

## **The use of scores anchored at the scale midpoint in reporting individuals' traits**

WILLEM K. B. HOFSTEE\*†  
and A. A. JOLIEN HENDRIKS  
*University of Groningen, The Netherlands*

### *Abstract*

*Standardized scores including factor scores are insufficient indicators of individuals' positions on personality traits when mean scores deviate considerably from the scale origin. Anchored scores, that is, raw scores expressed as deviations from the scale midpoint and divided by the standard deviation, retain this discrepancy. We argue in favour of reporting in terms of anchored scores in addition to standard scores, and present a solution to the specific problem of finding anchored factor (i.e., principal-component) scores. Such scores are obtained through dividing factor weights by the standard deviation of the variable in question, and applying the corrected weights to the matrix of raw scores, expressed as deviations from the scale midpoint (e.g. -2, -1, 0, +1, +2). © 1998 John Wiley & Sons, Ltd.*

## **INTRODUCTION**

Imagine a person who has filled out a personality questionnaire, and whose scores on a number of dimensions are reported to him or her. Typically, these scores will have been centred at the mean of the population. Consequently, if most individuals in the population have reported, for example, a preponderance of Social over Asocial behaviour, and if our subject scores somewhat below the mean on this dimension, we report that this subject is somewhat Asocial. However, the subject might object that the reporting contains no more than a half-truth: a score below the mean may well be above the midpoint of the scale. Using a scale centred at the scale midpoint, the person would thus have been reported to be on the Social side.

The logic of the objection is not affected by the details of the example. First, the receiver of the reporting may be a third person, for example, a parent, a teacher, or a

\*Correspondence to: Willem K. B. Hofstee, The Heymans Institute, University of Groningen, 9712 TS Groningen, The Netherlands.

†E-mail address: w.k.b.hofstee@ppsw.rug.nl

personnel selection officer. No lay person should be expected to spontaneously interpret psychological reportings in terms of relative scores (see Chaplin and Buckner, 1988). Second, the source of the assessment may consist of third persons rather than the target person himself or herself; they might be equally surprised to hear that this person was reported to be somewhat Asocial.

In the above argument, the midpoint of the rating scale acquires an absolute value of zero. The reasoning implies that personality scales are bipolar rather than unipolar; for example, that Asocialness is conceived as the opposite of Socialness, not as the mere absence of Socialness. The negation (a-, un-, dis-) is thus conceived as a *litotes* (*Oxford Dictionary*: expressing of an affirmative by the negative of its contrary). Arguments for this bipolar conception of personality scales are presented below.

Mean scores are generally on the socially desirable side of the scale midpoint, as follows from the high correlation between social desirability and the probability that a trait will be endorsed (Edwards, 1953), and may also be verified by inspecting any norm tables of published personality questionnaires. Thus the overall effect of using relative scales is that subjects are described in less socially desirable terms than they describe themselves and each other. To the extent that this psychological view of humankind is merely an unintended side-effect of a technical convention, it should be corrected.

It goes without saying that scores anchored at the scale midpoint ('anchored' scores<sup>1</sup> for short) have no special place in correlational research. However, personality assessment is not merely an interchange among investigators of personality. One of its products consists of scores to be communicated to others, like the assessed person or a third party. Our argument pertains to this context of application. After elaborating the general argument, we present new procedures leading to anchored scores on principal components, as principal-component analysis provides the most efficient way of summarizing the information in personality assessments.

## ARE PERSONALITY SCALES BIPOLAR?

An analytic argument for the bipolar conception of personality scales runs as follows (see Hofstee, 1990); if, for example, Unselfishness denoted the mere absence of Selfishness (unipolar conception), it would occupy the same position in the personality space (namely, at the origin) as Asocialness, whereas the two are largely antonymous. As Socialness is largely synonymous with Unselfishness, and Asocialness with Selfishness, all such meanings would regress toward the origin of the trait space, which would be absurd.

