Psychological Assessment
1992, Vol. 4, No. 1, 26–42

1040-3590/92/$3.00

The Development of Markers for the Big-Five Factor Structure

Lewis R. Goldberg
University of Oregon
and Oregon Research Institute
Eugene, Oregon

To satisfy the need in personality research for factorially univocal measures of each of the 5 domains that subsume most English-language terms for personality-traits, new sets of Big-Five factor markers were investigated. In studies of adjective-anchored bipolar rating scales, a transparent format was found to produce factor markers that were more univocal than the same scales administered in the traditional format. Nonetheless, even the transparent bipolar scales proved less robust as factor markers than did parallel sets of adjectives administered in unipolar format. A set of 100 unipolar terms proved to be highly robust across quite diverse samples of self and peer descriptions. These new markers were compared with previously developed ones based on far larger sets of trait adjectives, as well as with the scales from the NEO and Hogan personality inventories.

Over the past decade, a quiet revolution has been occurring in personality psychology, and an age-old scientific problem has recently begun to look tractable. Gradually, agreement has been growing about the number of orthogonal factors needed to account for the interrelations among English-language trait descriptors. Across a wide variety of studies involving trait-descriptive terms, five broad factors have consistently been found. Indeed, the empirical evidence for this "Big-Five" representation has been obtained in analyses of the relations among trait terms when they are used by subjects to describe themselves and others (e.g., Goldberg, 1990), as well as in analyses of judgments of the semantic similarity of the descriptors (e.g., Peabody & Goldberg, 1989). The history of this five-factor representation has been the subject of several recent reviews (e.g., Digman, 1990; John, 1990; Wiggins & Trapnell, in press). At present, investigators have begun to extend its application into new domains (e.g., Barrick & Mount, 1991; Botwin & Buss, 1989), especially into clinical contexts (e.g., McCrae & Costa, 1986, 1989a; Wiggins & Pincus, 1989).

To facilitate such applications and extensions, investigators need access to alternative sets of Big-Five markers, including ones that vary in their length, and thus in their demands on subject testing time. Evidence for the psychometric characteristics of such sets of factor markers should be available, including the extent to which they provide converging operationalizations of the Big-Five structure. This article, in conjunction with a closely related one (Goldberg, 1990), is meant to serve that purpose.

This project was supported by Grant MH-39077 from the National Institute of Mental Health.

The NEO–PI test booklets and answer sheets used in Studies 2 and 3 were provided by Paul T. Costa, Jr. The test booklets, answer sheets, and subjects' scale scores for the Hogan Personality Inventory used in Study 3 were provided by Robert Hogan. I am also indebted to Sarah E. Hampson, Willem K. B. Hofstee, Kevin Lanning, Dean Peabody, Tina K. Rosolack, and especially Oliver P. John, for their careful, thoughtful, and critical suggestions.

Correspondence concerning this article should be addressed to Lewis R. Goldberg, Oregon Research Institute, 1899 Willamette Street, Eugene, Oregon 97401.

Previous Markers of the Big-Five Structure

The Big-Five factor representation was originally discovered by Tupes and Christal (1961), on the basis of reanalyses of various data sets using bipolar variables constructed by Cattell (1957). The five factors have traditionally been numbered and labeled as follows: I. Surgency (or Extraversion); II. Agreeableness; III. Conscientiousness (or Dependability); IV. Emotional Stability (vs. Neuroticism); and V. Culture, Intellect, or Openness. From the analysis of Tupes and Christal (1961), Norman (1963) selected the four variables with the highest factor loadings on each of the five factors; using peer nominations on these 20 variables in four samples, he showed that the same five factors could be recovered in all samples. Although Norman used the complete Cattell variables in their original and complex format, in his published report he included only some "abbreviated descriptions" of the variables. For example, the Cattell variable

Insistently Orderly Tidy, over-precise, especially over details. Drives other people to be the same. Strict, fussy, pedantic. Insists on everything being orderly. (In these respects rather "uncomfortable to live with.") Seems unable to relax. Miserly.	*vs.*	Relaxed, indolent Rather careless of detail. Lazy. Careless over expenditures. Has no difficulty in relaxing. Enjoys ease.

was abbreviated by Norman (1963) as Fussy, tidy vs. Careless.

The complete versions of these 20 variables have been used as Big-Five markers in some later studies by Norman (e.g., Norman, 1969; Norman & Goldberg, 1966; Passini & Norman, 1966) as well as in cross-cultural studies by Guthrie and Bennett (1971) and by Bond and his collaborators (e.g., Bond, 1979, 1983; Bond & Forgas, 1984; Bond, Nakazato, & Shiraishi, 1975; Nakazato, Bond, & Shiraishi, 1976). In addition, some or all of Norman's abbreviated scale labels have been used in a wide variety of other studies (e.g., Borkenau, 1988; D'Andrade, 1965; Hakel, 1969; Smith & Kihlstrom, 1987; Watson, 1989; Weiss & Mendelsohn, 1986; N. Wiggins & Blackburn, 1976).

However, both sets of variables have some inherent limitations as markers of the Big-Five structure. First of all, there are only four variables included in each of these broad domains, and therefore many facets of each domain are not represented. In addition, because these variables were all selected from the original Cattell set, they include many of the idiosyncratic elements that characterize those variables (for a detailed analysis of the problematic nature of the Cattell variables, see John, Angleitner, & Ostendorf, 1988). The primary purpose of the present article is to provide more adequate sets of factor markers than those provided by Norman (1963).

At present, the major alternative set of Big-Five markers is the NEO Personality Inventory (NEO–PI), developed by Costa and McCrae (1985). That inventory provides scales measuring five domains, labeled *Neuroticism* (Factor IV), *Extraversion* (Factor I), *Openness* (Factor V), *Agreeableness* (Factor II), and *Conscientiousness* (Factor III), as well as measures of six facets of each of the Neuroticism, Extraversion, and Openness domains. Domain scores from this inventory have been used as a framework for integrating a wide variety of other self-report scales, including those developed by Eysenck (McCrae & Costa, 1985), Jackson (Costa & McCrae, 1988), Spielberger (Costa & McCrae, 1987), and Wiggins (McCrae & Costa, 1989d), as well as the scales included in the Minnesota Multiphasic Personality Inventory (Costa, Busch, Zonderman, & McCrae, 1986) and the Myers-Briggs Type Indicator (McCrae & Costa, 1989b). One of the purposes of the present article is to provide a comparison of the scales in the NEO–PI with the Big-Five marker variables developed by the present author.

The Structure of English Trait-Descriptive Adjectives

Building on the earlier work of Allport and Odbert (1936) and Cattell (1957), later investigators such as Norman (1967) and Goldberg (1982, 1990) have been able to analyze a far larger and more representative pool of English trait terms than has been studied in the past, thereby providing more compelling evidence concerning the structure of the personality lexicon. For example, in the first of three studies, Goldberg (1990) used 1,431 trait adjectives grouped into 75 clusters to establish the across-method generality of trait factor structures. Virtually identical Big-Five representations emerged in 10 analyses, each based on a different factor-analytic procedure (5 methods for factor extraction, each rotated orthogonally and obliquely). In a second study of 479 quite common terms grouped into 133 synonym clusters, the Big-Five structure was obtained in each of two samples of self-ratings and each of two samples of peer ratings. None of the factors beyond the fifth generalized across the samples, thus establishing the generality of this structural representation across both targets and samples. In a third study, these synonym clusters were further refined by internal-consistency analyses, culminating in a set of 100 clusters derived from 339 trait terms.

