4/ On the generality and comprehensiveness of the Five-Factor model of personality

Evidence for five robust factors in questionnaire data

Fritz Ostendorf
University of Koblenz-Landau, Germany

Alois Angleitner
University of Bielefeld, Germany

Introduction

In recent years, the development of personality trait taxonomies has led to increasing research on the most reliable and most important personality factors. As a result, a 'new' model of personality structure (but see Wiggins and Trapnell, in press) has joined the old debate between the well-known but vehemently competing factor models proposed by Cattell, Eysenck, Guilford, and others. Assuming five factors, which are believed to provide a sufficient description of personality structure, the new model takes an intermediate position between the more simple three-factor PEN model proposed by Eysenck and the more complex multifactorial systems proposed by Cattell and Guilford. Because the Five-Factor model has been shown to be robust across a diversity of studies, the five factors have also been called the Big Five. With reference to the work of Norman (1963, 1967), Goldberg (1980), and McCrae and Costa (1985a,b,c), the Big Five are frequently labelled: (1) *surgency* or *extraversion*; (2) *agreeableness*; (3) *conscientiousness*; (4) *emotional stability* or, conversely, *neuroticism*; and (5) *culture, intellect*, or *openness to experience*.

Although different researchers have varied in their choice of labels, there are substantial meaning overlaps as well as strong empirical relationships between the different operationalizations of the five factors. To elucidate their meaning, Table 4.1 presents some marker variables of the Big Five selected from adjective rating inventories published by Norman (1963), McCrae and Costa (1987), and Peabody and Goldberg (1989). In addition, Table 4.1 shows the loadings of the adjective rating scales on five Varimax rotated factors that emerged from principal component

Table 4.1 Big-Five marker scales from various rating inventories and their factor loadings in analysis of self- and peer-rating data

	SU/EX		AG		CO		ES		OP/IN	
	I	II	I	II	I	II	I	II	I	II
Factor I (surgency or extraversion)										
Talkative–silent	**78**	**71**	−05	03	02	06	−00	−06	09	05
Sociable–reclusive	**79**	**76**	09	17	02	−01	18	04	02	04
Fun-loving–sober	**61**	**53**	19	22	−24	**−30**	06	−06	18	27
Spontaneous–inhibited	**73**	**73**	02	−01	−09	−11	26	15	14	24
Frank–secretive	**76**	**65**	23	**33**	−03	−01	03	06	14	16
Active–inactive	**64**	**58**	−00	−05	**38**	**30**	19	16	17	27
Factor II (agreeableness)										
Mild, gentle–headstrong	−03	01	**69**	**68**	−05	−02	11	20	14	23
Good-natured–irritable	−13	−05	**58**	**57**	−02	12	**43**	**38**	07	09
Soft-hearted–ruthless	12	13	**71**	**68**	07	19	−13	−09	15	19
Forgiving–vengeful	25	18	**64**	**67**	10	24	17	25	07	14
Trustful–distrustful	28	25	**55**	**56**	−02	04	07	16	−02	−04
Lenient–harsh	04	−01	**68**	**69**	01	08	13	**30**	22	13
Factor III (conscientiousness)										
Responsible–undependable	15	12	17	24	**68**	**73**	−04	03	25	18
Scrupulous–unscrupulous	−15	−13	−04	−04	**68**	**65**	−03	−11	−03	11
Conscientious–negligent	13	14	07	18	**80**	**77**	07	04	07	12
Hardworking–lazy	16	20	11	17	**67**	**69**	07	04	06	06
Serious–frivolous	−25	−23	10	10	**64**	**70**	−01	09	09	14
Orderly–disorderly	04	05	07	14	**76**	**76**	18	16	−09	−01
Factor IV (emotional stability)										
Poised–nervous, tense	11	06	24	**30**	04	02	**75**	**73**	05	09
Calm–anxious	21	09	03	09	11	17	**64**	**70**	19	13
Hardy–vulnerable	12	16	−07	−03	15	10	**64**	**59**	−12	−20
Calm–worrying	11	−08	12	11	04	−03	**64**	**61**	06	02
Relaxed–tense	08	06	24	**31**	−02	−08	**70**	**63**	02	11
Contented–discontented	**39**	26	26	**32**	17	21	**55**	**54**	08	06
Factor V (culture, intellect, or openness to experience)										
Artistically sensitive–artistically insensitive	−04	05	09	15	−14	−04	−08	−06	**57**	**60**
Intellectual–unreflective, narrow	−07	−02	03	04	25	**36**	09	10	**66**	**69**
Creative–uncreative	12	19	06	14	04	04	09	02	**60**	**64**
Broad interests–narrow interests	18	**32**	09	16	18	**30**	21	13	**58**	**63**
Intelligent–unintelligent	08	16	01	07	20	**31**	09	13	**68**	**67**
Imaginative–unimaginative	27	**31**	10	12	00	05	04	05	**63**	**67**

Notes All loadings ≥0.30 are listed in **bold**. The first two adjective rating scales listed under the heading of each factor are from Norman (1963), the next two scales from McCrae and Costa (1987), and the last two from Peabody and Goldberg (1989). The factor loadings reported in the table are loadings on five Varimax rotated principal components based on analyses of 179 rating scales (see Ostendorf, 1990) published by Norman (1963), McCrae and Costa (1987), Peabody and Goldberg (1989), Goldberg (1983, 1989), and John (1983; see John *et al.*, 1984). I: peer-ratings (N = 383), II: self-ratings (N = 401). SU/EX = surgency or extraversion, AG = agreeableness, CO = conscientiousness, ES = emotional stability, OP/IN = openness or intellect.

analyses of all 179 rating scales contained in the rating inventories from Norman (1963), McCrae and Costa (1987), Peabody and Goldberg (1989), Goldberg (1983, 1989), and John (1983; see John *et al.*, 1984). Although the adjectives listed in Table 4.1 were not selected according to their factor loadings, the patterns there are strikingly congruent and clear. Details of both analyses, which were based on samples of 383 (peer-ratings) and 401 subjects (self-ratings) are reported in Ostendorf (1990).

Most previous confirmations of the robustness of these five factors have come from studies of adjective rating data. For example, the validity of the Five-Factor model has been demonstrated on the basis of representative samples of personality-descriptive terms in various languages: Goldberg (1990) for Anglo-American; De Raad *et al.* (1988) for Dutch; and Ostendorf (1990) for German. Similar variants of the Big Five were found in studies of non-western personality languages; for example, Chinese (Yang and Bond, 1990), Filipino (Church and Katigbak, 1989), and Japanese (Isaka, 1990). These and other studies have shown that the five factors were not always the only factors that could underlie a specific data set. However, they were the only robust factors that could be replicated reliably across different languages, adjective samples, groups of raters, rating formats, and variations in the method of factor analysis.

Because it is quite likely that the range of all possible personality descriptions based on trait adjectives is not equivalent to that based on personality-descriptive sentences, it seems advisable to test the structural validity of the Five-Factor model on other data sources as well; for example, on the basis of questionnaire scales or items. If these factors are universal factors of personality language, they should be revealed not only on the basis of personality-descriptive adjectives in rating data but also on the basis of personality-descriptive phrases in questionnaire data.

McCrae and Costa explored this issue in a series of studies. For each separate study, they selected one popular personality inventory that was representative of one other major personality theory (for example, the Personality Research Form-E: Costa and McCrae, 1988; the California Q-Sort: McCrae *et al.*, 1986; the Myers-Briggs Type Indicator: McCrae and Costa, 1989a; the Adjective Check List: Piedmont *et al.*, 1991; the EPI and Psychoticism scales: McCrae and Costa, 1985b). To evaluate the comprehensiveness of the Five-Factor model, the scales of each personality inventory were correlated with the scales of the NEO Personality Inventory (NEO-PI), a questionnaire explicitly constructed for measuring the Big Five. In most studies, the common structure of both inventories was subsequently analysed in one common factor analysis. However, a more suitable test of the Five-Factor structure in questionnaire data would require the inclusion of a more comprehensive sample of questionnaires, such as a simultaneous factor analysis of scales or items from a very broad spectrum of personality questionnaires.

A basic principle underlying the major studies on the Five-Factor model in rating data has been the *comprehensiveness* and *representativeness* of the item pool analysed: proceeding from the assumption that most important individual differences are already encoded in everyday language, one first extracts a comprehensive and reasonably representative sample of the personality-descriptive terms from this language. Then,

the most important factors describing personality can be determined on the basis of this representative sample of terms. Ideally, representative samples of personality-descriptive adjectives, verbs, or nouns are drawn from dictionaries containing the complete vocabulary of the language in question.

Unfortunately, no comparable sampling procedure is available for compiling personality-descriptive phrases such as questionnaire items. In principle, it is possible to construct an infinitely large number of personality-descriptive phrases. So, the number of possible questionnaire items is, at least theoretically, infinite. Previous attempts to test the validity of the Five-Factor model on the basis of questionnaire data have thus restricted their studies to a specific selection of items or scales from well-known personality questionnaires. Consequently, such a selection of questionnaire items does not represent the population of all possible personality-descriptive phrases but instead reflects the specific focuses and preferences of different researchers.

None the less, a relatively convincing solution could be provided by simultaneously analysing the items from a large set of personality questionnaires, even if representativeness cannot be ultimately achieved. With regard to the validity of the Five-Factor model, some such studies have already been performed with varying success.

In the following, we want to review a selection of such studies. Our selection is limited to studies in which a large sample of items or scales from several inventories have been factored in a common analysis. First, we will report results from factor analyses of questionnaire *scales*, which have been interpreted within the conceptual framework of the Five-Factor model either by the authors of the study in question or by other researchers. The next section explores how far the results of former extensive *item*-factor analyses have shown evidence for the Big-Five factors. Finally, we report the results of our own study, in which we examined the structural validity of the Five-Factor model on the basis of 576 items from inventories of different prominent personality theories.

Questionnaire scales and the Big Five

An overview of the many factor analytic studies of personality questionnaires is beyond the scope of this chapter. The only studies that are relevant to the structural validity of the Five-Factor model are those in which a large and comprehensive number of scales, preferably stemming from multidimensional personality inventories, have been analysed. We have also chosen a selection from this set of studies: in the following, we will report only studies whose results were originally discussed within the framework of the Five-Factor model. Table 4.2 presents a broad view of the results of such studies.

The results of a study by Amelang and Borkenau (1982) may lead to the assumption that the five rating factors can also be replicated in the domain of questionnaire data as long as a sufficiently large number of questionnaire scales from different inventories

Table 4.2 Classification of factors derived from comprehensive factor analyses of questionnaire scales

	Extraversion (I)	Agreeableness (II)	Conscientiousness (III)	Neuroticism (IV)	Openness (V)
Amelang and Borkenau (1982)	Extraversion	Dominance^{R}	Self-control	Neuroticism	Independence of opinion
Noller *et al.* (1987)	Extraversion	Agreeableness	Conscientiousness	Neuroticism	—
Boyle (1989)	Extraversion	Tough poise^{R}	Control	Neuroticism	Independence
Montag and Comrey (1990)	Extraversion	Agreeableness	Conscientiousness	Neuroticism	—
Conn and Ramanaiah (1990)	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness to experience
Matthews *et al.* (1990)	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness to experience
Zuckerman *et al.* (1988)	Sociability	Aggressive Sensation-seeking^{R}	Impulsive Unsocialized Sensation-seeking^{R}	Neuroticism Emotionality	—
Zuckerman *et al.* (1991)	Sociability	Aggression Hostility^{R}	P-Impulsive Unsocialized Sensation-seeking^{R}	Neuroticism Anxiety	—

Note: The superscript ^{R} means that a factor is reverse-scored in the direction opposite to that of the Big-Five label listed in the column head.

are entered into a common factor analysis. In this German study, the scales of Cattell's Sixteen Personality Factor Questionnaire (16PF; Schneewind, 1977), the Freiburg Personality Inventory (FPI; Fahrenberg and Selg, 1970), the Eysenck Personality Inventory in its two parallel forms (Eggert, 1974; Eysenck and Eysenck, 1968), and the scales of several inventories constructed by Guilford and his co-workers (GAMIN, STDCR, and GMPI, cited in Guilford *et al.*, 1976) were factor-analysed together. The extracted factors were interpreted as neuroticism, extraversion, dominance (probably similar to non-agreeableness), independence of opinion (culture/openness), and self-control (conscientiousness). Advocates of the Five-Factor model (such as Digman, 1990; McCrae and Costa, 1985b) frequently cite the results of this study as evidence for the generality and robustness of the Big Five. Other comparisons, using the factor models of Cattell, Guilford, and Eysenck, have led to outcomes that cannot be interpreted so clearly as corresponding to the Five-Factor model.