<sup>1</sup>We apologize to an anonymous reviewer and other readers of an earlier version for initially having designated anchored scales as bipolar 'ratio' (as opposed to mere interval) scales. In the use of Likert scales for personality assessment, few attempts are usually made to secure a numerical interpretation of scale units, for example, by using behaviour counts; thus the term 'ratio scale' is not appropriate. We do not agree that true units of measurement could not be constructed in personality assessment, for lack of equivalence. Certainly, being late for a party is different from being late for a job interview; but by that reasoning one could not even say that one building contains twice as many bricks as another, since each brick is as unique as Eysenck's proverbial old shoe. We do agree that any use of words such as 'scale' or 'absolute' (we purposely and carefully avoid 'measurement') in the present context is informal. Most importantly, however, we disagree with our reviewer that the metric issue has any impact on our argument for a bipolar conception of personality scales. One may have metric and nonmetric scales with or without a natural origin.

A first empirical argument relies in the social desirability values of trait terms. For opposite terms, these values are generally complementary (see e.g. Hofstee, 1990): for example, Asocial is about as undesirable as Social is desirable. If Asocialness denoted the mere absence of Socialness, it should logically be neither desirable nor undesirable. The finding that an important facet of the meaning of trait terms, namely their social desirability, behaves in a bipolar manner, argues for the bipolar conception.

Only a minority of the trait terms have logical opposites (unless one wishes to enrich the language with terms like Unsilent, Unharsh, Unsloppy, and Unmoody), and questionnaire items generally do not. However, the general case may be studied by having subjects assess the social desirability of modified terms, that is, by asking questions such as 'How desirable is it to be extremely (c.q. rather, little, extremely little) Moody'. Hofstee (1992) performed such an experiment. Subjects were given the 60 combinations of 15 trait terms and four modifiers. The trait terms were selected at five levels of social desirability ranging from clearly undesirable to clearly desirable. Subjects were instructed to judge each combination by itself, and to associate the modifiers with values on a 100-point scale, as follows: very little, 10; little, 30; rather, 70; extremely, 90. For their social desirability judgements of the modifier–trait combinations, they used a scale ranging from 'clearly negative' to 'clearly positive'.

To be little or very little Cordial (or any of the desirable traits) was judged to be negatively desirable, and to be little or very little Nervous (or any of the undesirable traits) was judged to be positively desirable. For all five levels of trait social desirability, the desirability in function of modifier intensity cut the intensity scale at about the midpoint (50). In other words, to be judged neither socially desirable nor undesirable falls on the midpoint, not on the 'very little' point. 'Very little Cordial' or 'little Nervous' again appear to be interpreted in the manner of a litotes, indicating the opposite of the trait rather than its mere absence. The bipolar interpretation is thus not restricted to cases in which trait terms have logical opposites.

Another empirical argument for a bipolar interpretation of trait scales is found in trait inference research. Relevant behaviours—for example, Taking the floor—that are judged prototypical for a trait (e.g. dominance), appear to be perceived as disconfirming instances for the opposite trait (i.e. submissiveness; Borkenau, 1988). When behaviours are observed in natural settings such as a group discussion, differences in overall activity among participants may obscure this pattern; however, it reappears upon partialling out the overall activity variable (Borkenau and Müller, 1992). These findings cannot be explained upon assuming a unipolar conception of traits.

All this is not to exclude that there are contexts in which a denial of a trait refers to its mere absence. For example, A may say to B that he finds C extroverted, and B may disagree; that does not necessarily mean that B finds C introverted. With respect to the context of responding to questionnaires, however, we have established that the dominant conception of trait scales is bipolar.

## IS THE AVERAGE PERSON SOCIALLY DESIRABLE?

To the objection that population-referenced scales discard relevant absolute information, one might reply that no such information exists in personality assessment, and that absolute standards cannot be found. Surely in a sense, all judgment is relative. To

say that people are tall would be meaningless except with reference to our ancestors, other animals, or the like. So-called absolute norms for educational tests can only mean that the grading does not depend on the particular class of students; such norms would be adapted if they were found to be unrealistic in a larger population, and would thus appear to be relative after all. According to a classical philosophical tenet, humans are the measure of all things. So, does it make sense at all to say that people on average are Social? A definitive answer, positive or negative, is not available and may not even be conceivable. However, a number of practical arguments for retaining absolute in addition to relative information can be brought forward.