The findings reported in Goldberg (1990) have stimulated the search for shorter and more easily administered markers of the Big-Five structure, for use in research contexts in which broad coverage of personality traits is needed but where subject time is at a premium. Because the sets of 133 and 100 synonym clusters were based on a representative sampling of English-language trait terms, they can be used as "criteria" for evaluating the convergent validity of the much shorter sets of factor markers presented in this article.

At the outset, however, it is important that readers distinguish between the criteria necessary to establish the utility of *factor markers*, which are intended solely as a means of locating other measures within a comprehensive structural representation, and *personality scales*, which are intended as measures of individual differences to be used for decision making in applied contexts. The primary criterion for the utility of factor markers is their ability to generate the target structure when the responses from large samples of individuals are factor analyzed. In contrast, the criteria that must be used to evaluate the utility of personality scales are far more diverse and demanding (American Psychological Association, 1985).

Strategies for Developing Factor Markers

The primary goal of the present studies is to discover a relatively small set of variables that will uniformly produce the Big-Five factor structure; such a marker set should include five reasonably homogeneous subsets of variables, each subset being roughly orthogonal to all the others. If such a factor-univocal set of marker variables were included in diverse studies, it would be possible to locate alternative theoretical positions and alternative personality measures within one comprehensive framework.

However, the search for factor-univocal variables to represent the natural language of personality traits is complicated by the fact that trait descriptors are not neatly clustered in multivariate space. That is, the lexicon of trait terms does not include only semantically isolated sets of near-synonyms and near-antonyms. Rather, most terms share some features of their meanings with one set of terms while they share other features with another set. Just as the stars appear to the naked eye in a clear sky, terms can be found nearly everywhere one looks in semantic space.

Also like the stars in the sky, however, there are some more densely concentrated regions, some semantic galaxies, within the widespread distribution of personality descriptors. Such semantic galaxies can be viewed as natural categories (e.g., Rosch, 1978) and thus as having a core of prototypical exemplars, plus a graded structure of increasingly more peripheral ones spreading out from the core; eventually one reaches exemplars that are as highly associated with another category as they are with the original one. Factor markers, then, might logically include those prototypical variables close to the category core and exclude those variables that fall in the interstitial regions between two or more categories.

In contrast to those who search for factor-univocal markers, those who work within the "circumplex" tradition (e.g., Kiesler, 1983; Wiggins, 1979; 1980) attempt to sample the semantic space in as uniform a fashion as possible. Working with a pair of dimensions, they project the locations of variables onto a circular representation and then select exemplars at equally spaced locations around the circle. Because all regions of semantic space are sampled uniformly, regions of dense concentration (i.e., the prototypical cores) are systematically undersampled, and more sparsely populated interstitial regions are oversampled.

In summary, then, one can contrast three general approaches to the selection of variables for developing factor markers: (a) *representative* sampling of the total domain, a strategy used for the development of Big-Five markers by Peabody (1987) and by Goldberg (1990); (b) *uniform* sampling, which implies oversampling sparsely populated regions and undersampling dense ones, as used by investigators in the circumplex tradition (e.g., Wiggins, 1979, 1980); and (c) *cluster* sampling, which aims for factor-univocal variables by the systematic omission of those located in interstitial regions between the clusters, as used by Norman (1963).

Whereas each of the three strategies is useful in particular contexts, cluster sampling has two inherent advantages for the purpose of developing factor markers. First, the same number of variables can be selected to mark each factor, thus reducing the probability of large differences in their reliabilities. In addition, because cluster sampling provides a simple-structured set of variables, markers based on this strategy can be expected to be more robust across different subject samples than those produced by either of the other strategies.

To develop an optimal set of cluster-based markers, however, one must use the other two strategies in the initial stages. First, one must map out the geography of the domain, for which one needs a reasonably representative sample of variables; this first step is necessary to locate the major clusters of variables. Once the major clusters have been identified, however, it is useful to adopt a circular perspective to discover the relative positions of variables located between pairs of clusters. In the final stage, one begins at the core of each cluster and selects variables equally from each side of the core. Ideally, to sample as broadly as possible without sacrificing fidelity, one would continue selecting away from the core, until one begins to merge too closely with adjoining clusters.

In a sense, then, the development of a set of factor markers is like the construction of a window. The two fundamental problems in constructing factor markers are (a) locating the center of each window and (b) establishing its width. Ideally, the center of each window should be located at the core of the cluster, and the windows should be as wide as possible without sacrificing factorial robustness across samples. There are two additional practical problems, (c) determining the number of variables to include in the total set and (d) determining the proportion of that set to be used to mark each factor. In practice, the total number of variables will be a compromise between length (thus increasing reliability) and brevity (thus decreasing subject testing time). If possible, however, it is desirable to include the same number of variables as markers of each of the factors, so as to measure each with roughly equal reliability.

Overview

Findings are presented from four studies of Big-Five marker sets, each set based on the cluster strategy of variable selection. In two studies of bipolar scales, the scales were grouped together by the factor with which they have been associated in past research and were administered, along with the factor label, in a highly "transparent" format. Self-ratings using this format were then compared with those using the more typical randomized or "opaque" arrangement of the same rating scales.

The implicit rationale for using bipolar rather than unipolar rating scales is to specify more precisely the dimension to be measured by the scale, rather than allowing one pole of the scale to be interpreted idiosyncratically (Goldberg & Kilkowski, 1985). Unfortunately, there is little in the way of evidence for the comparative utility of bipolar and unipolar variables when the two types are used to develop structural representations of personality-trait descriptors. One of the goals of the present studies is to compare the factor univocality of marker sets obtained from these two formats.

Accordingly, the polar terms that had been included in each of the bipolar scales were administered in unipolar format. In two studies, the factor univocality of the set of bipolar scales was compared with that of their unipolar equivalents. In a final study, a set of 100 unipolar adjectives, 20 marking each of the Big-Five factors, were shown to provide highly reliable and highly univocal markers in diverse samples of self- and peer descriptions. These new factor markers were compared with the synonym clusters reported in Goldberg (1990), as well as with the scales from the NEO and the Hogan (HPI; Hogan, 1986) personality inventories.

Study 1

Method

Sets of unipolar and bipolar variables were administered to the same sample of subjects, who used them for self-descriptions. In the bipolar set, there were 90 scales, including 50 potential Big-Five markers, 10 for each of the 5 factors. The bipolar scales were ordered such that those expected to be associated with the same factor were separated by scales from each of the other four factors, and one half of the more desirable poles were listed on each side of the page.