Noller, Law, and Comrey (1987) explored the common factor structure of the scales of the following inventories: the Comrey Personality Scales (CPS; Comrey, 1970), the 16PF scales (Cattell *et al.*, 1970), and the EPI scales (Eysenck and Eysenck, 1968). The factor analysis included 26 personality scales, five validity and response distortion scales, and scores for gender and age. Seven factors were extracted. According to the authors' interpretations, the most important part of the common variance of the questionnaire scales could be explained by four of the five

Norman factors. A factor identifiable as culture or openness could not be replicated. The evidence for the agreeableness factor was only weak. Costa and McCrae (1976), for example, interpreted a similar 'agreeableness' factor as openness to experience.

A quite similar picture resulted from a reanalysis of Noller *et al.*'s data conducted by Boyle (1989). Five factors resulted from a factor analysis of 25 EPI, CPS, and 16PF personality variables. In accordance with the 16PF secondaries reported by Krug and Johns (1986), Boyle (1989) interpreted the 'Big-Five' factors as extraversion, neuroticism, tough poise, independence, and control. Whereas Boyle (1989) did not discuss the possible correspondence of these factors to the Big Five, John (1990), starting from an inspection of the items of the 16PF, recently postulated a high equivalence between Cattell's second-order factors and the Big Five. We see a relatively high similarity between three factors reported by Boyle and the following three Big-Five factors: extraversion, neuroticism, and conscientiousness (Boyle's control factor). However, correspondence of the remaining two factors to the Big Five cannot be assumed without further empirical proof.

Montag and Comrey (1990) factored a subset of Noller *et al.*'s (1987) questionnaire scales in a different sample. In a common factor analysis of the 16PF and CPS scales, the authors clearly identified three of the Big-Five factors (extraversion, conscientiousness, and emotional stability). Some variants of agreeableness and openness were also identified, although not so clearly.

Conn and Ramanaiah (1990) found three factors underlying the Comrey Personality Scales (CPS) and five factors underlying the scales of the Personality Research Form-E (PRF; Jackson, 1984). The three Comrey factors were seen as similar to agreeableness, conscientiousness, and extraversion, and the five PRF factors were interpreted as being similar to those in the Five-Factor model. A combined factor analysis of the CPS and PRF factors yielded five second-order factors that were interpreted as extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. In our view, these results demonstrate that the CPS scales, and especially the PRF scales, assess a large proportion of the content of the Big Five. Besides obvious correspondences, there are also specific discrepancies. The CPS-extraversion factor, extracted by Conn and Ramanaiah (1990), appears to be a fusion of extraversion and neuroticism. However, no marker scales measuring the Big Five were included in the analysis, and, therefore, the concept of similarity used by Conn and Ramanaiah (1990) may be questioned.

Matthews, Stanton, Graham, and Brimelow (1990) analysed the structure of the Occupational Personality Questionnaire (OPQ; Saville & Holdsworth, 1984), which is based on a conceptual model claiming to provide a comprehensive coverage of personality. The authors conducted a factor analysis of the 31 OPQ scales, and the resulting pattern matrix of factor loadings showed a closer correspondence with the Big-Five dimensions of personality than with the structural model assumed by the authors of the OPQ. The similarity to the Big Five was impressive, although, as this interpretation is based on the inspection of factor loadings only, it needs further empirical proof.

Zuckerman, Kuhlman, and Camac (1988) factored a set of 46 questionnaire scales

mostly measuring the constructs of biologically oriented temperament theories.[1] As can be expected from such a specific selection of variables, the following factors were clearly identifiable: sociability (or extraversion) and emotionality (neuroticism). Another factor was called impulsive-unsocialized sensation seeking. This factor was mostly marked by scales that included not only aspects of openness but also features of conscientiousness. A fourth factor called aggressive sensation seeking primarily loaded through the scales aggression (PRF) and responsibility (Jackson Personality Inventory; Jackson, 1976). This factor might possibly represent the negative pole of agreeableness.

Quite similar factors were found by Zuckerman, Kuhlman, Thornquist, and Kiers (1991) in an analysis of a selection of 33 of the 46 scales originally used by Zuckerman *et al.* (1988). The authors, however, emphasized that their interpretations required further empirical evidence. Unfortunately, the studies of Zuckerman *et al.* (1988, 1991) only analysed a highly selective sample of scales from several more or less broad personality inventories. Scales that can be considered as relatively clear markers for the factors agreeableness (such as PRF-nurturance), conscientiousness (such as the PRF scales order, achievement, endurance, impulsivity), and openness (such as PRF-understanding) were not included.

In summary, at least variants of the Big Five have been found in the majority of studies. Nevertheless, there were some clear deviations from the postulated factor structure in single cases. As one could expect, the factors extraversion and neuroticism were the most robust. With only a few exceptions, these factors could be replicated reliably in all studies. It was nearly always these factors that explained the largest part of the variance.[2] Third in order of stability was conscientiousness. This factor could be replicated particularly well in factor analyses that included the scales of the 16PF, the CPS, or the PRF. Primary marker variables for conscientiousness were the CPS-scale orderliness, and the 16PF-scales superego strength, and self-control, as well as the PRF-scales order, cognitive structure, achievement, and non-impulsivity. There was less empirical evidence for the hypothesized factor agreeableness and only little evidence for an independent and clearly interpretable fifth factor. Of course, this pattern of results is most probably due to the idiosyncratic variable selections used in the different studies. The factor openness, for instance, was only represented in a few studies through a small number of questionnaire scales (such as the 16PF-M-scale imagination, the Q1-scale radicalism, and the PRF-scale understanding). Indeed, the results of the studies reflect the fact that most personality test constructors are interested in measuring at least the (most important?) personality factors extraversion and neuroticism.

Furthermore, a common characteristic of all reported studies was that assumed correspondence between questionnaire factors and the Big-Five rating factors were always inferred by subjective interpretations but not tested via empirical analyses. Most studies interpreted the results of explorative factor analyses solely on the basis of visual inspections of factor-loading patterns. Furthermore, all former studies analysed questionnaire data on the scale level. For several reasons, we prefer to factor-analyse questionnaire data on the item level.

Item-factor analysis as an appropriate method for analysing the structure of questionnaire data

It is well-known that different personality researchers have a preference for constructs of varying breadth, abstraction, or globality. Broad, abstract, or global constructs are defined in such a way that they refer to a larger number of behaviours than do narrow ones. In factor analyses, variations in the breadth of the measured traits could be represented adequately by using large numbers of scales or items for broad traits and only one or a few for narrow traits. In fact, broad constructs are usually operationalized through a larger number of items (for example, the NEO-extraversion scale contains 48 items). In factor analyses of scales, however, these items are summed up to only one single-scale score, with the result that broad scales unwarrantedly obtain the same weight as narrow ones. That is, the different breadths of the constructs are concealed. The resulting factors represent a mixture of different levels of abstraction. In extreme cases, a new factor appears because a very specific construct has been operationalized via several narrow scales, each composed of a small number of highly similar items. If, in contrast, the complete item pool is factored, the content of all the items can be represented more adequately. In an item-factor analysis, the contents will be represented most adequately if specific constructs are operationalized through fewer items than broad constructs. This tends to be the case, because global scales frequently comprise more items than narrow ones.

As well as accounting for a broader spectrum of contents at the item level, the method of item-factor analysis is more in line with the goal of the lexical approach: a comprehensive description of the structure of personality language should include as many different personality-descriptive terms or statements as possible.

Finally, for the purpose of assigning items to scales, a factor analysis of a large number of questionnaire items may lead to item weights that are much more valid than those given by the item keys of many commonly used personality inventories. This comparison also applies to factor-analytically derived questionnaire scales if these scales have been developed on the basis of only small homogeneous item pools. It is well known that the correlation between items and scales or factors heavily depends on the kind and size of the item pool. A collection of items that appears to be quite homogeneous in a specific item sample may turn out to be much more heterogeneous in a broader item sample.

A related argument derives from the observation that factor analyses of questionnaire scales are often difficult to interpret because the labels of the scales do not refer accurately to the item contents: similar labels refer to different contents; scales labelled differently measure similar domains. Therefore, in most scale-factor analyses, the interpretation of the factors cannot be based solely upon the factor pattern of the scales but must be grasped and guided indirectly through the interpretation of items. In factor analyses of items, this problem does not exist: the factors are directly interpretable according to their item loadings.

Actually there are a number of statistical problems associated with the application of item-factor analysis (see Nunnally, 1978). None the less, some of the leading

experts in factor analysis are convinced that the advantages of item-factor analysis outweigh its disadvantages. So, one can follow Cattell and Gibbons's (1968) recommendation that even the parcel-factor analysis appears to be an unsatisfactory compromise in comparison with complete item-factor analyses: 'any attempt to decide between the factor structure of two scales had best . . . break down the scales under examination ideally into single items or, if economy forbids this, into a fair number . . . [of] random "parcels" from each scale' (p. 118).

Studies based on comprehensive samples of questionnaire items

To the best of our knowledge, no study has been published in which the Five-Factor model was tested with a *comprehensive* selection of questionnaire items. Most factor-analytic studies of questionnaire items have been based on samples of items from the three prominent personality theories of Cattell, Guilford, and Eysenck. From the countless number of studies, we will review only those that have performed a factor analysis of large item samples taken from several personality questionnaires. Table 4.3 gives an overview of these studies.

Cattell and Gibbons (1968) compared the Guilford and Cattell factor markers in

Table 4.3 Classification of factors derived from comprehensive factor analyses of questionnaire items

	Extraversion (I)	Agreeableness (II)	Conscientiousness (III)	Neuroticism (IV)	Openness (V)
Cattell and Gibbons (1968)[a]	Exvia	Pathemia vs. cortertia	—	Anxiety	Independence
Sells *et al.* (1970)	Extraversion	Relaxed composure vs. suspicious excitability	Conscientiousness	Emotional stability^{R}	Artistic interests
Eysenck and Eysenck (1969)	Extraversion	—	—	Neuroticism	—
Vagg and Hammond (1976)	Sociability	Community-centred morality vs. self-centred independence		Neuroticism	Sensitivity vs. practicality
Eysenck (1978)	Extraversion	—	—	Neuroticism	—
McKenzie (1988)	Extraversion	—	Superego strength	Neuroticism	—
Johnson *et al.* (1984)	Surgency	Agreeableness	Conscientiousness	Emotional stability^{R}	Culture
Costa *et al.* (1985)	Extraversion	Cynicism^{R}	—	Neuroticism	Intellectual interests

Notes: The superscript ^{R} means that a factor is reverse-scored in the direction opposite to that of the Big-Five label listed in the column head. The factors found by Browne and Howarth (1977) are not reported in Table 4.2 (see text).

[a]Cattell and Gibbons (1968) performed an item-parcel factor analysis.

a common factor analysis. The data, sampled by Gibbons (1966) in a doctoral dissertation under Guilford, involved 424 items representing 14 Cattell factors and 15 Guilford factors. The authors conducted a parcel-factor analysis based on 68 variables representing the 424 items; that is, the 68 variables were constructed out of clusters of items with homogeneous content. Eighteen primary-order factors were extracted and rotated to an oblique factor pattern. Cattell and Gibbons (1968) concluded that all 14 Cattell factors were confirmed, and that the 14 Cattell factors and 15 Guilford factors had eight dimensions in common. However, we are more interested in the second-order factors found in these studies, because the primaries have been found to be highly unreplicable (for example, Eysenck and Eysenck, 1969; Howarth and Browne, 1971a,b; Kline and Barrett, 1983). Nine second-order factors were rotated. In agreement with former studies, the four factors explaining most variance were labelled: (1) exvia vs. invia, (2) anxiety, (3) cortertia, and (4) independence. As John (1990) has recently argued, Cattell's second-order factors show a high correspondence with the Big Five. According to John (1990, p. 88), there is a high correspondence between the factors extraversion and exvia, neuroticism and anxiety, agreeableness and pathemia vs. cortertia, openness to experience and independence, as well as between conscientiousness and superego strength. In line with John's interpretation, the four second-order factors in Cattell and Gibbons (1968) can be considered as equivalent to four of the Big Five. However, Cattell and Gibbons did not report a factor superego strength. Furthermore, the loading patterns of Cattell and Gibbons's second-order factors reveal some marked deviations from the results of other studies reported by Cattell's team (see Cattell and Gibbons, 1968, pp. 115–16). For example, the factor exvia–invia not only correlated with the typical scales for this factor but also showed substantial loadings of G (superego strength), O (untroubled adequacy vs. guilt proneness), and Q1 (conservatism vs. radicalism). Further pecularities were found in the loading patterns of the other factors. Therefore, neither a high stability of the second-order factors nor a perfect correspondence with the Big Five can be assumed. Consequently, the precise level of correspondence with the Big Five needs to be tested empirically through the inclusion of appropriate marker variables.