The trait labels used for reporting scores on personality variables may be conceived as summaries (Bem and Allen, 1974) of relevant behaviours. Standard scoring would thus imply that the average person is about as likely to behave in a socially undesirable as in a desirable manner. This is hard to believe. Almost by definition, most people would be judged to behave in a socially desirable manner most of the time.

Another way to articulate the common-sense position is by noting that some negative trait poles denote psychopathology. Standard scoring would convey the implausible message that about half of the normal population suffers from mental illness. The absurdity of that implication led Millon (1982), for example, to make the norms for the MCMI correspond to population base rates.

The most obvious objection to a statement such as 'most people are social' consists of the tenet that average scores above the midpoint of the scale arise through mere socially desirable responding, in the sense of misrepresentation on the part of the subject. However, elevated average scores are observed also in research settings (e.g. Hendriks, 1997), where there is no particular incentive for impression management. Moreover, ratings by others may even be slightly more elevated than self-ratings (see below).

A strong argument against the interpretation of socially desirable responding as a mere artifact relies in the simple distinction between rater and target effects. If a person describes himself or herself as Social, one may logically attribute that statement to the described (the person is indeed Social), the describer (the person is boasting), or both. Equally, if a third person rates an individual as Social, we cannot logically distinguish between a rater and a target effect. (In this case, the equivalent of socially desirable responding is usually termed 'leniency'). However, we may have targets rated by more than one assessor, and decide on empirical grounds to what extent a rating should be attributed to the rater and/or to the ratee. Philosophically, a significant and sizeable target effect does not prove that a person is really Social or Asocial; it could represent a shared illusion. Still, the most straightforward interpretation is that the person has that trait.

In the past, socially desirable responding has been exclusively associated with self-report. The multiple-raters design could not even be considered since the self is by definition single. However as a rater, the individual is only an element of the set of people who know the target person well; he or she may not even be the *primus inter pares* (Hofstee, 1994).

Finding a target effect on socially desirable responding thus leads to the proposition that some people are more socially desirable than others. In this proposition, 'social desirability' is used as a superordinate trait. The evidence in favour of the proposition is overwhelming: (i) correlations between knowledgeable raters, including

self, are typically of the order of 0.5–0.6 for most personality scales (for an overview, see Hofstee, 1994), indicating sizeable target effects; (ii) all trait dimensions are substantially correlated with social desirability; in fact, the first principal component of a representative set of trait ratings or scores on personality scales is usually seen as a desirability dimension. Therefore, social desirability shows sizeable target effects.

We used research materials collected by Hendriks (1997) to demonstrate the target desirability effect. For 116 target persons, we took the sum of their scores on five personality factors (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Autonomy/Intellect), self-rated *versus* peer-rated. Self and average peer correlated 0.56, which is highly significant. Self and average peer scores on the Agreeableness factor, which is probably an even better operationalization of individual differences in social desirability, correlated 0.60.

Finding a differential target effect does not prove that people in general are socially desirable; it shows that some people are more socially desirable than others. However, the substantial target variance in ratings of persons' social desirabilities does make the artifact interpretation of the general effect less plausible.

Probably the strongest argument for an artifact interpretation of elevated mean scores derives from social comparison experiments in which subjects are asked to rate themselves relative to the average other person in the group or sample (see e.g. Alicke, 1985). On average, subjects rate themselves higher than average, which would seem contradictory if it is held that the ratings reflect a true state of affairs. However, subjects may not adhere faithfully to the instructions given by the experimenter. When asked to compare themselves with others on Socialness, they may start by deciding that they are Social, in other words, may respond primarily with an absolute judgment (see Chaplin and Buckner, 1988). Subsequently, they may give insufficient weight to the reference group, even if perceived correctly.