The set of single trait adjectives included the 180 polar terms from the 90 bipolar scales, followed by 12 additional trait adjectives that were not included in any of these analyses. The 180 terms were ordered in the same manner as were the 90 bipolar scales, all of the terms from the left poles followed by all of those from the right, so that the two polar terms in each antonym pair were always 90 items apart. Items from both inventories were administered with 9-step rating scales. The bipolar rating scale ranged from (1) *very* (Trait A) through (5) *neither* (Trait A) *nor* (Trait B) to (9) *very* (Trait B). The unipolar rating scale ranged from (1) *extremely inaccurate* as a self-description through (5) *neither accurate nor inaccurate* or *don't know* to (9) *extremely accurate* as a self-description.

The subjects were 157 college students enrolled in an introductory psychology course who elected to complete a battery of aptitude, attitude, and personality measures for extra course credit. The measures were administered under semi-anonymous conditions, with subjects using only numbers for identification. Instructions for both the unipolar and bipolar inventories stressed candidness of response and comparisons with others of the same sex and approximate age.

Results

Three sets of varimax-rotated factors were compared, specifically those derived from (a) the 100 single-trait adjectives, (b) the

50 difference scores formed by subtracting each polar term from its antonym, and (c) the 50 bipolar scales. The first two sets of variables were based on subjects' responses to the unipolar inventory, whereas the last was based on their responses to the bipolar scales.

Within each of the three marker sets, quite clear versions of the Big-Five factors were found, and nothing else of substance was uncovered when more than five factors were rotated. When the factor scores from each of the three analyses were correlated across the 157 subjects, the structure derived from (a) the 100 terms was found to be virtually identical to that derived from (b) the 50 difference scores; the correlations between the corresponding factor scores from these two data sets were .998, .988, .995, .987, and .991 for Factors I through V, respectively.

The varimax factor loadings for the 50 bipolar scales and the 50 difference scores are available from the author. The correlations across the 157 subjects between the corresponding factor scores from the two analyses were .86, .63, .79, .70, and .69 for Factors I through V, respectively. Moreover, it was the representation based on difference scores that more closely mirrored the structure that was initially hypothesized. Among the difference scores, 49 of the 50 variables (98%) had their highest loadings on the targeted factor, in contrast to 41 of the 50 bipolar scales (82%). Indeed, the difference scores produced higher mean factor loadings across the 10 markers of each of their hypothesized factors; averaged across the 5 factors, the mean factor loading was .58 for the difference scores and .51 for the bipolar scales.[1]

The findings from Study 1 suggest that it may be quite difficult to select a set of bipolar scales that are so univocal as factor markers that they are robust across procedural variations. The aim of Study 2 was to investigate a format for collecting ratings on bipolar scales that might prove to be more robust than that used in earlier studies.

Study 2

In Study 1, the bipolar scales were ordered quasirandomly, and the direction of the scales was counterbalanced so that only one half of the desirable poles were on the same side of the page. Indeed, as a control for response biases, such ordering and counterbalancing is almost always used in studies using bipolar scales. However, does the use of such controls reduce the clarity of the resultant factor structure? If the separation of variables measuring the same factor serves to decrease within-factor homogeneity, then grouping the variables by the factor with which they are associated may produce more highly simple-structured representations. On the other hand, the juxtaposition of conceptually related variables might lead subjects to increase their sensitivity to small differences among them, thereby attenuating the clarity of the resultant factor structures. Given the relative fragility of bipolar markers as compared with unipolar ones in previous studies, it seemed reasonable to investigate whether a more transparent format might lead to increased factor robustness.

The fairly extensive literature in personality assessment on item context and ordering effects does not include any studies focused directly on factor robustness. Although it is clear from that literature that responses to early items in a test (or to early tests in a battery) differ from responses to later ones (e.g., Hofstee, 1969; Knowles, 1988; Osberg, 1985), numerous studies have shown that mean scale scores do not differ significantly when the items are administered randomly throughout an inventory or all grouped together (e.g., Kestenbaum & Hammersla, 1976; Lieberman & Walters, 1968; Perkins & Goldberg, 1964; Strahan & Wilson, 1976). The effects of both grouping items together and labeling the scale are not as clear. Hamsher (1969) showed no significant effects of such scale transparency on mean scale scores, whereas other investigators have demonstrated increases in reliability and convergent validity, and corresponding decreases in discriminant validity, with transparent formats (e.g., Schriesheim, 1981a, 1981b; Schriesheim & Denisi, 1980; Solomon & Kopelman, 1984). More recently, however, Schriesheim, Solomon, and Kopelman (1989) reported findings from two studies in which the transparent format was superior on all counts. As they noted,

> Randomization of the order of questionnaire items is often assumed as necessary or desirable for measuring instruments, without fully considering other options or the possible consequences of randomization. . . . If respondents are trying to provide accurate answers, randomization makes the task more difficult and may therefore impair the validity of responses. This suggests that grouping and labeling items measuring the same constructs may be advantageous; such a format may break monotony and lessen fatigue effects, for example, by presenting respondents with many short questionnaire subsections as opposed to several long subsections. It may also help convince respondents that many items are not simply duplicates, designed to "catch" them in inconsistencies . . . (p. 19).

Method

Appendix B provides an example of the format used in these studies to collect self-ratings on blatantly transparent sets of bipolar rating scales. The reader will note that "birds of a feather have been flocked together," along with the name of that "flock." Not only were the scales grouped systematically rather than ordered randomly, but the more desirable pole of each scale was always on the right. Included in Study 2 were the same 50 factor markers used in Study 1, 10 scales marking each of the Big-Five factors; consequently, the factor structure based on the transparent format can be compared directly with that from the more typical format used in the previous study.

Of the students in a large undergraduate course in personality psychology, 215 completed the transparent bipolar scales at the beginning of a class period; they were asked to describe themselves as accurately as possible, and their confidentiality was assured by the use of identification numbers instead of names. In addition, 99 of these students completed Form S of the NEO-PI.

Results

In Study 1, 41 (82%) of the 50 bipolar scales had their highest loadings on the factor with which they are normally associated. In contrast, when administered with the new transparent format in the present study, every one of the 50 scales loaded as

[1] Averaged across the five factors, the mean intercorrelation among the items within each factor were .26 for the 20 unipolar terms, .32 for the 10 difference scores, and .29 for the 10 bipolar scales. The corresponding coefficient alpha reliability estimates were .87 for the 20 unipolar terms, .81 for the 10 difference scores, and .79 for the 10 bipolar scales.

hypothesized, and there was little variation in the size of the factor loadings among the 10 markers of each factor.[2] As evidence of convergent and discriminant validity, the coefficient alpha reliability estimates of the 10-item composite scales were .88, .88, .85, .88, and .84 for Factors I through V, respectively, averaging .87; the correlations among the composite scales ranged from .22 (Factors I vs. III) to .40 (Factors I vs. II), averaging .33. Correlations with the corresponding scales from the NEO–PI were .68, .54, .51, −.65, and .48 for Factors I through V, respectively. All of the correlations with noncorresponding NEO–PI scales did not differ significantly from zero, with one exception (a correlation of −.34 between the markers of Factor III and the Openness scale from the NEO–PI).[3]

The remarkable match between the hypothesized and obtained factor loadings for the transparent bipolar scales implies that this format may serve to direct subjects' attention to the common features of the scales related to each factor, rather than sensitizing them to the differences among scales within the same domain. However, because this is the first time such an effect has been discovered, it is necessary to try to replicate it. That was the aim of the next study.