Sells, Demaree, and Will (1970) conducted an item-factor analysis comparing the personality concepts of Cattell's and Guilford's models. This study is one of the largest item-factor analyses ever reported. A total of 600 items was factorized; 300 marker items comprised 78 marker clusters for 15 Guilford factors, and 300 items represented marker items for 17 Cattell factors. Twenty-three factors were extracted from the matrix of item correlations, and both 15 and 18 factors were rotated using both the Varimax and Promax methods. The 18 Promax factors were interpreted, but neither Guilford's nor Cattell's model was confirmed. The two most dominant factors were clearly identified as *(1) emotional stability* (primarily marked by items from the Guilford [G] and Cattell [C] scales cycloid disposition [G:C], nervousness [G:N], depression [G:D], guilt proneness [C:O], ergic tension [C:Q4], and as *(2) extraversion* (which was primarily loaded by items of the scales sociability [G:S], Parmia [C:H: venturesome, bold], surgency [C:F], and ascendance [G:A]). In terms

of explained variance, the next most importance factors were *(3) artistic interests* (this factor corresponds almost completely with Guilford's Artistic Interest Scale), *(4) conscientiousness* (which shows loadings from cultural conformity [G:CC] and superego strength [C:G]) and *(5) relaxed composure vs. suspicious excitability* (primarily loaded by excitability [C:D], protension [C:L: suspicious] vs. alaxia [trusting, adaptable], agreeableness [G:A], and self-control [C:Q3].

It is probable that at least the three most dominant factors (emotional stability, extraversion, conscientiousness) are very similar to Big-Five factors, as well as to Cattell's postulated second-order factors (exvia–invia, anxiety, superego strength). Usually, artistic interests is considered to be an important facet of the fifth factor in the Five-Factor model (Costa and McCrae, 1985). For this reason, one would expect a close correspondence — although no identity — between these factors. Likewise the factor relaxed composure vs. suspicious excitability should correspond to agreeableness. The Cattell factors excitability (for example, the uncontrolled expression of anger) and suspiciousness vs. trust can be interpreted as facets of the broader factor agreeableness.

However, our speculations are just as empirically unfounded as the stability of the factors reported by Sells *et al.* (1970). For instance, the decision to extract 18 factors was made rather arbitrarily. Even using marker variables, Howarth and Browne (1971a) were able to replicate only some of the factors reported by Sells *et al.* (1970). It has often been shown that large numbers of factors are usually required to explain a reasonable proportion of the variables' common variance in a specific sample. In general, however, only those factors explaining most of the variance are found to be replicable over different studies. These factors are mostly similar to those that can be found at a higher-order level (McCormick *et al.*, 1991). A reanalysis of the data in Sells *et al.* (1970) could perhaps reveal more evidence for the Big Five by considering only the most dominant and robust factors or by extracting only higher-order factors.

An interesting detail of this study, which supports the utility of item-factor analysis, was the finding 'that analysis at the item level is highly destructive to the factors previously assembled without adequate concern for their loadings in large matrices' (Sells *et al.*, 1970, p. 419). For example, Sells *et al.*'s analysis of Cattell's and Guilford's scales indicated 'a clear need for reclassification of at least 400 of the 600 source items' (p. 421).

A further extensive analysis is reported by Eysenck and Eysenck (1969). They factored a selection of marker items from the following personality inventories: STDCR and GAMIN (Guilford), 16PF (Cattell), and EPI (Eysenck). In this study, Guilford's model was represented by 109 items, Cattell's model by 99 items, and Eysenck's model by the 114 items from the EPI. All items were administered to 600 male and 600 female students. First-, second-, and third-order Promax factor solutions were calculated separately for each inventory, and only factors from a higher level of analysis were carried over into a common factor analysis. As Vagg and Hammond (1976, p. 122) noted, 'only three factors, at most, were carried into common factoring, and since it was to be expected that two of these would be E

and N, there was little chance for another factor to appear'. Unfortunately, Eysenck and Eysenck (1969) did not describe their analysis procedure in sufficient detail. None the less, the higher-order-level analyses yielded only two robust factors. One additional factor found in the male sample and two further factors in the female sample were relatively unimportant and difficult to interpret. The primary goal of Eysenck and Eysenck's (1969) study seems to have been to confirm the factors neuroticism and extraversion. This may be underlined with a citation from Eysenck and Eysenck (1985), regretting 'that at the time these analyses were done . . . no P scale was in existence; hence the analysis is concerned only with two factors. It would be interesting to repeat the work . . . and to attempt to find a third factor representing psychoticism' (p. 137). Kline and Barrett (1983) have also pointed out procedural problems in the above-mentioned study.

Vagg and Hammond (1976) designed their study as a partial replication of Eysenck and Eysenck (1969). However, they employed methods that gave smaller factors a chance to emerge. That is, Vagg and Hammond (1976) based their analysis on correlations between a larger number of primary factors extracted from each of the three questionnaires, whereas Eysenck and Eysenck's (1969) inter-inventory analysis relied on the correlations of a small number of higher-order factors from each questionnaire. As in the original study, the matrix of item correlations of each inventory was factored for the arbitrary number of 20 principal components, and each of these solutions was subsequently transformed by a Promax rotation. To arrive at a second-order solution, the authors did not calculate factor scores, but obtained scale scores for each person on the 60 primary factors by giving unit weight to each item assigned to a factor. Ten second-order factors were extracted from the correlation matrix of the 60 scales, and a four-factor Varimax solution was found to be robust across gender groups. The largest and most stable factors could be identified clearly as neuroticism and sociability or social extraversion. Factor III was called sensitivity vs. practicality, and was primarily loaded by items from the Cattell scales M+ (praxernia; imagination), I+ (harria; tender-mindedness), A− (sizothymia), and the Guilford scale T (introspectiveness). The authors speculated that this factor might also be related to the artistic interests factor of Sells *et al.* (1970). A comparison reveals that Factor III has much in common with the factor called *Unabhängigkeit der Meinungsbildung* (independence of opinion) found by Amelang and Borkenau (1982) in their common-factor analysis of the Cattell, Eysenck, and Guilford scales. McCrae and Costa (1985b), for example, viewed this factor (*Unabhängigkeit der Meinungsbildung*) as evidence for the robustness of their factor openness to experience. An inspection of the factor loadings and the item contents reported by Vagg and Hammond (1976, p. 126) reveals that this factor covers all important aspects of Costa and McCrae's (1985) factor openness to experience: openness to ideas, to aesthetics, to feelings, values, actions, and openness to fantasy. Factor IV was called community-centred morality vs. self-centred independence, and was seen as equivalent to the conscientiousness factor found by Sells *et al.* (1970) and as similar to Cattell's conception of Promethean will. In addition, the authors expected significant relations to a factor called friendliness, which was found by Bendig (1962)

in a factor analysis of the Guilford–Zimmerman Temperament Survey. Therefore, community-centred morality vs. self-centred independence may be related to both the conscientiousness and the agreeableness factor of the Five-Factor model. However, further studies are needed to test these hypotheses empirically.

Because the studies of Eysenck and Eysenck (1969) and Sells *et al.* (1970) intended to provide a direct comparison of the personality systems of Cattell, Guilford, and Eysenck, their results only apply to these factor models. In contrast, Browne and Howarth (1977) started — similarly to the classic work of French (1953, 1973) — with a comprehensive sampling of scales from several personality theories.

In summary, Browne and Howarth (1977) sampled 3,029 items from 17 personality inventories. After eliminating repeated items, the pool amounted to 1,726 items. The authors selected 400 items on the basis of 20 putative factor hypotheses (PFHs). These PFHs were determined as being representative of the factor structure indicated by previous item-factor studies as well as a review of the literature.[3] Twenty factors were extracted from the correlation matrix of the 400 personality inventory items, rotated to an oblique criterion, and interpreted according to their psychological content. The authors concluded that they had recovered some 15 of the 20 PFHs with which they entered their study. We consider that the contents of these factors are mostly related to the domains of neuroticism and extraversion, and only a few may be related to conscientiousness and agreeableness. Aspects of openness to experience are probably only represented by one factor, called optimal arousal, which combines the PFHs sensation seeking (need for external arousal) and thoughtfulness (need for internal, intellectual arousal). The latter interpretation is in accordance with results reported by McCrae and Costa (in press) who found that some of the aspects of sensation seeking (Zuckerman, 1979) are highly related to openness (see also Angleitner and Ostendorf, 1991).

Eysenck (1978) used the factor inter-correlations reported by Browne and Howarth (1977) to perform a second-order factor analysis. The resulting factors were identified by Eysenck (1978) as neuroticism, extraversion, and psychoticism. The best fit resulted for N and E. Eysenck (1978) attributed the poorer fit for P to the fact that 'traditional inventories have always been preoccupied with N and E variables, and have not paid much attention to Psychoticism' (p. 478). The persuasiveness of Eysenck's results, however, is reduced by the fact that he used a target rotation procedure that probably led to a forced adjustment of the factors from Browne and Howarth (1977) to the Eysenckian model. One indication for this is the fact that the target-rotated factors correlated quite substantially (for example, E and P: −0.32). However, an inspection of the comprehensive sampling of personality inventory items by Browne and Howarth (1977, pp. 417–25) shows that one of Eysenck's arguments is highly convincing: most personality inventories primarily pick out extraversion and neuroticism as a central theme. As long as other personality traits are measured only occasionally, they have only a reduced chance of appearing as independent and significant personality factors.

In a more recent study, McKenzie (1988) subjected the items of the 16PF and the EPQ to a common-factor analysis. He concluded that the Eysenckian and Cattelian

personality model is based on three common, robust factors. But, contrary to the result that the reader might expect, McKenzie (1988) found that Cattell's second-order factor superego, rather than Eysenck's psychoticism, 'may best lay claim to join neuroticism and extraversion in what may be termed the great triumvirate of the personality sphere' (p. 850). Superego is primarily defined by items from the 16PF scales superego strength (G) and self-control (Q3), and this pattern is highly compatible with an interpretation of this factor as conscientiousness. The P-dimension is only partly related to the factor superego; a finding that is in accordance with empirical results from Goldberg (1991) and McCrae and Costa (1985b), who found that psychoticism — as a measure of both Factor 2 (agreeableness) and 3 (conscientiousness) — is related more strongly to agreeableness.

Despite the psychopathological emphasis of the Minnesota Multiphasic Personality Inventory (MMPI), it is also frequently viewed as a comprehensive personality inventory. The total pool of 556 MMPI-items may be large enough to measure a broad spectrum of personality characteristics. For example, the fact that the item pool has been used to develop more than 500 sub-scales used in research and clinical applications may lead to the conviction that one could measure almost everything with the MMPI item base.

Johnson, Null, Butcher and Johnson (1984) factored the whole pool of 556 items from the MMPI and found 21 Varimax-rotated factors to be replicable across random halves of a sample of 11,138 psychiatric patients. The factors were categorized rationally according to Norman's (1963) Five-Factor model, and Johnson *et al.* (1984) concluded that the MMPI factors 'reflect a large number or range of personality traits rather than just those related to emotional stability' (p. 112). However, none of the MMPI factors could be classified to the content domain of the factor conscientiousness of the Five-Factor model. The observed similarities, however, were only defined via rational classifications. The relation between the MMPI factors and the Big Five could have been interpreted much more adequately if Johnson *et al.* (1984) had applied some sort of higher-order factor analysis.

Such an analysis was undertaken by Costa, Zonderman, McCrae, and Williams (1985). They were unable to replicate the 21 MMPI factors postulated by Johnson *et al.* (1984) in a new sample of 1,576 normal subjects. They found only 11 interpretable factors and they accepted a solution with nine Varimax-rotated factors, because this was the psychologically most interpretable solution. Five of the nine components — neuroticism, cynicism, extraversion, religious orthodoxy, and intellectual interests — correspond closely to five factors found by Johnson *et al.* (1984). From these, neuroticism and extraversion correspond to two factors of the Five-Factor model. The cynicism factor may be related to the agreeableness factor, and the fifth factor of the Five-Factor model may include the MMPI intellectual interests factor. Then, nine scale scores were computed for each subject by summing up the items with absolute loadings exceeding 0.3 on each factor. These scale scores were inter-correlated, and a component analysis followed by Varimax rotation of the nine scales yielded one major factor that was defined by five psychopathology scales. Two additional minor factors were defined by the scales (1) masculinity versus femininity and religious orthodoxy, and (2) extraversion and intellectual interests.