In conclusion, socially desirable responding cannot be explained satisfactorily as a mere rater's artifact. Edwards (1953) has already noted that the correlation between endorsement and social desirability '... does not necessarily mean that the subjects are misrepresenting themselves on the inventory. It may be that traits which are judged as desirable are those which are fairly widespread or common among members of a culture or group.' (p. 92). This is not to deny that the elevation of ratings may be influenced by strategic considerations elicited by situational demands, for example, in personnel selection settings. Also, our argument does not detract from the relevance of population-normed information as such. This information, however, may be retained by expressing the central tendency of the pertaining population on the same scale as the individual's score. A proper reporting would thus consist of statements such as 'You are (this person is) on the Social side, but less so than the average (modal, median) person in this population'.

In the domain of personality questionnaires, representative coverage of a scale label by the items has to be insured for effective communication. Of course, this is a general requirement. Upon retaining absolute information, however, it holds especially with respect to the social desirability of the scale label. For example, if items tapping Conscientious behaviour are labelled as a Compulsiveness scale, or Assertiveness items as a Dominance scale, the distribution of individuals on the trait would differ from their distribution on the scale, and the reporting would be deficient. Careful item production and selection, and matching the social desirability of the trait label to that of the average item, provide a practical solution to that problem.

Carrying the above reasoning a step further, an anonymous reviewer states that personality assessments *could* be interpreted in absolute terms if they represented summaries of elementary behaviours; for example, if a person is late for appointments, in finishing jobs, and the like more than half of the time, that person would be Unconscientious in an absolute sense, irrespective of other individuals. Consequently, if the world population on average displayed these behaviours more than half of the time, it could be concluded that people are Unconscientious. However, one would have to define the set of relevant behaviours for a trait, and the set of situations in which to observe them. These definitions would unavoidably be arbitrary to some extent. Also, the fifty-fifty definition of the scale origin is contestable: we reserve the right to call a person Murderous even if he has killed only a small minority of the persons who came his way. Thus the relative character of the scores returns through the back door.

The present argument for retaining absolute information should not be confused with a plea for idiographic or idiothetic (Lamiell, 1981) measurement. Clearly, we share Lamiell's concern for a faithful description of personality. However, there is no emphasis here on the dynamics of intra-personal comparison, or on any opposition between an individual differences paradigm and the study of personality. Moreover, our approach focuses on the midpoint of bipolar rating scales, which are treated as unipolar in Lamiell's reconstruction.

## SCORING PROCEDURES

The argument thus far offers no grounds for choosing between absolute and standardized bipolar scales anchored at the midpoint. Both approaches are elaborated here, with special emphasis on the development of absolute and anchored factor scores. On the basis of these elaborations, concrete arguments for and against either approach will be documented.

### Absolute scores

In the simplest case, we have an assessment instrument (e.g. a personality questionnaire) consisting of a number  $n$  of items keyed in the same direction and responded to on an  $m$ -point scale. Assuming equal intervals between these points, responses should be coded as deviations from the scale midpoint (e.g.  $-2$ ,  $-1$ ,  $0$ ,  $1$  and  $2$  for a five-point scale;  $-1.5$ ,  $-0.5$ ,  $0.5$ , and  $1.5$  for a four-point scale). Giving equal weights to all  $n$  items, the absolute score on the scale is obtained by summing the item scores and dividing by the number of items. For example, a person would obtain a score of  $-0.93$  on an Extraversion scale consisting of items rated on a five-point scale, and would be judged to be moderately (as opposed to extremely) Introverted.

In the general case, item weights may be negative (as for the item 'Do you stay away from parties?' in an Extraversion scale), and their sizes may differ. To obtain the absolute score, the weighted sum of the item scores (or each item weight) should be divided by the sum of the absolute values of the item weights. This again places the total scores on the same scale as the items: extreme responses in the Extroverted direction on all items would give an extreme total score. Dividing by the number of items is a special case of this general rule.