Study 3

Given the encouraging findings from the 50 bipolar scales in Study 2, Study 3 was designed to compare the new transparent ordering of the scales with the traditional opaque ordering, using the same subjects. In addition, this study was designed to compare the Big-Five factor markers derived from both types of bipolar scales with (a) those derived from two sets of synonym clusters (Goldberg, 1990), (b) the factors from a new set of 100 unipolar variables, and (c) the scales included in the NEO and the Hogan personality inventories.

Method

A few of the bipolar scales that had been included as markers of Factor I (Surgency) and Factor IV (Emotional Stability) in Study 2 were factorially more complex than might be desirable. In an effort to provide more univocal factor markers, five new scales were added to the original 50, two as potential markers of Factor I and three as potential markers of Factor IV. The resulting 55 bipolar rating scales (55–BRS) were assembled in two formats: (a) The transparent format was identical to that used in Study 2, except that the new scales were listed at the end of the original Factor I and Factor IV scales; (b) Four versions of the traditional opaque format were developed, in each of which the more desirable pole alternated scale-by-scale from left to right. There were two quasirandom orders, with all of the scales reflected in one of the two versions of each order.

The subjects in this study were university undergraduates enrolled in an introductory course in personality psychology who elected to complete a few short questionnaires in class and a few longer inventories at home. Subjects were assured of confidentiality, and they used only numbers for identification. At the beginning of an early class period, the subjects completed one of the four opaque versions of the 55-BRS; two days later, at the beginning of the next class session, they completed the 55-BRS in its transparent format. Of the 200 students in the course, 175 completed both the transparent and one of the opaque versions of the 55-BRS. In addition, most subjects completed the inventory of 479 trait adjectives (Goldberg, 1990), the NEO–PI, and the Hogan Personality Inventory.

Results

The 55 bipolar scales. As a test of the cross-sample robustness of the transparent format, the factor loadings were compared from Study 2 to Study 3 for those 50 scales that were included in both studies. Whereas in Study 2 all 50 (100%) of the scales had their highest loading on the factor with which they are normally associated, in Study 3 this was reduced to 46 (92%). Nonetheless, the factor structures based on these 50 common scales were quite similar, with most of the variables functioning as relatively univocal factor markers in both studies.[4]

Of greater importance is the comparison between the findings from the transparent and the opaque formats in the same sample of subjects. Table 1 compares these factor structures, based on the 10 most univocal markers of each factor among the 55 scales included in Study 3. In the transparent format, 48 (96%) of the 50 scales had their highest loadings on the targeted factor, whereas in the opaque format, 46 (92%) did so. The correlations between the factor scores from the two types of format were slightly higher for Factor I (.81), Factor II (.79), and Factor III (.81) than for Factor IV (.77) and Factor V (.74).

Another way to compare the two types of formats is to examine the convergent and discriminant properties of the sets of scales marking each of the five factors, after reflecting scales in the opaque format so that they are all oriented in the favorable direction. Across all 55 variables, the transparent format elicited mean self-ratings that were significantly ($p < .001$) more favorable than those elicited by the opaque format. Because self-ratings tend to be highly favorable to begin with, the increase in mean favorability led to a ceiling effect, as indicated by a significant ($p < .001$) decrease in the variances of the ratings across the subjects. Both of these effects were relatively constant across the 55 scales: The correlations of the 55 means between the two formats was .98, and the correlation between the two sets of standard deviations was .91.

There was a substantial difference between the two formats in their average within-set homogeneity; the mean item intercorrelation was .35 in the transparent formats, compared with .27 in the opaque formats. This difference, however, varied greatly by domains: The greater homogeneity of the transparent format was particularly pronounced for Factor II (.42 versus .23), whereas it was actually reversed for Factor V (.25 versus .27). For each of the five factors the respective coefficient alpha reliability estimates were .87, .87, .84, .88, and .76 (averaging .84) in the transparent format versus .81, .72, .78, .85, and .79 (averaging .79) in the opaque format.

Finally, as an index of discriminant validity, subjects' mean responses to the scales in each of the five marker sets were correlated across the 175 subjects. For the transparent format, these interdomain correlations ranged from .13 (Factors I vs. III) to .42 (Factors II vs. IV), averaging .27. For the opaque format, they ranged from .01 (Factors I vs. III) to .36 (Factors I

---

[2] A table presenting the factor loadings is available from the author.

[3] Correlations with the NEO–PI facet scales are available from the author.

[4] A table comparing the factor loadings in the two studies is available from the author.

Table 1
A Comparison of the Factor Loadings for 50 Bipolar Scales When Administered in Transparent (T) and Opaque (O) Orders (N = 175)