Costa *et al.* (1985) concluded that the MMPI appears to represent only one of the five Normal dimensions (neuroticism) and partly to represent three others. However, these speculations also require empirical proof.

As in the case of *scale*-factor analyses, the results of these extensive *item*-factor analyses lead us to conclude that at least variants of the Big-Five factors have been found in many of the studies mentioned. However, there were once more clear discrepancies in individual cases. In many studies, various facets described by the five factors were actually represented, but often to a differing extent depending on the factor. Factors represented only by a few of their facets had little chance of appearing. Nevertheless, no important and robust factor was found beyond the Big Five.

The resulting item factors and their interpretation varied not only according to the kind of sample of variables chosen but also according to the author of the study and the methods of factor analysis applied. These aspects can be seen in the discussion on the factor systems of Cattell, Guilford, and Eysenck. Whereas, for example, the study by Cattell and Gibbons (1968) could confirm only vaguely the Cattell factors exvia, anxiety, pathemia, and independence, which in our opinion correspond only weakly with the Big Five, the most dominant factors published by Sells *et al.* (1970) proved to be surprisingly similar to the Big Five. Sells *et al.*'s study was also based on a representative selection of Cattell's and Guilford's items.

Similar discrepancies can be found between the publications of those authors who examined samples of questionnaire items or scales from the three major factor analytical personality systems (Amelang and Borkenau, 1982; Eysenck and Eysenck, 1969; Vagg and Hammond, 1976). Whereas Eysenck and Eysenck (1969) discovered only the factors neuroticism and extraversion (what else was to be expected in 1969?), Vagg and Hammond (1976) found four robust factors in their replication study, three of which seem to show a high correspondence with the factors neuroticism, extraversion, and openness. Facets of the remaining two Big-Five factors were also represented in the study, although not comprehensively enough to withstand analysis as separate factors. The results of the study by Amelang and Borkenau (1982) offer the clearest evidence that the superstructure of the factor models of Cattell, Guilford, and Eysenck can be well described by the Big Five.

As Table 4.3 shows, neuroticism and extraversion appear as remarkably robust factors even in item-factor analyses. Compared to the results of the scale-factor analyses, the evidence for an independent item factor conscientiousness is relatively weak. This is most probably due to the different samples of variables analysed in the respective studies. For example, in scale analyses, the factor conscientiousness was usually represented adequately by scales from the PRF, the CPS, and the 16PF. In contrast, questionnaire items for the traits order, achievement, endurance, reliability, and self-control, which are the typical marker items for the factor conscientiousness, were poorly represented even in those studies in which extensive and supposedly representative samples of questionnaire items were factor-analysed (such as Browne and Howarth, 1977; Eysenck and Eysenck, 1969; Sells *et al.*, 1970).

In comparison with the dominant extraversion and neuroticism factors, empirical evidence for the factors agreeableness and openness is also weak. However, the

relative weakness of the factors conscientiousness, agreeableness, and openness does not necessarily mean that they are less important personality dimensions. On the contrary, this weakness could possibly be explained simply by the strong *overrepresentation* of neuroticism and extraversion items in personality questionnaires. This overrepresentation may, in turn, merely be a sign of the idiosyncratic preference of many researchers for the factors extraversion and neuroticism. For example, studies of the structure of representative sets of personality-descriptive terms have revealed that the factor neuroticism plays a much less important role in ordinary personality language than a review of many personality questionnaire studies might lead us to anticipate (see Ostendorf, 1990; Peabody, 1987; Peabody and Goldberg, 1989). In these representative studies, neuroticism generally explained the lowest percentage of variance compared to the other four Big-Five factors (for example, only 25–58 per cent of the variance explained by the factor agreeableness or conscientiousness). Consequently, many personality inventories do not refer adequately to some of the most important individual differences that ordinary people observe in their daily transactions.

As in the case of scale-factor analyses, a typical characteristic of the reported item-factor analyses is that the labelling of the factors depends solely on the imagination of the authors. Empirical methods such as the Recaptured Item Technique suggested by Meehl, Lykken, Schofield, and Tellegen (1971), which could enable the idiosyncratic interpretations to be made more objectively, are not yet in use even in the most up-to-date studies. Similarly, the deduced correspondence between factors found in a study and those found in earlier studies is always the result of an interpretation on behalf of the researchers involved and is not based on empirical evidence — for instance, on the use of marker items.

Of course, this criticism also applies to our own review. That is, the taxonomy of factors presented in Tables 4.2 and 4.3 is also found solely on our subjective interpretations of factor patterns reported in the studies discussed. The only way to carry out a rigorous test of the validity of the proposed classification would be to replicate all reported studies. To test the putative factor structure empirically, it would then be necessary to include additional items or scales, known as marker variables, measuring the Big Five. Unfortunately, such a comprehensive study would take a lot of time and money to perform. Instead, until now, we have chosen to test the Five-Factor model with our own set of data, which was compiled earlier for other research purposes. In the present studies, we investigated the structural validity of the Five-Factor model on the basis of 576 items taken from personality questionnaires related to different major theories of personality.

Testing the Big-Five-Factor structure on a comprehensive set of 576 questionnaire items

Each of the questionnaires used in our study purports to assess a collection of personality constructs that cover a major proportion of the personality sphere. To check the validity of the interpretation of the questionnaire factors, we correlated

these factors with factors calculated from marker variables of the Big Five in the domain of rating data. Two kinds of rating datum were used for this comparison: first, questionnaire factor scores were correlated with factor scores from various standard rating inventories measuring the Big Five. Then, the questionnaire factors were correlated with the five factors that had previously been found in analyses of self- and peer-rating data obtained for a representative set of 430 German personality-descriptive adjectives (Ostendorf, 1990).

Study I: Questionnaire study

Method

The rating and questionnaire data sets were collected in two different studies.[4]

Subjects

Subjects of the questionnaire study were 300 German adults (171 females, 129 males) with a mean age of 26.4 years (*SD* = 9.84 years). They were recruited through an announcement in the local newspapers of the City of Bielefeld. They had various occupational backgrounds and were paid for their participation.

Measures

Subjects were asked to fill out the following questionnaires in addition to performing other tests: the German version of the Personality Research Form A (PRF-A; Jackson, 1984; Stumpf *et al.* 1985), the Freiburg Personality Inventory-Revised Form (FPI-R; Fahrenberg *et al.*, 1984), a personality questionnaire widely used in Germany; the Eysenck Personality Inventory Form A (EPI-A; Eggert, 1974; Eysenck and Eysenck, 1968; and the NEO Personality Inventory (Borkenau and Ostendorf, 1986; Costa and McCrae, 1985).

For the purpose of this study, all items from the control scales of the questionnaires (*SD*-scales, Lie-scales) were dropped from the analysis. All 224 items in the German adaptation of the PRF (which contains 14 of the original American scales) were included.[5] All 125 items were taken from the 11 scales of the FPI-R.[6] In addition, we used all 47 items from the neuroticism and extraversion scales of the EPI-A and all 180 items of the NEO Personality Inventory. Altogether, we analysed a sample of 576 questionnaire items.

In addition to the questionnaire items, subjects in Study I completed 20 bipolar adjective rating scales. These rating scales were translations of the scales suggested by Norman (1963). Because a different translation of the 20 Norman scales was administered five months later to subjects in Study II, we labelled the rating scales in Study I the *Norman-A scales*.

Results

To determine the factor structure of the item sample, we applied principal component analysis to the correlation matrix of the 576 questionnaire items (for the sake of simplicity, we will call components 'factors').

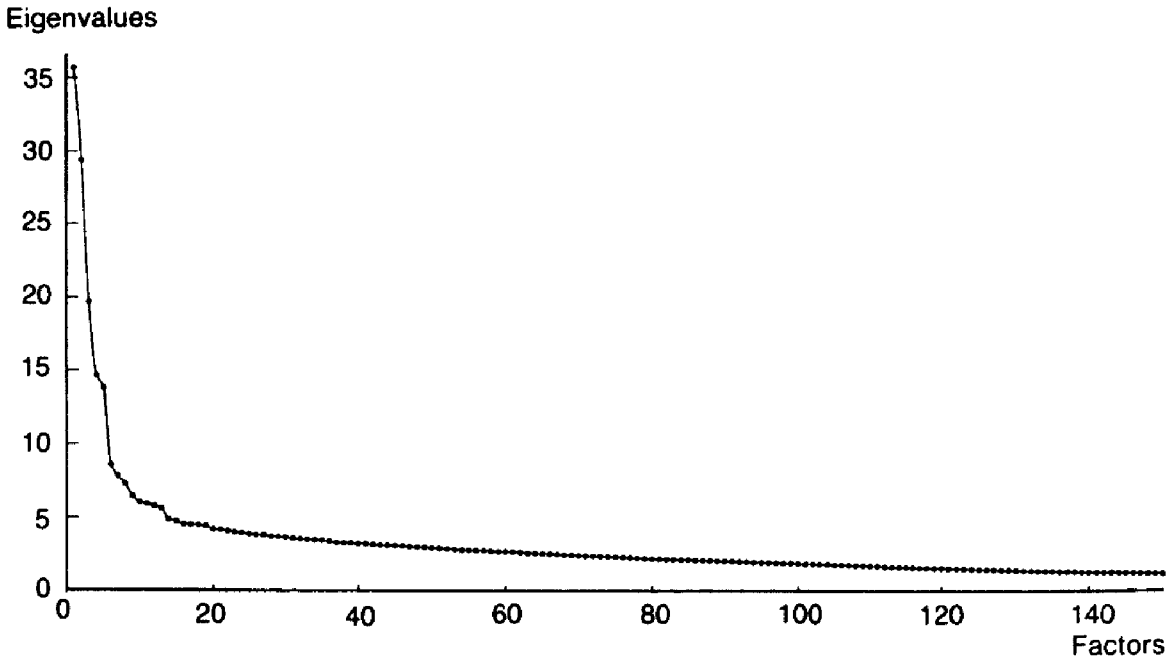

Figure 4.1 Analysis of 576 questionnaire items: plot of the first 150 eigenvalues

As shown in Figure 4.1, the plot of eigenvalues indicated five large factors, with a clear 'break' at the sixth eigenvalue. The eigenvalues for the first ten unrotated components were 35.7, 29.4, 19.6, 14.6, 13.8, 8.6, 7.8, 7.3, 6.5, and 6.0, respectively.

The first five factors were well ahead of the others in terms of percentage of explained variance. Of greater interest than the number of important factors is the correspondence between the first five dominant factors and the Big Five. In order to answer this question, we extracted five factors and rotated them according to the Varimax criterion. Those questionnaire items that loaded highest on the five factors are reported in Tables 4.4–4.8. It can be seen that all five factors could be interpreted as clear variants of the Big Five.

Table 4.4 shows the 20 items with the highest loadings on the neuroticism factor. This factor was the most dominant principal component of the analysis, accounting for 27.8 per cent of the common variance explained by all five factors. Even items loading on the neuroticism factor but belonging — according to the scoring key of a questionnaire — to conceptually different traits could be seen to be descriptive of the neuroticism dimension; for example, the item belonging to the dominance scale of the PRF: 'I feel incapable of handling many situations.'