In rating instruments consisting of several subscales, factor scores (specifically, scores on rotated principal components) provide the most efficient way of

summarizing the information contained in the responses. The procedure for calculating absolute factor scores is straightforward:

- (i) perform a raw-score principle-component analysis of the item scores *expressed as deviations from the scale center* (with the *caveat* that standard programs will automatically turn raw scores into standard scores);
- (ii) calculate the matrix of factor weights;
- (iii) divide each column of the weights matrix by the sum of its absolute values;
- (iv) apply these weights to the matrix of raw item scores expressed as deviations from the scale center, to find the absolute factor scores.

### Anchored scores

The argument pertaining to the midpoint of a bipolar scale does not address the observed spread of raw scores. There is thus no objection in advance against standardizing scores with respect to spread. Anchored scores  $x'$ , that is, raw scores  $X$  expressed as deviations from the scale midpoint and divided by their standard deviations  $s$ , thus  $x' = X/s$ , are to be distinguished from deviation scores  $x = X - M_X$  and standard scores  $z = (X - M_X)/s$ . Naturally,  $x' = z + M_X/s$ , that is, the individual anchored score may be found by adding the mean anchored score to the standard score.

To construct anchored factor scores, a similar procedure might be applied as in constructing absolute factor scores, namely, principal-component analysis of the matrix of anchored scores, followed by adapting the weights matrix. However, a disadvantage of such factor scores—and of absolute scores—is that their factorial structure is not necessarily the same as the structure found upon standard principal-component analysis. For the purpose of retaining continuity with earlier substantive findings, 'compatible' anchored factor scores that do show the same factor structure as standard factor scores are presented below. The derivation is given in the appendix.

Compatible anchored factor scores are found by applying regular factor weights to the matrix of anchored scores (i.e. raw scores expressed as deviations from the scale midpoint and subsequently divided by their standard deviation). These anchored factor scores have unit standard deviations; they are not orthogonal but they are uncorrelated. They consist of regular factor scores plus weighted mean anchored scores. An equivalent way of constructing these factor scores is to divide each row of the matrix of regular factor weights by the standard deviation of the variable in question, and to apply the resulting weights to the matrix of raw scores expressed as deviations from the scale midpoint as above. Finally and most easily, anchored factor scores are found by (i) calculating a regular factor score, (ii) calculating the regular factor score  $z_c$  of a fictitious subject who has endorsed the midpoint of the scale (e.g. three on a five-point scale) on all items, and (iii) subtracting the algebraic value  $z_c$  (which will be negative if the scale is scored in a socially desirable direction) from all regular factor scores.

### Evaluation of anchored and absolute scores

A number of points should be made in comparing absolute and anchored scores including factor scores. On the one hand, absolute scores may seem preferable: if, for example, an individual obtains a score of +0.9 on a five-point Extraversion scale, this score can be interpreted in the terms in which the responses were given ('close to

moderately'). However, the extremeness of a score would be a function of the number of items. The larger the number, the more the weighted or unweighted mean score across items would regress towards the midpoint of the scale: with factor scores based on large numbers of items, scores outside the  $-1$  to  $+1$  range on a five-point scale ( $-2$  to  $+2$ ) appeared to be extremely rare in our materials. It may not be acceptable that extremeness is a function of the number of items.

A special feature of compatible anchored factor scores is that they yield a structure identical to regular, standard-score analysis. A practical advantage of this approach is continuity: it preserves earlier substantive results with respect to personality structure, for example, the 'Big-Five' structure.

### **An empirical example of anchored scoring**

The above procedure for calculating anchored factor scores was applied to data collected by Hendriks (1997), consisting of scores of 159 students and staff members of psychology at the University of Groningen, on 225 personality-descriptive adjectives. Both self-ratings and independent averaged ratings by a maximum of four others who knew these persons well were available. No special incentives for socially desirable responding were present in this research setting. Figure 1 presents the average profiles for self- and others' ratings of this group of subjects, in terms of anchored factor scores, on the 'Big Five' (see e.g. Goldberg, 1992) rotated principal components taken from these ratings: I, Extraversion; II, Agreeableness; III, Conscientiousness; IV, Emotional Stability; V, Autonomy/Intellect.