	I		II		III		IV		V	
Scale	T	O	T	O	T	O	T	O	T	O
					Factor I. Surgency					
Introverted–extraverted	**.77***	**.67***	.01	−.00	.08	.02	.12	.01	−.01	.08
Unenergetic–energetic	**.65***	**.59***	.27	**.30**	−.01	.17	.13	.02	.05	.06
Silent–talkative	**.65***	**.64***	−.05	−.00	.09	−.05	−.07	−.15	.04	.01
Unenthusiastic–enthusiastic	**.65***	**.64***	.26	.29	.10	.23	.11	.08	.15	.04
Timid–bold	**.63***	**.64***	−.18	−.25	.07	−.00	.20	.19	.18	.16
Inactive–active	**.64***	**.54***	.12	.11	.06	.10	.12	.16	.02	.07
Inhibited–spontaneous	**.47***	**.49***	.17	.13	−.20	−.34	.26	.25	.27	.09
Unassertive–assertive	**.54***	**.59***	−.18	−.03	.29	.21	.27	.26	.28	.22
Unadventurous–adventurous	**.49***	**.50***	.05	−.00	−.05	−.02	−.02	.10	.18	.14
Unsociable–sociable	**.72***	**.42***	.22	.23	−.06	−.06	.13	.07	.17	.22
					Factor II. Agreeableness					
Cold–warm	**.31**	**.31**	**.66***	**.59***	.02	.07	.09	.06	.00	−.00
Unkind–kind	.13	.06	**.74***	**.50***	.10	.12	.08	−.13	.11	.19
Uncooperative–cooperative	−.02	−.03	**.63***	**.63***	**.31**	.22	−.03	.12	−.08	.04
Selfish–unselfish	−.10	−.05	**.62***	**.46***	.11	−.08	.28	.29	.15	−.10
Rude–polite	−.03	−.06	**.63***	**.59***	.26	.14	.17	−.04	.11	.16
Disagreeable–agreeable	.02	.12	**.75***	**.37***	−.18	−.06	.01	−.03	−.13	−.02
Distrustful–trustful	.25	.03	**.41***	**.49***	.12	.14	.17	.03	**−.38**	**−.35**
Stingy–generous	.11	.12	**.62***	**.53***	.02	−.08	.20	.02	.12	−.02
Inflexible–flexible	.07	.02	**.65***	**.39***	.10	−.13	.06	.10	−.05	.05
Unfair–fair	.10	.01	**.55***	**.47***	**.31**	**.34**	.24	.10	.08	.16
					Factor III. Conscientiousness					
Disorganized–organized	.19	.16	.06	−.04	**.69***	**.68***	−.12	−.08	−.02	−.03
Irresponsible–responsible	.19	−.02	.11	−.01	**.79***	**.70***	−.07	.18	−.04	−.03
Undependable–reliable	.19	.11	.18	.11	**.69***	**.57***	−.07	.12	.07	.09
Negligent–conscientious	.14	−.08	.25	.02	**.68***	**.60***	.01	.05	.20	**.38**
Impractical–practical	−.11	−.07	.04	.19	**.66***	**.44***	.21	.04	.11	.14
Careless–thorough	.05	.13	.06	−.08	**.74***	**.66***	−.00	.03	.15	.08
Lazy–hardworking	.25	.22	**.31**	.11	**.49***	**.53***	.20	.04	−.00	−.13
Extravagant–thrifty	−.19	−.23*	.02	−.05	**.37***	.12	.29	.17	.00	.00
Rash–cautious	−.24	**−.31**	.05	.24	**.63***	**.49***	.18	.00	.12	.10
Frivolous–serious	−.29	−.27	−.13	−.19	**.49***	**.50***	.12	−.07	.12	.05
					Factor IV. Emotional Stability					
Angry–calm	.13	.09	**.32**	**.47***	−.04	−.04	**.58***	**.33**	−.19	−.01
Tense–relaxed	.22	.26	.24	.25	−.13	−.14	**.70***	**.57***	−.04	−.08
Nervous–at ease	.23	**.40**	.12	.07	−.05	−.08	**.71***	**.61***	.04	.02
Envious–not envious	−.01	−.03	.19	.03	.07	.03	**.53***	**.54***	.22	.09
Unstable–stable	.17	.11	.22	.15	**.33**	**.31**	**.56***	**.48***	−.04	−.08
Discontented–contented	.28	−.07	.24	**.32**	.19	.17	**.57***	**.47***	.02	.09
Insecure–secure	.20	.28	−.03	.03	.25	.22	**.65***	**.66***	.24	.13
Emotional–unemotional	−.15	−.17	−.09	−.10	−.03	−.12	**.50***	**.54***	−.17	.02
Guilt-ridden–guilt-free	.04	.20	.01	−.01	.10	.10	**.57***	**.68***	.16	.03
Moody–steady	.13	.24	.26	**.38***	.02	.08	**.65***	**.38**	−.25	−.22
					Factor V. Intellect					
Unintelligent–intelligent	.06	.09	−.09	−.23	.22	.19	−.06	−.02	**.46***	**.53***
Imperceptive–perceptive	.20	.13	−.04	−.03	.19	.08	.04	.19	**.59***	**.39***
Unanalytical–analytical	−.00	.05	−.25	−.20	.19	.10	−.03	−.04	**.36***	**.38***
Unreflective–reflective	.14	.02	−.03	.18	.11	−.04	−.09	−.08	**.56***	**.54***
Uninquisitive–curious	**.36**	**.35**	.09	.13	−.02	−.02	.05	−.15	**.51***	**.50***
Unimaginative–imaginative	.13	.12	.19	.06	−.03	.06	.10	.06	**.75***	**.68***
Uncreative–creative	.14	.17	.20	.10	.01	−.00	.01	−.02	**.76***	**.65***
Uncultured–cultured	.01	−.11	.16	.25	**.32***	.10	.09	.16	.29	**.40***
Unrefined–refined	.03	−.13	**.32**	.25	**.33***	.09	.12	**.34***	.21	.24
Unsophisticated–sophisticated	.09	.13	.22	.02	.17	−.02	.12	.16	.25*	**.43***

Note. Values equal to or larger than |.30| are listed in boldface type.
* Highest factor loading of each scale.

vs. IV), averaging .20. In summary, then, it is obvious that the transparent format elicits substantially increased scale intercorrelations, both within and across the domains. However, it is the increased within-domain homogeneity of the transparent format that must provide its slight advantage in factor univocality over the opaque format.

A new set of 50 difference scores. The findings from Study 1 suggested that it may be easier to select relatively univocal factor markers from trait descriptors administered in unipolar format than from bipolar scales, and one of the aims of Study 3 was to replicate this finding. Most of the subjects in this sample had completed the inventory of 479 trait-descriptive adjectives (Goldberg, 1990). To eliminate any individual differences in subjects' use of the rating scale when responding to this relatively long inventory, the responses of each subject were standard (*Z*) scored across the 479 items. From this item pool, 10 antonym pairs (20 single terms) were selected as markers of each of the Big-Five domains; 60 of the 100 terms were the same as those included in Study 1. Four analyses will be compared, contrasting the 100 single terms and the 50 difference scores, each based on the original (raw) and on the standardized (*Z*) item responses.

In each of these four analyses, the Big-Five factors emerged quite distinctly. Across 192 subjects, the correlations between the factor scores based on the 100 single terms and those based on the 50 difference scores were .98, .97, .98, .95, and .97 using the original responses and .98, .98, .99, .98, and .98 using the standardized responses, for Factors I to V, respectively. The virtually perfect similarity of the factor structures derived from single terms as compared with difference scores has now been confirmed in two studies, each using a somewhat different set of trait descriptors.

The Big-Five factors derived from the 50 difference scores, based on both the original and standardized responses, are presented in Table 2. The correlations across the 192 subjects between the factor scores from these two analyses were .97, .93, .95, .97, and .91 for Factors I to V, respectively, indicating that the two structures are virtually the same. Moreover, the very slight differences between them suggest the relative superiority of the structure based on standardized responses. For example, using the original responses, 48 (96%) of the 50 variables had their highest loading on the factor that they were expected to mark, whereas this value was 100% using the standardized responses.

Putting It All Together: An Improved Set of Unipolar Markers

The cumulative pattern of findings across these three studies suggests that relatively small sets of variables can serve as markers of the Big-Five factor structure. Moreover, the findings also suggest that variables administered in unipolar format appear to be more robust across samples than are bipolar scales. Building on these findings, a final study was undertaken to develop a particularly robust set of unipolar factor markers, to test its generality across more diverse data sets than those that had been used in the earlier studies, and to examine its convergence with other factor markers, including those described in Goldberg (1990) and the scales from the NEO and Hogan personality inventories.

To the extent possible, the new factor markers were designed to possess three major properties: (a) They should be at least as robust as the transparent bipolar scales developed in the previous studies, without capitalizing on the halo-inducing response biases that could be associated with those transparent scales; (b) they should show substantial concordance with Big-Five markers based on much larger and more representative sets of variables; and (c) each of the five factors elicited by these markers should be of roughly similar size.