The other factors could also be interpreted quite clearly: the factor extraversion explained 16.5 per cent of the common variance and was primarily marked by items from the different extraversion scales of the personality inventories as well as by items from the PRF-Affiliation scale (see Table 4.5). Of the 20 items with the highest loadings, 13 were from extraversion scales, four from the PRF-affiliation scale, two from the PFR-exhibition scale, and one from the PRF-play scale. Extraverted subjects with high scores on this factor 'really enjoy talking to people', they 'consider

Table 4.4 Varimax structure of 576 questionnaire items. 20 items with highest loadings on factor IV (neuroticism)

Questionnaire scale	Item	E	A	C	N	O
NEO-N	I am seldom sad or depressed.	06	−04	−01	**−67**	−19
NEO-N	I rarely feel lonely or blue.	10	−01	02	**−63**	−24
NEO-N	I often feel helpless and want someone else to solve my problems.	08	09	−25	**60**	−04
FPI-CO	I'm hardly ever in a depressed, unhappy mood.	14	00	00	**−57**	−25
NEO-N	I'm pretty stable emotionally.	00	02	19	**−56**	−15
NEO-N	Sometimes things look pretty bleak and hopeless to me.	−07	03	−10	**56**	21
EPI-N	Are you often troubled with feelings of guilt?	−07	09	−06	**56**	13
NEO-N	I feel I am capable of coping with most of my problems.	01	01	18	**−55**	03
NEO-N	It takes a lot to get me mad.	−22	06	03	**−55**	04
PRF-DO	I feel incapable of handling many situations.	−02	20	**−35**	**55**	−07
NEO-N	I rarely feel fearful or anxious.	−12	−02	−02	**−53**	−06
EPI-N	Are you troubled with feelings of inferiority?	−07	08	−12	**53**	11
FPI-PS	I'm often nervous, because too much seems to happen at once.	−02	08	17	**53**	07
NEO-N	When I'm under a great deal of stress, sometimes I feel like I'm going to pieces.	12	−00	−11	**52**	04
NEO-N	I have fewer fears than most people	00	−14	05	**−52**	06
FPI-IR	In general, I'm a very calm person and it's not easy to arouse me.	−23	27	02	**−52**	09
EPI-N	Are you an irritable person?	03	−12	13	**51**	06
FPI-CO	All in all, I am very happy with my previous life.	17	09	02	**−51**	−16
NEO-N	Too often, when things go wrong, I get discouraged and feel like giving up.	−06	04	**−32**	**50**	03
NEO-N	I am easily frightened.	16	15	−13	**50**	−09

Notes: N = 300. All loadings ≥0.30 are listed in **bold**. NEO = NEO Personality Inventory: FPI = Freiburg Personality Inventory; EPI = Eysenck Personality Inventory; PRF = Personality Research Form; Questionnaire scales: N = neuroticism; CO = contentedness; IR = irritability; PS = proneness to stress; DO = dominance.

themselves as especially light-hearted' (NEO-extraversion), they have many friendships (PRF-affiliation), and they report that other people do *not* consider them 'as a serious, reserved person' (PRF-play).

Conscientiousness was not an appropriate label for the third factor. Digman's (1990) 'will to achieve' offered a better description. This factor explained 19.1 per cent

Table 4.5 Varimax structure of 576 questionnaire items. 20 items with highest loadings on factor I (extraversion)

Questionnaire scale	Item	E	A	C	N	O
NEO-E	I like to have a lot of people around me.	**62**	−12	−09	−04	−03
FPI-E	I only make friends slowly.	**−59**	16	00	03	01
PRF-AF	My friendships are many.	**55**	−08	−00	−10	−09
NEO-E	I really enjoy talking to people.	**53**	07	03	−10	07
PRF-PL	People consider me as a serious, reserved person.	**−50**	13	08	05	−02
PRF-EX	At a party I enjoy entertaining others.	**50**	−16	−01	−01	12
NEO-E	I don't consider myself especially 'light-hearted'.	**−48**	09	−02	25	20
EPI-E	Do other people think of you as being very lively?	**48**	−28	07	−10	13
PRF-AF	I spend a lot of time visiting friends.	**48**	−07	−23	−12	−03
NEO-E	I usually prefer to do things alone.	**−47**	−09	13	−00	02
FPI-E	I'm quite a lively person.	**47**	−23	03	−15	15
NEO-E	I laugh easily.	**46**	04	−11	−11	09
NEO-E	Many people think of me as somewhat cold and distant.	**−46**	−10	05	−03	13
NEO-E	I'm known as a warm and friendly person.	**46**	21	08	−08	−17
EPI-E	Can you usually let yourself go and enjoy yourself a lot at a gay party?	**46**	−23	09	−16	25
NEO-E	I am a cheerful, high-spirited person.	**45**	01	−04	**−35**	−10
PRF-EX	Others think I am lively and witty.	**45**	−22	00	−14	05
PRF-AF	I choose hobbies I can share with other people.	**45**	04	−08	−12	−05
PRF-AF	I try to be in the company of friends as much as possible.	**45**	−08	−18	−01	−07
NEO-E	Sometimes I bubble with happiness.	**44**	02	−00	07	27

Notes: $N = 300$. All loadings ≥ 0.30 are listed in **bold**. NEO = NEO Personality Inventory; FPI = Freiburg Personality Inventory; EPI = Eysenck Personality Inventory; PRF = Personality Research Form; Questionnaire scales: E = extraversion; AF = affiliation; PL = play; EX = exhibition.

of the common variance. It was marked primarily by items from the NEO-conscientiousness scale (6 items), the PRF-endurance scale (5 items), the PRF-achievement scale (3 items), the PRF-play scale (3 items with negative loadings), and 2 items from the PRF-order and the PRF-impulsivity scales (see Table 4.6). One of the 20 items with the highest loadings was from the NEO-scale extraversion, which is clearly misclassified in the NEO-inventory: 'I have a leisurely style in work and play.' Items such as 'I work hard to accomplish my goals' and 'I strive for

Table 4.6 Varimax structure of 576 questionnaire items. 20 items with highest loadings on factor III (conscientiousness or will to achieve)

Questionnaire scale	Item	E	A	C	N	O
NEO-C	I waste a lot of time before settling down to work.	02	−09	**−56**	26	29
NEO-C	I work hard to accomplish my goals.	09	−08	**55**	−09	−15
NEO-C	I have a clear set of goals and work toward them in an orderly fashion.	07	−13	**49**	−23	−07
PRF-PL	Even if I had the money and the time, I would not feel right just playing around.	−11	19	**47**	10	−01
PRF-PL	I spend a good deal of time just having fun.	23	−20	**−47**	−04	07
PRF-EN	If I run into great difficulties on a project, I usually stop work rather than try to solve them.	12	−05	**−47**	17	−16
NEO-C	I strive for excellence in everything I do.	−07	−04	**46**	−02	−10
PRF-EN	The mere prospect of having to put in long hours working makes me tired.	01	−14	**−44**	20	04
NEO-C	I am a productive person who always gets the job done.	08	19	**43**	−21	**−32**
PRF-PL	Most of my spare moments are spent relaxing and amusing myself.	06	−19	**−43**	01	−05
PRF-IM	Often I stop in the middle of an activity in order to start something else.	09	−02	**−43**	**32**	24
PRF-EN	I do not like leaving anything unfinished.	−03	04	**43**	−19	−28
PRF-OR	I rarely clean out my bureau drawers.	−13	−03	**−42**	01	**37**
PRF-EN	When other people give up working on a problem, I usually quit too.	07	15	**−42**	15	−21
PRF-AC	I do not mind working while other people are having fun.	−22	11	**42**	−15	01
PRF-AC	I enjoy difficult work.	03	06	**42**	−01	03
PRF-EN	I will continue working on a problem even with a severe headache.	−08	−01	**41**	−14	06
PRF-AC	I enjoy work more than play.	−15	11	**41**	01	−03
NEO-E	I have a leisurely style in work and play.	−17	12	**−40**	02	05
NEO-C	I try to perform all the tasks assigned to me conscientiously.	11	24	**40**	−11	−25

Notes: N = 300. All loadings ≥0.30 are listed in **bold**. NEO = NEO Personality Inventory: PRF = Personality Research Form. Questionnaire scales: C = conscientiousness; PL = play; EN = endurance, IM = impulsivity; OR = order; AC = achievement; E = extraversion.

Table 4.7 Varimax structure of 576 questionnaire items. 20 items with highest loadings on factor II (agreeableness)

Questionnaire scale	Item	E	A	C	N	O
NEO-A	If necessary, I am willing to manipulate people to get what I want.	−06	**−61**	−02	−07	14
PRF-AG	I am reluctant to distress someone even if I do not like him.	04	**53**	−03	−09	13
PRF-DO	I do not have a forceful or dominating personality.	−16	**49**	−23	04	−09
FPI-RE	I prefer to remain in the background at social and public events.	**−38**	**48**	−05	07	−09
NEO-E	I have often been a leader of groups I have belonged to.	08	**−47**	18	−09	23
PRF-EX	The idea of acting in front of a large group does not appeal to me.	−23	**46**	−04	−04	−13
NEO-A	Some people think I'm selfish and egoistical.	−06	**−45**	−01	09	21
NEO-E	I like to be where the action is.	**33**	**−45**	−02	09	14
NEO-N	I am known as hot-blooded and quick-tempered.	27	**−45**	02	16	02
FPI-E	I like to take command of group activities.	13	**−43**	18	−19	12
PRF-EX	I like to be in the spotlight.	**30**	**−43**	11	01	17
PRF-DO	I have little interest in leading others.	−15	**43**	−25	05	14
NEO-A	I generally try to be thoughtful and considerate.	13	**42**	03	17	07
PRF-EX	I am more a listener than a talker.	−09	**42**	−16	−00	−17
PRF-DO	I would make a poor military leader.	03	**42**	−20	11	10
NEO-A	I would rather cooperate with others than compete with them.	09	**41**	−19	07	17
PRF-DO	I feel confident when directing the activities of others.	03	**−41**	24	−20	−05
PRF-AG	I would never start a fight with someone.	20	**41**	02	−10	−10
PRF-AG	I try to show self-restraint to avoid hurting other people.	08	**40**	15	04	03
NEO-A	Some people think of me as cold and calculating.	−28	**−39**	07	−02	15

Notes: N = 300. All loadings ≥0.30 are listed in **bold**. NEO = NEO Personality Inventory: PRF = Personality Research Form; FPI = Freiburg Personality Inventory. Questionnaire scales: A = agreeableness; AG = aggression; DO = dominance; RE = restraint; E = extraversion; EX = exhibition; N = neuroticism.

excellence in everything I do' (NEO-conscientiousness) were typically representative for the meaning of the third factor. This third questionnaire factor primarily described only one facet of the conceptually broader factor conscientiousness in the Five-Factor model: will to achieve. Further central facets, such as orderliness, honesty, punctuality, and self-control were only marginally covered by this third factor.

The factor agreeableness explained 20 per cent of the extracted variance and was marked primarily by items from the following scales: NEO-agreeableness (5 items), PRF-aggression (3 items), PRF-dominance (4 items), and PRF-exhibition (3 items; see Table 4.7). Traits such as dominance and exhibition are usually associated more strongly with the extraversion factor. Most of the items in the extraversion, dominance, and exhibition scales that loaded on the agreeableness factor referred to — at least in terms of German culture — socially *un*desirable, aggressive dominance over others (for example, 'I have a forceful or dominating personality', 'I would make a good military leader', PRF-dominance; 'I like to be in the spotlight', PRF-exhibition; 'I like to take command of group activities', FPI-extraversion). Subjects who agreed with such items as 'I would rather cooperate with others than compete with them' and 'I'm not willing to manipulate people to get what I want' from the NEO-agreeableness scale and who described themselves as 'unselfish, not egoistical, thoughtful, and considerate' tended to negate items from the extraversion, dominance, and exhibition scales. These results corresponded with our findings on the rating data domain showing that aggressive dominance and its opposite pole, unassertiveness, were important components of the agreeableness factor (Ostendorf, 1990). Such results may in fact reflect actual cultural trait differences. Whereas dominance is a more socially desirable trait in American culture and can be said to be an elementary part of North-American lifestyle, it is generally associated with aggression and seen as a negative character trait in Germany. Here, dominance is seen as an uncooperative way of imposing one's will without regard for the interests of others.

The greatest differences in opinion between authors are seen in the interpretation of the fifth factor of the Norman taxonomy. Norman (1963) himself has labelled this factor culture, while McCrae and Costa (1987, in press) refer to openness to experience. In our own studies, based on self- and peer-ratings, we have found a factor intellect (Angleitner and Ostendorf, 1989; Ostendorf, 1990) analogous to the results of Goldberg (1990) and Peabody and Goldberg (1989). In our present study, the fifth factor was also the most difficult to interpret. It explained 16.6 per cent of the common variance. As Table 4.8 shows, in line with our hypotheses, the factor was marked in most cases by items belonging to the openness construct of Costa and McCrae (1985).