With the exception of Factor IV, Emotional Stability, the mean scores are on the socially desirable side of the scale midpoint. The profiles for self-ratings and others' ratings are strikingly parallel, the latter being slightly more elevated. The most

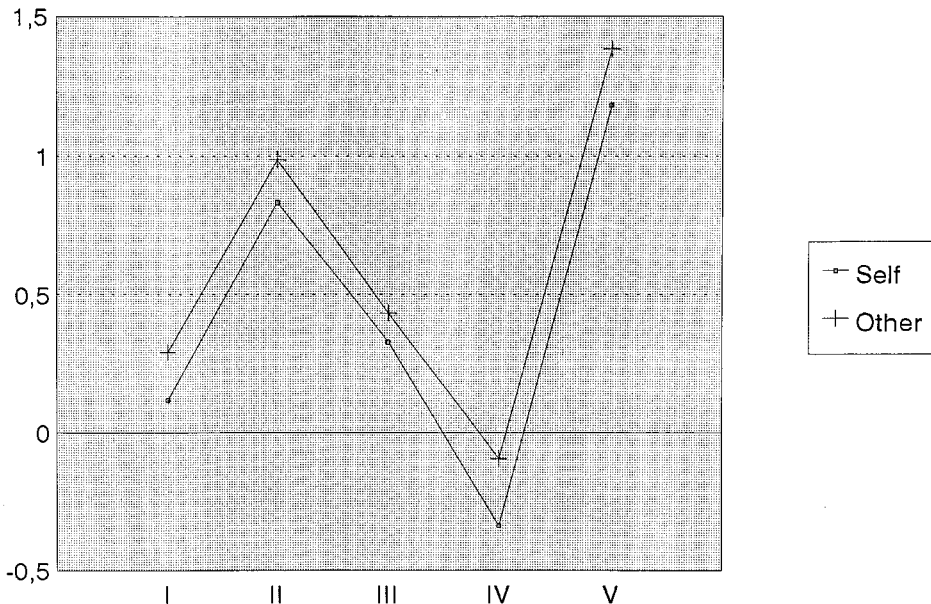


Figure 1. Average profiles for self- and others' ratings, in terms of anchored factor scores ( $s.d. = 1$ ), on the Big-Five factors of personality



straightforward interpretation of the profile is that, on average, this group of people is slightly on the Extroverted side, clearly Agreeable, somewhat Conscientious, slightly on the Emotionally Unstable side (which may be explained by the often-observed correlation of this factor with sex, the majority of the respondents being female), and clearly autonomous/Intellectual. Although the general population may be expected to score higher on Factor IV and lower on V, there is no way in which its profile would regress entirely to the scale midpoint.

## CONCLUSION

In a classic argument, Cattell (1944) has called attention to 'the Motley, Deceptive and Segregated Character of Psychometric Data' (p. 301). In his words, '... it is more valuable to know that Shakespeare at nine had a recognition vocabulary of thirty-eight thousand words than that he stood at the ninety-second percentile in the County of Warwick Vocabulary Test of that day' (pp. 299/300). Part of the problem raised by Cattell is the neglect of absolute information through the use of scales with arbitrary origins. We have presented detailed procedures to promote more informative reporting of scores on personality variables.

Our recommendations may be summarized and specified as follows. (i) Scores on personality variables, including item scores, should be expressed on a bipolar scale. (ii) Scale labels should be chosen so as to cover the content of the items, expressly including the social desirability aspect; upon using factor scores, the social desirability of the scale is the weighted mean of the item desirabilities. (iii) An individual's total score should be expressed as an anchored (factor) score. (iv) The individual score should be pitted against the central tendency of the scores of a relevant norm group, expressed on the same anchored scale.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the comments by Alois Angleitner, Lars R. Bergman, Henk A. L. Kiers, Robert R. McCrae, Fritz Ostendorf, Rainer Riemann, Eddy E. Ch. I. Roskam, and Jos F. M. Ten Berge on a draft of this paper.