Study 4

The Selection of New Variables

To avoid large differences in reliability between markers of the five factors, it is reasonable to select the same number of variables for each of them. Two decades of basic research on the structure of personality-trait adjectives have demonstrated, however, that the Big-Five factors are represented quite disproportionately in the English lexicon (e.g., Peabody, 1987; Goldberg, 1990). Specifically, there are more adjectives associated with Factor II (Agreeableness) than with any of the others, and there are substantially more adjectives associated with that factor, as well as with Factor I (Surgency) and Factor III (Conscientiousness), than there are with either Factor V (Intellect) or Factor IV (Emotional Stability). Moreover, whereas there are roughly similar frequencies of adjectives associated with each of the two poles of four of the factors, there are extremely few terms marking the desirable pole of Factor IV.

Indeed, in contrast with adjectives in general and trait adjectives in particular, terms describing symptoms of emotional instability (e.g., *anxious, high-strung, moody, nervous, temperamental, tense, touchy*) have no direct antonyms. For this reason, it is quite difficult to find more than a half-dozen commonly used adjectives that are univocal markers of the positive pole of Factor IV, and if one were to insist on equal numbers of markers of each pole of the five factors, one would be limited to 50 or 60 variables. To increase marker reliability, it may be necessary to relax the limitation on equal representation of the two poles of each factor, if only for Factor IV.

Moreover, in each of the preceding studies in this series the unipolar variables included only antonym pairs, specifically the two terms included in each bipolar scale. However, when antonyms are presented in unipolar format, they are not always treated as opposites in subjects' self-descriptions (Goldberg & Kilkowski, 1985). Indeed, the unipolar terms in some antonym pairs may even be associated with different factors. For example, whereas subjects have no difficulty describing targets using the bipolar contrast *moody–steady*, in unipolar format the first term is highly associated with the negative pole of Factor IV (Emotional Instability), and the second term is associated with the positive pole of Factor III (Conscientiousness).

Both of these peculiarities of the personality-trait lexicon were taken into account in the development of a new set of 100 factor markers, 10 for each pole of all factors other than Factor IV, which was marked by 6 positive and 14 negative terms. All 100 terms had been included in the inventory of 566 common trait adjectives, described in Goldberg (1982). Factor analyses of various subsets of these terms served to establish the overall

Table 2
The Big-Five Factors Derived From Difference Scores Based on Self-Ratings of Single Terms: Original (Raw) Versus Standardized (Z) Responses (N = 192)

	I		II		III		IV		V	
Terms	Raw	Z	Raw	Z	Raw	Z	Raw	Z	Raw	Z
Factor I. Surgency										
Silent–talkative	**.67***	**.65***	.09	.06	−.05	−.05	−.27	−.25	−.03	−.10
Timid–bold	**.64***	**.64***	−.02	−.06	.09	−.02	.14	.10	.13	.15
Compliant–assertive	**.59***	**.56***	−.18	−.22	.10	.05	.10	.04	.09	.14
Inhibited–spontaneous	**.56***	**.54***	.05	−.02	−.16	−.22	.28	.26	.25	.26
Passive–active	**.62***	**.63***	.06	−.06	.16	.06	.10	.05	.19	.09
Reserved–demonstrative	**.59***	**.53***	−.10	−.11	−.17	−.20	−.13	−.12	.01	.01
Lethargic–energetic	**.58***	**.60***	.23	.16	.24	.19	.16	.13	.13	.06
Apathetic–enthusiastic	**.41***	**.41***	.20	.10	**.30**	.24	.11	.09	.17	.09
Unadventurous–adventurous	**.44***	**.40***	.00	−.01	−.15	−.23	.14	.14	**.31**	.23
Unsociable–sociable	**.58***	**.57***	**.47**	**.43**	.11	.07	−.18	−.16	.12	.07
Factor II. Agreeableness										
Cold–warm	.24	.23	**.68***	**.67***	.06	−.04	.08	.10	.28	.20
Unkind–kind	.08	.01	**.62***	**.64***	.27	.15	.03	.03	**.45**	**.36**
Uncooperative–cooperative	−.08	−.16	**.79***	**.74***	.25	.17	−.02	−.05	−.07	−.14
Selfish–unselfish	−.03	−.05	**.51***	**.47***	.15	.14	**.33**	**.34**	−.01	.02
Rude–polite	.05	−.01	**.68***	**.64***	.29	.18	.05	.02	.23	.10
Distrustful–trustful	.18	.16	**.55***	**.50***	.15	.12	−.04	−.05	−.23	−.26
Stingy–generous	.02	−.03	**.59***	**.54***	−.01	−.13	.27	.26	.22	.16
Stubborn–flexible	−.01	−.02	**.47***	**.43***	−.23	−.21	.23	.27	−.01	−.01
Inconsiderate–considerate	.07	−.05	**.58***	**.59***	**.35**	.20	.08	.01	**.44**	**.38**
Quarrelsome–agreeable	−.12	−.13	**.65***	**.55***	−.07	−.17	.28	**.32**	.06	−.01
Factor III. Conscientiousness										
Disorganized–organized	.24	.24	.01	−.04	**.66***	**.62***	−.14	−.18	.03	−.04
Undependable–dependable	.25	.15	.29	.21	**.63***	**.57***	−.11	−.21	.05	−.08
Unconscientious–conscientious	.08	.02	.19	.09	**.68***	**.64***	.14	.12	.25	.19
Impractical–practical	.03	−.04	.05	.00	**.65***	**.64***	.17	.15	.15	.20
Careless–thorough	.12	.14	.15	.07	**.74***	**.69***	−.06	−.07	.10	−.06
Extravagant–thrifty	−.29	−.29	−.07	−.03	**.53***	**.53***	.15	.11	−.03	−.02
Rash–cautious	−.20	−.24	.29	.17	**.49***	**.42***	.05	.04	.15	.09
Frivolous–serious	−.27	−.26	−.09	−.17	**.60***	**.57***	.07	.06	.14	.06
Wasteful–economical	.00	−.05	−.01	−.08	**.66***	**.63***	.20	.21	.08	.10
Unreliable–reliable	.18	.11	.27	.23	**.62***	**.58***	−.02	−.08	.15	.04
Factor IV. Emotional Stability										
Emotional–unemotional	−.12	−.10	−.25	−.20	−.17	−.07	**.41**	**.43***	**−.42***	**−.35**
Envious–unenvious	.10	.09	.20	.15	−.04	−.01	**.58***	**.60***	.09	.08
Nervous–relaxed	**.32**	**.33**	.21	.12	−.04	−.07	**.60***	**.58***	.01	.02
Subjective–objective	.04	.00	.12	.14	.11	.10	**.40***	**.36***	.00	.03
High-strung–imperturbable	−.14	−.16	.02	−.01	.05	.00	**.57***	**.54***	−.03	−.07
Demanding–undemanding	−.12	−.15	.28	.25	−.26	−.20	**.49***	**.54***	−.18	−.16
Fretful–placid	−.03	.05	.16	.10	.10	.08	**.54***	**.53***	−.06	−.05
Volatile–peaceful	−.06	−.10	**.36**	**.30**	.15	.10	**.53***	**.52***	.09	.05
Suggestible–independent	.07	.09	−.15	−.25	.19	.06	**.53***	**.41***	.26	.20
Fearful–uninhibited	.29	.28	−.13	−.16	.01	−.02	**.58***	**.53***	.09	.14
Factor V. Intellect										
Unintelligent–intelligent	.09	.01	−.02	.00	.24	.12	.09	.04	**.64***	**.62***
Imperceptive–perceptive	.16	.09	.14	.09	.19	.13	**.33**	**.31**	**.46***	**.45***
Uninquisitive–curious	.18	.13	.07	.04	.14	.00	−.08	−.17	**.62***	**.53***
Unimaginative–imaginative	.10	.10	.10	.09	−.07	−.11	.05	.01	**.70***	**.63***
Uncreative–creative	.17	.14	.06	.06	−.04	−.02	.07	.04	**.63***	**.60***
Unsophisticated–sophisticated	.17	.17	.07	.05	.05	.05	−.13	−.09	**.32***	**.36***
Ignorant–knowledgeable	.19	.10	.10	.07	.16	.11	.26*	.22	.25	.28*
Unintellectual–intellectual	.04	−.01	−.12	−.09	.20	.13	−.06	−.11	**.62***	**.65***
Shallow–deep	−.10	−.15	.06	−.02	.15	.06	.05	.07	**.61***	**.55***
Provincial–cultured	.08	.12	.13	.08	.06	.00	.06	.09	**.45***	**.35***