However, at the same time, some items describing a lack of orderliness correlated with this factor. These correlations deviated from the Five-Factor structure usually found in rating data in which items from the trait sphere order load primarily on the third factor conscientiousness (Digman, 1989; Goldberg, 1990; Hofstee *et al.*, 1992; McCrae and Costa, 1987; Ostendorf, 1990). Consequently, it may be difficult

Table 4.8 Varimax structure of 576 questionnaire items. 20 items with highest loadings on factor V (openness)

Questionnaire scale	Item	E	A	C	N	O
PRF-OR	I feel comfortable in a somewhat disorganized room.	−02	−01	**−30**	−09	**47**
NEO-O	I often enjoy playing with theories or abstract ideas.	−17	−13	−07	−06	**46**
NEO-O	I find philosophical arguments boring.	12	03	−24	−16	**−44**
NEO-O	I follow the same route when I go somewhere.	−02	01	−02	08	**−42**
PRF-UN	Abstract ideas are of little use to me.	03	09	−07	−04	**−42**
PRF-OR	I seldom take the time to hang up my clothes neatly.	−15	−02	**−39**	01	**42**
NEO-O	I have a wide range of intellectual interests.	−11	−06	14	01	**41**
NEO-O	I believe that laws and social policies should change to reflect the needs of a changing world.	−08	21	−04	09	**41**
NEO-O	I believe letting students hear controversial speakers can only confuse and mislead them.	07	14	−01	−01	**−41**
NEO-C	I never seem to be able to get organized.	−04	−01	**−36**	**31**	**41**
PRF-SR	Nothing would hurt me more than to have a bad reputation.	27	14	10	10	**−41**
NEO-O	I believe that the different ideas of					

cont.

to discriminate between the fifth questionnaire factor and the rating factor conscientiousness. Subjects with high scores on the fifth questionnaire factor can be described as intellectual, imaginative, and liberal. They think critically about socio-political issues, are interested in arts, and do not take much care of their appearance or their personal belongings (such as, clothing, home). The prototype can be described as the 'absent-minded professor'.

In order to provide empirical support for the interpretation of the five questionnaire factors from the NEO Personality Inventory, the NEO-PI facets were factored and rotated by the Validimax method, as recommended by McCrae and Costa (1989b). Correlations between questionnaire-item and Validimax-factor scores ranged from 0.73 (for the conscientiousness factor) to 0.89 (for the Neuroticism factor) with a mean of 0.81. As expected, the highest hetero-factor correlation was −0.33 between the fifth questionnaire factor (openness) and the third factor from the NEO-PI (conscientiousness). On the whole, these results demonstrated that the questionnaire factors could readily be interpreted in accordance with the Big-Five factor structure.

Table 4.8 *cont.*

Questionnaire scale	Item	E	A	C	N	O
	right and wrong that people in other societies have may be valid for them.	−06	08	−08	−08	**40**
NEO-O	Sometimes when I am reading poetry or looking at a work of art, I feel a chill or wave of excitement.	−03	13	−00	13	**40**
NEO-C	I like to keep everything in its place so I know just where it is.	−14	12	12	−10	**−40**
NEO-O	I experience a wide range of emotions or feelings.	22	−01	16	11	**40**
NEO-C	I keep my belongings neat and clean.	00	12	25	−09	**−40**
PRF-EX	When I am in the crowd, I want others to notice me.	15	−21	−04	**−30**	**40**
NEO-O	I would have difficulty just letting my mind wander without control or guidance.	−04	08	13	−10	**−38**
FPI-SO	As the state already takes care of welfare, I don't have to help people personally.	−20	−24	−07	−22	**−38**
PRF-SR	I constantly try to make people think highly of me.	**39**	12	07	05	**−38**

Notes: N = 300. All loadings ≥0.30 are listed in **bold**. PRF = Personality Research Form; NEO = NEO Personality Inventory: FPI = Freiburg Personality Inventory. Questionnaire scales: OR = order; O = openness; UN = understanding; C = conscientiousness; SR = social recognition; EX = exhibition; SO = social orientation.

Study II: Convergence of questionnaire and adjective rating factors

Method

Subjects

Six months later, 95 of the 300 subjects from the questionnaire study took part in a second study to test the Five-Factor model on the basis of adjective rating data (Ostendorf, 1990). A total of 414 subjects participated in this rating study (170 males and 239 females; 5 did not report gender). Again, subjects were recruited through announcements in local newspapers in Bielefeld and neighbouring cities. They filled out the test material at home and were not paid for their participation. Their ages ranged from 15 to 81 years with a mean of 32.6 years (SD = 13.3).

Measures

The subjects filled out adjective rating inventories constructed by various authors. The marker variables consisted of German translations of the rating inventories

developed by Norman (1963; Norman-B scales), Goldberg (1983, 1989), McCrae and Costa (1987), and Peabody and Goldberg (1989). The Norman-B scales contained German translations that differed from those of the Norman-A scales used in Study I. The final list was made up of 179 separate bipolar adjective rating scales. This also contained German translations of 80 adjectives from Norman's Five-Factor taxonomy (see Goldberg, 1990) published by John (1983; see John *et al.*, 1984). Because all of these Big-Five standard rating inventories use bipolar scales, we have labelled them the Bipolar Adjective Rating Scale (BARS).

Additionally, ratings on Unipolar Adjective Rating Scales (UARS) were obtained for a representative sample of 430 trait adjectives previously selected from a total pool of 5,160 German personality-descriptive adjectives (see Angleitner *et al.*, 1990; Ostendorf, 1990). A detailed description of the rating scales administered can be found in Ostendorf (1990). Because each of the rating inventories involved a separate operationalization of the Five-Factor model by a different author, diverse indicators were available to test the convergence of rating and questionnaire factors.

Results

Taking the total sample of 414 subjects as a basis, we first performed a separate principal component analysis for each list of adjective rating scales, extracted five factors, and rotated them according to the Varimax criterion. The five postulated factors were found very clearly in each of these analyses (see Ostendorf, 1990). Then, we calculated factor scores on the basis of the rating factors from Study II and the questionnaire factors from Study I. Factor scores were also computed for each subject on the basis of the self-rating inventories from Study I (Norman-A factors). Then, all adjective and questionnaire factor scores were correlated in order to perform a higher-order principal component analysis. This was done to test the congruence and discrimination of the five factors across the different methods.

The eigenvalues of the first six principal components were 9.1, 7.9, 7.6, 6.7, 6.1, and 1.7. The first five of the second-order factors explained 83 per cent of the total variance. Table 4.9 presents the loadings of the primary rating and questionnaire factors on five Varimax-rotated second-order components.

Table 4.9 shows that the second-order factors could be interpreted clearly in terms of the Big Five. All rating and questionnaire factors loaded highest on the appropriate factors and generally had only relatively low loadings on the other factors. This was particularly true for the factors from the bipolar adjective rating scales. Factors from the Norman-A scales showed somewhat smaller loadings, which — compared to the loadings of the Norman-B scales — could be attributed not only to the different kind of translation but also to temporal changes in the characteristics. Although the Norman-A scales were applied five months earlier under different conditions in Study I, the Norman-A factors generally had impressive convergent and discriminant validities.

Table 4.9 shows that factors from the total list of all bipolar rating scales (BARS179) had particularly high loadings on the secondary factors. This finding was to be

expected, as the total list contained the scales from all previously mentioned rating inventories (with the exception of the Norman-A scales). None the less, the operationalizations of the five factors differed greatly: for example, Peabody's list of 57 rating scales (Peabody, 1987; Peabody and Goldberg, 1989) was developed without reference to the Five-Factor model in order to compile a representative selection of trait terms for the total vocabulary of American trait adjectives.

Similarly, the list of 430 unipolar adjective rating scales (UARS) was compiled on the basis of lexical analyses of the German language — completely independently from the Five-Factor model. This list represented a sample of prototypical trait adjectives that could be viewed as representative for the entire range of trait-descriptive terms in the German language. Ostendorf (1990) has shown that the Big Five provide an adequate description of the structure of this representative trait list. As Table 4.9 shows, with the exception of the fourth factor (emotional stability), each primary factor had a high and relatively unequivocal loading on its corresponding secondary factor. The weaker correspondence for emotional stability was probably due to the small number of German adjectives describing it. As in American studies, emotional stability explained only a small proportion of the common variance of representative adjective-rating samples in German-speaking countries (Ostendorf, 1990; Peabody and Goldberg, 1989).

An interpretation of the loadings of the primary questionnaire factors in Table 4.9 has to take account of the five-month time interval between the presentation of the questionnaires and most of the rating inventories. Three of the five questionnaire factors (extraversion/surgency, agreeableness, and neuroticism vs. emotional stability) exhibited a pattern of high convergent and discriminant validity; that is, they had their highest loadings on the corresponding second-order factors, whereas they correlated only marginally with factors from which they were expected to discriminate. In comparison, the loading of the third questionnaire factor on the second-order factor conscientiousness was lower. This outcome was expected, as the questionnaire factor primarily assesses a specific facet (will to achieve) of the conscientiousness factor. None the less, the more specific questionnaire factor could be interpreted unequivocally, as there were no significant secondary loadings on conceptually unrelated secondary factors. The fifth questionnaire factor, which primarily marked the secondary factor intellect vs. openness, exhibited an anticipated, significantly negative relationship to the second-order factor conscientiousness.

Discussion

In summary, we may conclude that our studies have revealed highly unequivocal variants of the Big Five in the domain of questionnaire data. In contrast to the common practice of factor interpretation, our findings are supported by not only a subjective inspection of factor-loading patterns but also objective empirical tests of the convergence of questionnaire and rating factors.

Our studies have examined the common structure of questionnaire items from the

Table 4.9 Higher-order Varimax-rotated factor structure calculated from first-order adjective-rating and questionnaire factors

First-order self-rating adjective factors	Higher-order factors I SU/EX	II AG	III CO	IV N	V IN/OP
Factors from bipolar adjective rating scales					
Norman-A (20 scales)					
Surgency	**80**	−16	−05	−08	−04
Agreeableness	00	**75**	05	−16	−14
Conscientiousness	−09	11	**82**	−08	00
Emotional stability	−11	−13	−03	**82**	−11
Culture	−01	−07	−00	−00	**43**
Norman-B (20 scales)					
Surgency	**94**	−02	09	01	11
Agreeableness	−13	**89**	05	−06	09
Conscientiousness	−03	−01	**90**	11	−06
Emotional stability	−09	01	−14	**88**	−04
Culture	−02	13	04	22	**85**
NEO-R (80 scales)					
Surgency	**96**	−10	−03	−02	−04
Agreeableness	01	**97**	10	01	−03
Conscientiousness	03	05	**98**	02	10
Neuroticism	−04	04	04	**97**	02
Openness	02	06	−08	13	**96**
Peabody and Goldberg (57 scales)					
Surgency	**96**	−04	14	−10	−02
Agreeableness	−01	**95**	06	04	19
Conscientiousness	−13	11	**95**	−01	−06
Emotional stability	16	07	06	**87**	19
Intellect	01	−13	01	00	**95**
Goldberg (40 scales)					
Surgency	**91**	−27	−06	−06	07
Agreeableness	15	**95**	−02	08	00
Conscientiousness	08	11	**96**	06	−06
Emotional stability	−04	00	−02	**96**	10
Intellect	−07	06	04	05	**97**

cont.

PRF, the FPI, the EPI, and the NEO-PI. This selection of items represents a *broad* spectrum of personality characteristics as assessed by popular personality questionnaires. None the less, it certainly cannot be viewed as a representative sample drawn from the population of all conceivable personality-descriptive items. By drawing on commonly used questionnaires, it reflects far more the specific research foci of the various scientists who developed them.