## APPENDIX

The derivation of anchored factor scores proceeds from the familiar formula

$$\mathbf{ZB} = \mathbf{F} \quad (1)$$

with  $\mathbf{Z}$  a matrix of regular standard scores of  $N$  subjects,  $\mathbf{B}$  a matrix of regular factor weights of the variables on principle components, any subset of which may have been orthogonally rotated, and  $\mathbf{F}$  a matrix of regular factor scores.

The relation between  $\mathbf{Z}$  and the matrix of raw scores  $\mathbf{X}$  is given by

$$\mathbf{Z} = (\mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}^T)\mathbf{X}\mathbf{D}^{-1} \quad (2)$$

with **I** a unit matrix, **1** a unit vector, and **D** a diagonal matrix with standard deviations. Substituting (2) into (1) gives

$$(\mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}^T)\mathbf{X}\mathbf{D}^{-1}\mathbf{B} = \mathbf{F} \quad (3)$$

or

$$\mathbf{X}\mathbf{D}^{-1}\mathbf{B} - N^{-1}\mathbf{1}\mathbf{1}^T\mathbf{X}\mathbf{D}^{-1}\mathbf{B} = \mathbf{F} \quad (4)$$

Therefore

$$\mathbf{F}^1 = \mathbf{X}\mathbf{D}^{-1}\mathbf{B} = \mathbf{F} + N^{-1}\mathbf{1}\mathbf{1}^T\mathbf{X}\mathbf{D}^{-1}\mathbf{B} \quad (5)$$

Thus, the matrix of anchored factor scores  $\mathbf{F}^1$  is found by (i) expressing item scores **X** as deviations from the scale midpoint, (ii) dividing them by their standard deviations, and (iii) weighting them by their regular factor weights; they consist of regular factor scores **F** plus the weighted mean anchored scores.

## REFERENCES

- Alicke, M. D. (1985). 'Global self-evaluation as determined by the desirability and controllability of trait adjectives', *Journal of Personality and Social Psychology*, **49**: 1621–1630.
- Bem, D. J. and Allen, A. (1974). 'On predicting some of the people some of the time: the search for cross-situational consistencies in behavior', *Psychological Review*, **81**: 506–520.
- Borkenau, P. (1988). 'The multiple classification of acts and the big five factors of personality', *Journal of Research in Personality*, **22**: 337–352.
- Borkenau, P. and Müller, B. (1992). 'Inferring act frequencies and traits from behavior observations', *Journal of Personality*, **60**: 553–573.
- Cattell, R. B. (1944). 'Psychological measurement: normative, ipsative, interactive', *Psychological Review*, **51**: 292–303.
- Chaplin, W. F. and Buckner, K. E. (1988). 'Self-ratings of personality: a naturalistic comparison of normative, ipsative, and idiothetic standards', *Journal of Personality*, **56**: 509–530.
- Edwards, A. L. (1953). 'The relationship between the judged desirability of a trait and the probability that the trait will be endorsed', *Journal of Applied Psychology*, **37**: 90–93.
- Goldberg, L. R. (1992). 'The development of markers for the big-five factor structure', *Psychological Assessment*, **4**: 26–42.
- Hendriks, A. A. J. (1997). 'The construction of the Five-Factor Personality Inventory (FFPI)', Doctoral dissertation, University of Groningen.
- Hofstee, W. K. B. (1990). 'The use of everyday language for scientific purposes', *European Journal of Personality*, **4**: 77–88.
- Hofstee, W. K. B. (1992). 'Complexiteit en eenvoud in relaties tussen persoonlijkheidseigenschappen' ['Complexity and simplicity in relations among personality traits'], *Nederlands Tijdschrift voor de Psychologie*, **47**: 278–287.
- Hofstee, W. K. B. (1994). 'Who should own the definition of personality?', *European Journal of Personality*, **8**: 149–162.
- Lamiell, J. T. (1982). 'Toward an idiothetic psychology of personality', *American Psychologist*, **36**: 276–289.
- Millon, T. (1982). *Millon Clinical Multiaxial Inventory (MCMI)*, Interpretive Scoring Systems, Minneapolis, MN.

Copyright of European Journal of Personality is the property of John Wiley & Sons Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.