Note. Values equal to or larger than |.30| are listed in boldface type.
* Highest factor loading of each variable.

Table 3

The Big-Five Factors Derived From the 100 Markers in Study 4: Varimax-Rotated Factor Loadings in Self-Descriptions and Descriptions of Liked Peers (Z-Scored Responses)

Term	I		II		III		IV		V	
	Peers	Self	Peers	Self	Peers	Self	Peers	Self	Peers	Self
Factor I. Surgency										
I+										
Extraverted	.60*	.68*	.10	.08	−.05	−.09	.06	−.02	−.02	.03
Talkative	.65*	.67*	.20	.13	−.21	−.10	−.18	−.19	−.12	−.02
Assertive	.58*	.51*	−.13	−.12	.09	.16	.08	.17	.11	.08
Verbal	.50*	.60*	.04	−.06	−.10	−.11	−.10	−.05	.00	.07
Energetic	.45*	.45*	.18	.25	.21	.12	.12	.14	.04	.00
Bold	.60*	.48*	−.10	−.11	.02	.00	.21	.24	.01	.04
Active	.43*	.42*	.10	.26	.27	.12	.05	.14	.08	−.10
Daring	.43*	.45*	−.04	−.03	.00	.02	.11	.12	.15	−.02
Vigorous	.38*	.39*	.06	.08	.07	.13	.12	.12	.08	−.05
Unrestrained	.54*	.45*	.16	.10	−.13	−.06	.05	.01	.01	.02
I−										
Introverted	−.70*	−.69*	−.02	−.07	.04	.03	−.08	.00	.01	.03
Shy	−.73*	−.73*	.12	.11	.04	.04	.04	−.04	.07	−.10
Quiet	−.68*	−.71*	.00	.00	.16	.13	.18	.11	.14	.02
Reserved	−.66*	−.64*	.02	.04	.23	.17	.15	.10	.04	−.04
Untalkative	−.68*	−.68*	−.20	−.06	.17	.09	.20	.20	.03	−.01
Inhibited	−.50*	−.57*	.02	−.07	.04	.04	−.15	−.11	−.17	−.07
Withdrawn	−.68*	−.65*	−.14	−.12	−.03	−.09	−.01	−.04	−.01	.03
Timid	−.69*	−.59*	.28	.17	−.02	−.01	−.05	−.18	−.09	−.20
Bashful	−.65*	−.67*	.20	.15	.05	.04	.01	−.06	−.01	−.06
Unadventurous	−.42*	−.34*	.11	−.13	.03	−.01	−.02	−.15	−.28	−.15
Factor II. Agreeableness										
II+										
Kind	.02	−.06	.66*	.63*	.14	.11	−.12	−.02	−.01	.03
Cooperative	−.12	−.05	.52*	.54*	.26	.16	.21	.10	−.01	−.09
Sympathetic	−.08	−.05	.66*	.65*	−.01	−.07	−.08	−.11	.09	−.07
Warm	.16	.17	.63*	.55*	.09	.06	−.10	−.05	.03	−.01
Trustful	.01	.00	.55*	.49*	.20	.11	.19	.23	−.05	−.11
Considerate	−.16	−.04	.56*	.49*	.17	.14	.15	.05	.07	.02
Pleasant	.06	.04	.58*	.47*	.04	.07	.09	.19	−.12	.00
Agreeable	−.13	−.20	.46*	.45*	.00	.04	.18	.03	−.02	.03
Helpful	.06	.02	.57*	.52*	.22	.23	.15	.04	.14	.07
Generous	.08	.09	.47*	.38*	.00	.02	.14	.13	.08	−.02
II−										
Cold	−.15	−.27	−.63*	−.58*	.09	−.03	−.01	.04	−.08	.07
Unkind	−.03	−.10	−.52*	−.54*	−.09	−.06	.13	.08	−.12	−.17
Unsympathetic	−.03	−.10	−.56*	−.61*	−.03	.05	.21	.02	−.19	−.07
Distrustful	−.08	−.19	−.39*	−.43*	−.21	−.15	−.07	−.15	.07	.12
Harsh	.13	.08	−.48*	−.48*	.02	.01	−.20	.00	−.14	−.01
Demanding	.30	.24	−.33	−.39*	.07	.12	−.36*	−.15	−.02	.05
Rude	.24	.04	−.55*	−.38*	−.11	−.21	−.07	.08	−.11	.05
Selfish	.02	−.17	−.41*	−.38*	−.16	−.11	−.22	−.21	.05	.16
Uncooperative	−.01	.04	−.36*	−.40*	−.17	−.19	−.11	−.01	−.08	.01
Uncharitable	−.17	−.04	−.39*	−.46*	−.05	.07	.00	−.07	−.16	−.11
Factor III. Conscientiousness										
III+										
Organized	−.06	−.06	.00	−.04	.75*	.80*	−.04	−.02	−.05	−.02
Systematic	−.10	−.06	.00	−.04	.61*	.64*	.10	.14	.03	−.02
Thorough	−.07	−.05	−.08	.04	.57*	.61*	.20	.21	.10	.06
Practical	−.05	.00	.15	.12	.57*	.48*	.19	.09	−.01	−.09
Neat	−.14	−.04	.10	.13	.64*	.71*	−.14	−.10	−.06	−.01
Efficient	.07	−.03	−.04	.17	.65*	.59*	.04	.02	−.01	.06
Careful	−.22	−.17	.20	.12	.49*	.43*	−.03	.03	.09	−.08
Steady	−.02	−.04	.11	.19	.48*	.38*	.40	.27	.02	−.19
Conscientious	−.01	−.19	.26	.14	.29*	.28*	.07	.09	.16	−.02
Prompt	.00	−.02	.11	.07	.57*	.41*	.06	−.08	−.03	−.10