Some general conclusions can be drawn from the review of comprehensive factor analyses and the results of our own studies:

1. Factors E and N are the most robust factors in nearly all of the reported studies, and in comparison to other item or scale factors, they frequently explain the

Table 4.9 *cont.*

First-order self-rating adjective factors	Higher-order factors I SU/EX	II AG	III CO	IV N	V IN/OP
John (40 scales)					
Surgency	**80**	**43**	−13	20	−05
Agreeableness	**−51**	**80**	−01	15	−10
Conscientiousness	17	19	**93**	09	−12
Emotional stability	−11	−03	01	**95**	12
Culture	02	14	03	−00	**94**
BARS179 (179 scales)					
Surgency	**98**	02	09	−03	−02
Agreeableness	−13	**98**	06	00	−00
Conscientiousness	−05	08	**98**	05	01
Emotional stability	01	10	04	**98**	04
Intellect	−01	06	−02	10	**98**
Factors from 430 unipolar adjective rating scales					
Surgency	**91**	−11	−10	−20	02
Agreeableness	10	**89**	12	−05	13
Conscientiousness	−05	−09	**89**	−27	−11
Emotional stability	18	−05	25	**63**	**30**
Intellect	−14	−02	−04	−14	**87**
First-order questionnaire factors					
I	**77**	05	−05	10	−12
II	−24	**78**	17	05	−01
III	04	02	**54**	−03	11
IV	−15	−02	−16	**84**	−16
V	29	−03	**−42**	−05	**54**

Notes: N = 98 subjects. All loadings ≥0.30 are listed in **bold**. Factors from bipolar rating scales: Norman-A: scales from Norman (1963), translated by Borkenau (1988); Norman-B: scales from Norman (1963), translated by Ostendorf (1990); NEO-R: NEO Rating Inventory (McCrae and Costa, 1985a); Peabody and Goldberg: scales from Peabody and Goldberg (1989); Goldberg: scales from Goldberg (1983); John: scales from John (1982); BARS179: 179 bipolar rating scales. Factors from unipolar rating scales: these are based on a representative set of 430 trait adjectives. Questionnaire factors = factors based on 576 questionnaire items.

largest proportions of variance in questionnaire data. The dominance of the factors E and N can be attributed to the overrepresentation of both E- and N-items in the domain of questionnaire data. This overrepresentation can be found even in studies designed to sample a broad spectrum of questionnaire items and in which the largest and most comprehensive pools of questionnaire items have been analysed (see Browne and Howarth, 1977; Sells *et al.*, 1970). The item samples utilized in these studies have generally been compiled on the basis of literature surveys; that is, from popular, broad-band personality questionnaires. This sampling strategy will almost inevitably lead to an overrepresentation of E- and N-items, because most personality inventories measure at least some aspects of E and N, but either assess the domains of other factors less systematically or neglect their measurement completely.

2. Neuroticism is found to be by far the most dominant factor. However, neuroticism usually explains the smallest proportion of variance when factor analyses are based on representative samples of trait-descriptive adjectives (Ostendorf, 1990; Peabody and Goldberg, 1989). This difference may be due to many questionnaire authors preferring to measure clinical aspects of personality, while these personality features apparently receive only little consideration in the ordinary language of western cultures. In German, for instance, there are roughly three times as many adjectives describing the domain of agreeableness as adjectives describing characteristics of neuroticism.
3. On the other hand, the measurement of some major personality characteristics seems to be neglected in the domain of Q data when the range of personality features measured by an average or typical personality inventory is compared with the bandwidth of the Big-Five rating factors. This particularly applies to the factor openness, which cannot be found in Eysenck's P-E-N model of personality and has only a relatively weak representation in Guilford's and Cattell's factor models. Important facets of this factor, such as self-reports on abilities, intelligence, creativity and fantasy, gifts, talents, artistic skills, and musical as well as cultural needs and interests are rarely measured by personality questionnaires. Although there are certainly good reasons for measuring abilities through objective tests, self- or peer-reports on these characteristics are in no way less interesting or less important. Compared with the domain of trait adjectives, questionnaire items measuring openness are strongly underrepresented in our present empirical studies. Apart from relevant items in the NEO-openness scale, only the PRF-understanding scale assesses central facets of the fifth factor. The underrepresentation of relevant items may explain, first, the somewhat lower level of congruence of the openness factors across the Q- and L-data media, and second, the relatively low discrimination between the questionnaire factor openness and the adjective-rating factor conscientiousness. A long list of personality traits that correlate with openness is reported by McCrae and Costa (1985c, in press). However, most of these traits are rarely measured by the scales of very specific questionnaires.
4. Scales measuring facets of agreeableness and conscientiousness are usually underrepresented in broad-band personality inventories as well. For example, traits like pro-social behaviour, altruism, nurturance, and so forth describe important individual differences that have long been neglected in personality research (see Rushton, 1981). This is documented by the fact that many popular textbooks on personality psychology at most refer only marginally to this trait domain. As our review and results have shown, the likelihood that an independent conscientiousness factor will emerge from a common-factor analysis of several broad-band personality inventories is usually greater than that for the factors agreeableness and openness. The reason that attributes of conscientiousness are measured more frequently by personality inventories may be traced back to the important role of achievement and competition behaviour in western industrial societies. Relevant questionnaire scales are frequently rooted in (1) Freudian

theory (for example, superego strength, self-control) and (2) achievement motivation research (for example, need for achievement, need for endurance). In contrast, scales measuring related characteristics like conscientiousness, honesty, reliability and will are only seldom included in personality inventories. For example, some central facets of the conscientiousness factor represented by adjectives such as discreet, honest, deliberate, firm, thrifty, responsible, reliable, punctual, self-disciplined, strong-willed, foresighted, organized, scrupulous, and thorough have no equivalents in our pool of questionnaire scales. In the analysis, only a few core features of the construct are represented by the PRF scales need for achievement, need for endurance, and need for order, and the FPI scale achievement orientation. This may explain why the narrow questionnaire factor conscientiousness shows a high discriminant validity but only a moderate convergence with the broader adjective rating factor.

By and large, the correspondence of the five factors across the two data media is quite substantial. Observed deviations from the hypothesized, ideal structure can be explained by the different bandwidths of the factors in each data medium. An inspection of the scales and items shows that the bandwidth of the rating factors seems to be broader than that of the questionnaire factors. Further empirical studies are required to determine how far these differences in bandwidth will affect the instruments' power to predict important life criteria.

Although we have analysed a broad spectrum of questionnaire items, we are not able to present any persuasive arguments confirming the representativeness of our item selection. (Actually, the selection of personality inventories was guided by the aims of a research programme that was completely unrelated to the Five-Factor model under study.)[4] More powerful tests of the Big-Five factor structure in questionnaire data will require much larger and more representative samples of questionnaire items. For example, future studies could start by using the large item pools compiled by French and his colleagues. Taking French's (1953, 1973) comprehensive survey of the literature as a basis, Dermen, French, and Harman (1978) wrote questionnaire items to assess all the discernible, homogeneous facets of 28 factors previously identified in at least three analyses carried out in at least two different laboratories. L.R. Goldberg administered more than 1,400 of these items to large samples of subjects. Unfortunately, only preliminary results of this study have been reported (Dermen *et al.*, 1974; French and Dermen, 1974). It would be an interesting task for future research to explore the relationships between the structure of this comprehensive pool of questionnaire items and the Five-Factor model. Likewise, it may be worth reanalysing the studies of Browne and Howarth (1977), Eysenck and Eysenck (1969), and Sells *et al.* (1970).

But we anticipate that the results of such large item studies would not differ greatly from our general conclusions: extraversion and neuroticism would emerge as the most dominant factors, while the structure of all the remaining traits would be largely described by conscientiousness, agreeableness, and openness to experience. However, it would be most interesting to search for additional robust questionnaire factors

beyond the first five and to test whether these factors carry a substantial meaning not already covered by the Big-Five personality factors originally found in the domain of rating data.

Authors' notes

This research was supported in part by a grant from the German science foundation to the second author (DFG-Az. 106/10-1).

We thank J. Harrow and F. Spinath for their assistance in preparing this manuscript.

Notes

1. For a discussion of differences between personality and temperament see Strelau (1987) and Hofstee (1991).
2. These restrictions refer to Conn and Ramanah's (1990) study that resulted in no clearly interpretable neuroticism and extraversion factors. Unfortunately, we cannot report any statistics on the amount of variance explained by the factors, as most authors do not provide the necessary information.
3. The use of PFHs was recommended by Guilford: 'The initial planning should emphasize the formation of hypotheses as to what factors are likely to be found in the selected domain' (1952, p. 36). Thus, Browne and Howarth (1977) expected 20 replicable factors, representative of the domain of questionnaire items.
4. We are indebted to Peter Borkenau for allowing us to use his questionnaire data.
5. PRF-scales: achievement, affiliation, aggression, dominance, endurance, exhibition, harm-avoidance, impulsivity, nurturance, order, play, social recognition, succourance, understanding.
6. FPI-R-scales: contentedness, social orientation, achievement orientation, restraint, irritability, aggression, proneness to stress, somatic complaints, health concerns, extraversion, emotionality.

References

Amelang, M. and 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 emotion lability], *Zeitschrift für Differentielle und Diagnostische Psychologie*, **3**, 119–46.

Angleitner, A. and Ostendorf, F. (1989), 'Personality factors via self- and peer-ratings based

on a representative sample of German trait descriptive terms', paper presented at the First European Congress of Psychology, Amsterdam, The Netherlands, 2–7 July, 1989.

Angleitner, A. and Ostendorf, F. (1991), 'Temperament and the Big-Five factors of personality', paper presented at the Conference on the Developing Structure of Temperament and Personality in Childhood, Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Wassenaar, The Netherlands, 17–20 June, 1991.

Angleitner, A., Ostendorf, F. and John, O.P. (1990), 'Towards a taxonomy of personality descriptors in German: a psycho-lexical study', *European Journal of Personality*, **4**, 89–118.

Barrett, P.T. and Kline, P. (1980), 'The location of superfactors P, E, and N within the unexplored factor space', *Personality and Individual Differences*, **1**, 239–47.

Bendig, A.W. (1962), 'Factor analysis of the Guilford–Zimmerman Temperament Survey', *Journal of General Psychology*, **67**, 21–6.

Borkenau, P. (1988), 'The multiple classification of acts and the Big-Five factors of personality', *Journal of Research in Personality*, **22**, 337–52.

Borkenau, P. and Ostendorf, F. (1986), 'The NEO-Personality Inventory', unpublished German translation, University of Bielefeld, FRG.

Boyle, G.J. (1989), 'Re-examination of the major personality-type factors in the Cattell, Comrey and Eysenck Scales: were the factor solutions by Noller *et al.* optimal?', *Personality and Individual Differences*, **10**, 1289–99.

Browne, J.A. and Howarth, E. (1977), 'A comprehensive factor analysis of personality questionnaire items: a test of twenty putative factor hypotheses', *Multivariate Behavioral Research*, **12**, 399–427.

Cattell, R.B. and Gibbons, B.D. (1968), 'Personality factor structure of the combined Guilford and Cattell personality questionnaires', *Journal of Personality and Social Psychology*, **9**, 107–20.

Cattell, R.B., Eber, H.J. and Tatsuoka, M.M. (1970), *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*, Champain, IL: Institute for Personality and Ability Testing.

Church, A.T. and Katigbak, M.S. (1989), 'Internal, external, and self-report structure of personality in a non-western culture: an investigation of cross-language and cross-cultural generalizability', *Journal of Personality and Social Psychology*, **57**, 857–72.

Comrey, A.L. (1970), *Manual for the Comrey Personality Scales*, San Diego, CA: Educational and Industrial Testing Service.

Conn, S.R. and Ramanaiah, N.V. (1990), 'Factor structure of the Comrey Personality Scales, the Personality Research Form-E, and the Five-Factor model', *Psychological Reports*, **67**, 627–32.

Costa, P.T., Jr. and McCrae, R.R. (1976), 'Age differences in personality structure: a cluster analytic approach', *Journal of Gerontology*, **31**, 564–70.

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

Costa, P.T., Jr. and McCrae, R.R. (1988), 'From catalog to classification: Murray's needs and the Five-Factor model', *Journal of Personality and Social Psychology*, **55**, 258–65.

Costa, P.T., Jr., Zonderman, A.B., McCrae, R.R. and Williams, R.B., Jr. (1985), 'Content and comprehensiveness in the MMPI: an item factor analysis in a normal adult sample', *Journal of Personality and Social Psychology*, **48**, 925–33.

De Raad, B., Mulder, E., Kloosterman, K. and Hofstee, W.K.B. (1988), 'Personality-descriptive verbs', *European Journal of Personality*, **2**, 81–96.

Dermen, D., French, J.W. and Harman, H.W. (1974), *Verification of Self-report Temperament Factors*, Technical Report 6, Princeton, NJ: Educational Testing Service. Report No. ED 104 912.

Dermen, D., French, J.W. and Harman, H.W. (1978), *Guide to Factor-referenced Temperament Scales*, Princeton, NJ: Educational Testing Service, Report No. NR 150 329.

Digman, J.M. (1989), 'Five robust trait dimensions: development, stability, and utility', *Journal of Personality*, **57**, 195–214.

Digman, J.M. (1990), 'Personality structure: emergence of the five-factor model', *Annual Review of Psychology*, **41**, 417–40.

Eggert, D. (1974), *Eysenck-Persönlichkeits-Inventar, EPI: Handanweisung für die Durchführung und Auswertung* [Manual of the Eysenck Personality Inventory, EPI], Göttingen, FRG: Hogrefe.

