, or one may attempt to construct new marker variables for the five factors that makes them mutually orthogonal. This, however, implies that the appropriate interpretation of the factors may change to some degree.

The finding that measures of Openness to Experience and of Culture load on the same trait factor, suggests that the 5-factor model of personality is empirically more definite than might be concluded from the various verbal interpretations of the five factors that suggest a lot of ambiguity. Indeed, Fiske (1949), Tupes and Christal (1961), Digman and Takemoto-Chock (1981), Amelang and Borkenau (1982), Costa and McCrae (1985), and Goldberg (1989) suggest somewhat different names for at least some of the factors. The present study shows that for Norman's factor Culture

and for Costa and McCrae's factor Openness to Experience, the different concepts do not result in a corresponding discriminant validity of the measures, at least as far as our German translation of these measures is concerned. Thus the present study replicates the 5-factor model of personality not only across sexes, across instruments, across observers, and across languages, but also across somewhat different conceptualizations of the five major factors of personality.

Acknowledgements—The research reported in this paper was supported by a grant from Deutsche Forschungsgemeinschaft to Peter Borkenau (Az. Bo 774/2-1). We thank Lew Goldberg, Robert McCrae, John Digman, and especially Wim Hofstee and Rainer Riemann for their helpful comments on earlier drafts of this article.

REFERENCES

Amelang, M. & Borkenau, P. (1982). Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität [On the factor structure and external validity of some questionnaire scales measuring dimensions of extraversion and emotional lability]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 3,* 119–146.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 3,* 588–606.

Borkenau, P. & Ostendorf, F. (1990). Untersuchungen zum Fünf-Faktoren Modell der Persönlichkeit und seiner diagnostischen Erfassung [Studies on the five-factor model of personality and its diagnostic assessment]. *Zeitschrift für Differentielle und Diagnostische Psychologie.* In press.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cattell, R. B. (1950). *Factor analysis in the human sciences.* New York: Harper.

Cattell, R. B. (1965). *The scientific analysis of personality.* London: Penguin.

Costa, P. T. & McCrae, R. R. (1985). *The NEO Personality Inventory. Manual Form S and Form R.* Odessa, Florida: Psychological Assessment Resources.

Digman, J. M. & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50,* 116–123.

Digman, J. M. & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Reanalysis and comparison of six major studies. *Multivariate Behavioral Research, 16,* 149–170.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18,* 197–218.

Eysenck, H. J. & Eysenck, S. B. G. (1969). *Personality structure and measurement.* London: Routledge.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44,* 329–344.

Goldberg, L. R. (1989). An alternative "description of personality": The Big-Five factor structure. Unpublished manuscript: Oregon Research Institute.

Guilford, J. P. (1975). Factors and factors of personality. *Psychological Bulletin, 82,* 802–814.

John, O. P., Angleitner, A. & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2,* 171–203.

Jöreskog, K. G. & Sörbom, D. (1984). *LISREL VI.* Chicago: National Educational Resources.

Judd, C. M., Jessor, R. & Donovan, J. E. (1986). Structural equation models and personality research. *Journal of Personality, 54,* 149–198.

McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

McCrae, R. R., Costa, P. T. & Busch, C. M. (1986). Evaluating comprehensiveness in personality systems: The California Q-set and the five-factor model. *Journal of Personality, 54,* 430–446.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22,* 267–305.

Noller, P., Law, H. & Comrey, A. L. (1987). Cattell, Comrey, and Eysenck personality factors compared: More evidence for the five robust factors? *Journal of Personality and Social Psychology, 53,* 775–782.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes. Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574–583.

Peabody, D. & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57,* 552–567.

Schmitt, N. & Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement, 10,* 1–22.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrica, 38,* 1–10.

Tupes, E. C. & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF Technical Report ASD-TR-61-97.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement, 9,* 1–26.

Zuckerman, M., Kuhlman, D. M. & Camac, C. (1988). What lies beyond E and N? Factor analyses of scales believed to measure basic dimensions of personality. *Journal of Personality and Social Psychology, 54,* 96–107.