Eysenck, H.J. (1978), 'Superfactors P, E and N in a comprehensive factor space', *Multivariate Behavioral Research*, **13**, 475–81.

Eysenck, H.J. and Eysenck, M.W. (1985), *Personality and Individual Differences: A natural science approach*, New York, NY: Plenum Press.

Eysenck, H.J. and Eysenck, S.B.G. (1968), *Manual of the Eysenck Personality Inventory*, San Diego, CA: Educational and Industrial Testing Service.

Eysenck, H.J. and Eysenck, S.B.G. (1969), *Personality Structure and Measurement*, London: Routledge & Kegan Paul.

Fahrenberg, J. and Selg, H. (1970), *Das Freiburger Persönlichkeitsinventar (FPI)* [The Freiburg Personality Inventory], Göttingen, FRG: Hogrefe.

Fahrenberg, J., Hampel, R. and Selg, H. (1984), *Das Freiburger Persönlichkeitsinventar (FPI-R)* [The Freiburg Personality Inventory (FPI-R)], Göttingen, FRG: Hogrefe.

French, J.W. (1953), *The Description of Personality Measurements in Terms of Rotated Factors*, Princeton, NJ: Educational Testing Service.

French, J.W. (1973), *Toward the Establishment of Noncognitive Factors through Literature Search and Interpretation*, Princeton, NJ: Educational Testing Service.

French, J.W. and Dermen, D. (1974), *Seeking Markers for Temperament Factors among Positive and Negative Poles of Temperament Scales*, Princeton, NJ: Educational Testing Service, Report No. ED 104 909.

Gibbons, B.D. (1966), 'A study of the relationships between factors found in Cattell's 16PF questionnaire and factors found in the Guilford personality inventories', unpublished doctoral dissertation, University of Southern California.

Goldberg, L.R. (1978), *Language and Personality: Developing a taxonomy of personality-descriptive terms. A progress report and research proposal*, University of Oregon and Institute for Measurement of Personality, 1201 Oak Street, Eugene, OR 97401, USA.

Goldberg, L.R. (1980), 'Some ruminations about the structure of individual differences: developing a common lexicon for the major characteristics of human personality', a contribution to the symposium 'Personality: Beyond and Beneath the Factors', Honolulu, Hawaii, May, 1980.

Goldberg, L.R. (1983), 'The magical number five, plus or minus two: some considerations on the dimensionality of personality descriptors', paper presented at a Research Seminar, Gerontology Research Center, NIA/NIH, Baltimore, MD, June 1983.

Goldberg, L.R. (1989), 'Standard markers of the Big-Five factor structure', paper presented at the Invited Workshop on Personality Language, Groningen, The Netherlands, 26–30 June 1989.

Goldberg, L.R. (1990), 'An alternative "description of personality": the Big-Five factor structure', *Journal of Personality and Social Psychology*, **59**, 1216–29.

Goldberg, L.R. (1991), 'Comparing the Big-Five factor structure with its competitors: I. Eysenck's P-E-N model', paper presented at the Conference on the Development Structure of Temperament and Personality in Childhood, Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Wassenaar, The Netherlands, 17–20 June 1991.

Guilford, J.P. (1952), 'When not to factor analyze', *Psychological Bulletin*, **49**, 26–37.

Guilford, J.S., Zimmerman, W.S. and Guilford, J.P. (1976), *The Guilford–Zimmerman-Temperament Survey Handbook: Twenty-five years of research and application*, San Diego, CA: Edits Publishers.

Hofstee, W.K.B. (1991), 'The concepts of personality and temperament', in J. Strelau and

A. Angleitner (eds.), *Explorations in Temperament: International perspectives on theory and measurement*, New York, NY: Plenum Press, pp. 177–88.

Hofstee, W.K.B., De Raad, B. and Goldberg, L.R. (1992), 'Integration of the Big Five and circumplex approaches to trait structure', *Journal of Personality and Social Psychology*, **63**, 146–63.

Howarth, E. (1980), 'Major factors of personality', *Journal of Personality*, **104**, 171–83.

Howarth, E. and Browne, J.A. (1971a), 'Investigation of personality factors in a Canadian context. I. Marker structure in personality questionnaire items', *Canadian Journal of Behavioral Science*, **3**, 161–73.

Howarth, E. and Browne, J.A. (1971b), 'An item-factor-analysis of the 16PF', *Personality*, **2**, 117–39.

Isaka, H. (1990), 'Factor analysis of trait terms in everyday Japanese language', *Personality and Individual Differences*, **11**, 115–24.

Jackson, D.N. (1976), *JPI–Jackson Personality Inventory — Manual*, Goshen, NY: Research Psychologists Press.

Jackson, D.N. (1984), *Personality Research Form Manual* (3rd edn.), Goshen, NY: Research Psychologists Press.

John, O.P. (1982), 'German Adjective List' (GAL), unpublished assessment instrument, University of Bielefeld, Experimental and Applied Psychology Section, FRG.

John, O.P. (1990), 'The "Big-Five" factor taxonomy: dimensions of personality in the natural language and in questionnaires', in L.A. Pervin (ed.), *Handbook of Personality: Theory and Research*, New York, NY: Guilford Press, pp. 66–100.

John, O.P., Goldberg, L.R. and Angleitner, A. (1984), 'Better than the alphabet: taxonomies of personality-descriptive terms in English, Dutch and German', in H.C.J. Bonarius, G.L.M. Van Heck and N.G. Smid (eds.), *Personality Psychology in Europe. Theoretical and empirical developments*, Lisse, The Netherlands: Swets & Zeitlinger, vol. 1, pp. 83–100.

Johnson, J.H., Null, C., Butcher, J.N. and Johnson, K.N. (1984), 'Replicated item level factor analysis of the full MMPI', *Journal of Personality and Social Psychology*, **47**, 105–14.

Kline, P. and Barrett, P. (1983), 'The factors in personality questionnaires among normal subjects', *Advances in Behavioral Research and Therapy*, **5**, 141–202.

Krug, S.E. and Johns, E.F. (1986), 'A large scale cross-validation of second-order personality structure defined by the 16PF', *Psychological Reports*, **59**, 683–93.

McCormick, I.A., Green, D.E. and Walkey, F.H. (1991), 'A comparison between first, second, and third-order factor analysis in the multi-scale questionnaire — the Eysenck Personality Inventory', *Personality and Individual Differences*, **12**, 43–8.

McCrae, R.R. (1989), 'Why I advocate the five-factor model: joint factor analyses of the NEO-PI with other instruments', in D.M. Buss and N. Cantor (eds.), *Personality Psychology: Recent trends and emerging directions*, New York, NY: Springer, pp. 237–45.

McCrae, R.R. and Costa, P.T., Jr. (1985a), 'Updating Norman's adequate taxonomy: intelligence and personality dimensions in natural language and in questionnaires', *Journal of Personality and Social Psychology*, **49**, 710–21.

McCrae, R.R. and Costa, P.T., Jr. (1985b), 'Comparison of EPI and Psychoticism scales with measures of the five-factor model of personality', *Personality and Individual Differences*, **6**, 587–97.

McCrae, R.R. and Costa P.T., Jr. (1985c), 'Openness to experience', in R. Hogan and W.H. Jones (eds.), *Perspectives in Personality*, Greenwich, CT: JAI Press, vol. 1, pp. 145–72.

McCrae, R.R. and Costa, P.T., Jr. (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. and Costa, P.T., Jr. (1989a), 'Reinterpreting the Myers–Briggs Type Indicator from the perspective of the Five-Factor model of personality', *Journal of Personality*, **57**, 17–40.

McCrae, R.R. and Costa, P.T., Jr. (1989b), 'Rotation to maximize the construct validity of factors in the NEO Personality Inventory', *Multivariate Behavioral Research*, **24**, 107–24.

McCrae, R.R. and Costa, P.T., Jr. (in press), 'Conceptions and correlates of openness to experience', in S.R. Briggs, R. Hogan and W.H. Jones (eds.), *Handbook of Personality Psychology*, New York, NY: Academic Press.

McCrae, R.R., Costa, P.T., Jr. and Busch, C.M. (1986), 'Evaluating comprehensiveness in personality systems: the California Q-Set and the five-factor model', *Journal of Personality*, **54**, 430–46.

McKenzie, J. (1988), 'Three superfactors in the 16PF and their relations to Eysenck's P, E and N', *Personality and Individual Differences*, **9**, 843–50.

Matthews, G., Stanton, N., Graham, N.C. and Brimelow, C. (1990), 'A factor analysis of the scales of the occupational personality questionnaire', *Personality and Individual Differences*, **11**, 591–6.

Meehl, P.E., Lykken, D.T., Schofield, W. and Tellegen, A. (1971), 'Recaptured-Item Technique (RIT): a method for reducing somewhat the subjective element in factor naming', *Journal of Experimental Research in Personality*, **5**, 171–90.

Montag, I. and Comrey, A.L. (1990), 'Stability of major personality factors under changing motivational conditions, in J.W. Neuliep (ed.), *Handbook of Replication Research in the Behavioral and Social Sciences*, Special Issue, *Journal of Social Behavior and Personality*, **5**, 265–74.

Noller, P., Law, H. and 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**, 775–82.

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**, 574–83.

Norman, W.T. (1967), '2800 personality trait descriptors: normative operating characteristics for a university population', Department of Psychology, University of Michigan.

Nunnally, J.C. (1978), *Psychometric Theory* (2nd edn.), New York, NY: McGraw-Hill.

Ostendorf, F. (1990), *Sprache und Persönlichkeitsstruktur. Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit* [Language and Personality Structure: On the structural validity of the Five-Factor model of personality], Roderer, FRG: Regensburg.

Peabody, D. (1987), 'Selecting representative trait adjectives', *Journal of Personality and Social Psychology*, **52**, 59–71.

Peabody, D. and Goldberg, L.R. (1989), 'Some determinants of factor structures from personality-trait descriptors', *Journal of Personality and Social Psychology*, **57**, 552–67.

Piedmont, R.L., McCrae, R.R. and Costa, P.T., Jr. (1991), 'Adjective Check List scales and the five factor model', *Journal of Personality and Social Psychology*, **60**, 630–7.

Rushton, J.P. (1981), 'The altruistic personality', in J.P. Rushton and R.M. Sorrentino (eds.), *Altruism and Helping Behavior*, Hillsdale, NJ: Erlbaum, pp. 251–66.

Saville & Holdsworth, Ltd (1984), *Occupational Personality Questionnaires Manual*, Esher: Saville & Holdsworth, Ltd.

Schneewind, K.A. (1977), 'Entwicklung einer deutschsprachigen Version des 16 PF-Tests von Cattell' [Development of the German version of Cattell's 16 PF-test], *Diagnostika*, **23**, 188–91.

Schrueger, J.M. and Allen, L.C. (1986), 'Second-order factor structure common to five personality questionnaires', *Psychological Reports*, **58**, 119–26.

Sells, S.B., Demaree, R.G. and Will, D.P., Jr. (1970), 'Dimensions of personality: I. Conjoint factor structure of Guilford and Cattell trait markers', *Multivariate Behavioral Research*, **5**, 391–422.

Strelau, J. (1987), 'The concept of temperament in personality research', *European Journal of Personality*, **1**, 107–17.

Stumpf, H., Angleitner, A., Wieck, T., Jackson, D.N. and Beloch-Till, H. (1985), *Deutsche Personality Research Form (PRF)*, Göttingen, FRG: Hogrefe.

Vagg, P.R. and Hammond, S.B. (1976), 'The number and kind of invariant personality (Q) factors: a partial replication of Eysenck and Eysenck', *British Journal of Social and Clinical Psychology*, **15**, 121–9.

Wiggins, J.S. and Trapnell, P.D. (in press), 'Personality structure: the return of the Big Five', in S.R. Briggs, R. Hogan and W.H. Jones (eds.), *Handbook of Personality Psychology*, New York, NY: Academic Press.

Yang, K. and Bond, M.H. (1990), 'Exploring implicit personality theories with indigenous or imported constructs: the Chinese case', *Journal of Personality and Social Psychology*, **58**, 1087–95.

Zuckerman, M. (1979), *Sensation Seeking: Beyond the optimal level of arousal*, Hillsdale, NJ: Erlbaum.

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

Zuckerman, M., Kuhlman, D.M., Thornquist, M. and Kiers, H. (1991), 'Five (or three) robust questionnaire scale factors of personality without culture', *Personality and Individual Differences*, **9**, 929–41.