Table 3 (*continued*)

	I		II		III		IV		V	
Term	Peers	Self	Peers	Self	Peers	Self	Peers	Self	Peers	Self
Factor III. Conscientiousness (*continued*)										
III−										
Disorganized	.05	−.02	.03	.04	−.77*	−.76*	.03	.07	.01	.02
Careless	.09	.04	−.14	.01	−.65*	−.55*	−.07	−.04	−.03	−.08
Unsystematic	−.06	.00	.02	.10	−.62*	−.58*	.01	−.02	.01	−.11
Inefficient	−.17	−.17	−.05	−.10	−.63*	−.57*	.01	.08	−.04	−.07
Undependable	−.02	.00	−.17	−.24	−.44*	−.42*	−.07	.11	.07	.01
Impractical	−.08	−.05	−.18	−.02	−.53*	−.42*	−.04	−.12	−.07	.02
Negligent	−.15	−.01	−.24	−.02	−.49*	−.47*	−.07	−.10	−.09	−.07
Inconsistent	.04	−.03	−.09	−.16	−.50*	−.53*	−.23	−.18	−.15	−.02
Haphazard	.13	−.03	−.12	−.04	−.52*	−.51*	−.07	−.10	−.17	−.19
Sloppy	.11	−.08	−.03	−.08	−.59*	−.56*	.17	.11	.09	−.03
Factor IV. Emotional Stability										
IV+										
Unenvious	−.02	−.02	.09	.10	.05	−.02	.55*	.60*	−.05	.13
Unemotional	−.23	−.21	−.41	−.37	.05	.14	.45*	.47*	−.06	−.14
Relaxed	.05	.18	.26	.18	.00	14	.49*	.47*	−.10	−.11
Imperturbable	−.18	−.03	.11	.12	.08	.01	.33*	.42*	−.17	−.09
Unexcitable	−.33	−.27	−.29	−.24	.04	−.04	.46*	.40*	−.05	−.14
Undemanding	−.27	−.29	.32	.20	−.05	−.12	.43*	.32*	−.16	−.18
IV−										
Anxious	−.01	.07	.00	.00	.03	.05	−.39*	−.43*	−.09	−.04
Moody	−.12	−.11	−.23	−.09	−.07	−.07	−.59*	−.57*	.10	.09
Temperamental	.00	−.02	−.35	−.13	−.02	−.07	−.51*	−.56*	−.04	.01
Envious	−.13	−.09	−.01	−.08	−.10	−.04	−.44*	−.56*	−.03	−.18
Emotional	.10	.08	.45*	.33	−.10	−.09	−.44*	−.52*	.08	.04
Irritable	.05	−.02	−.44*	−.22	−.04	.01	−.40*	−.53*	−.07	−.02
Fretful	−.19	−.17	.07	−.04	−.06	−.08	−.46*	−.51*	−.05	−.10
Jealous	.00	−.09	−.06	−.10	−.09	−.02	−.53*	−.53*	−.11	−.14
Touchy	−.18	−.07	−.16	.09	.00	.04	−.48*	−.55*	−.12	.03
Nervous	−.19	−.23	.04	.02	.02	−.11	−.39*	−.47*	−.02	−.06
Insecure	−.42	−.34	.02	.02	−.25	−.22	−.46*	−.40*	.00	−.01
Fearful	−.26	−.24	.07	.14	−.12	.00	−.32*	−.37*	−.08	−.16
Self-pitying	−.36	−.33	−.14	−.04	−.14	−.05	−.43*	−.45*	−.08	−.01
High-strung	.18	.16	−.08	−.07	−.01	−.06	−.48*	−.35*	−.02	.06
Factor V. Intellect										
V+										
Intellectual	−.09	.00	.00	−.06	.18	.08	.22	.09	.49*	.59*
Creative	.01	−.04	.07	.07	−.06	.09	.13	.11	.57*	.69*
Complex	−.12	−.06	−.10	−.05	−.06	−.15	−.06	−.11	.51*	.50*
Imaginative	.08	.09	.01	.06	−.02	.00	.12	.03	.52*	.57*
Bright	.05	.01	.16	.03	.11	.00	.18	−.01	.41*	.45*
Philosophical	−.07	−.05	.05	−.01	.01	−.10	.12	.06	.53*	.52*
Artistic	−.12	−.07	.06	.02	.03	.11	.01	.08	.52*	.50*
Deep	−.06	−.12	.18	.19	−.04	−.14	.05	.07	.50*	.38*
Innovative	.12	.09	.05	−.08	.10	.10	.24	.17	.47*	.40*
Introspective	−.21	−.28	.06	−.03	−.07	−.10	−.01	.07	.34*	.34*
V−										
Unintellectual	−.01	−.06	.02	.08	−.16	−.06	.01	.07	−.47*	−.56*
Unintelligent	.01	−.02	−.10	.05	−.08	−.06	−.01	.01	−.47*	−.51*
Unimaginative	−.23	−.12	−.05	−.05	.02	−.09	.09	.02	−.56*	−.60*
Uncreative	−.12	−.06	.01	.01	.05	−.08	.09	−.03	−.53*	−.66*
Simple	−.18	−.13	.12	.21	.03	.10	.12	.08	−.33*	−.46*
Unsophisticated	−.17	−.15	−.05	.07	−.25	−.16	.25	.21	−.34*	−.32*
Unreflective	−.02	.07	−.08	−.04	.06	.04	.16	.01	−.39*	−.41*
Imperceptive	−.14	−.17	−.02	.10	−.06	−.04	.08	.10	−.41*	−.33*
Uninquisitive	−.27	−.14	−.10	−.02	−.03	.03	.09	.15	−.32*	−.32*
Shallow	−.18	−.09	−.21	−.09	−.05	.03	−.03	−.06	−.39*	−.37*

Note. Values equal to or larger than |.30| are listed in boldface type.
* Highest factor loading of each variable.

topography of this representative set of trait adjectives. To optimize factor robustness across diverse data sets, the marker subsets were deliberately selected to be somewhat less broad than those that had been used in the other studies in this series, and only about one half of the variables were direct antonym pairs. Of the 100 terms, 46 had been included in the unipolar set used in Study 1, and 54 had been included in the corresponding unipolar set used in Study 3.

The Six Data Sets

If a set of factor markers is to be most useful, it should elicit the same factors in self-descriptions as in descriptions of others. Peabody and Goldberg (1989), however, have shown that one major determinant of the size and nature of the factors derived from trait adjectives is the evaluative homogeneity of the targets being described. Specifically, factors based on samples that include both liked and disliked targets differ systematically from those based on samples of restricted evaluative range (e.g., self-descriptions, descriptions of one's friends). Consequently, the most difficult test of factor robustness is a comparison between self-descriptions (the most restricted case) and descriptions of evaluatively heterogeneous targets (the least restricted case). In the present study, the robustness of the 100 new Big-Five markers was tested in three samples: (a) self-descriptions, (b) descriptions of friends, and (c) pooled descriptions of liked, neutral, and disliked peers.

The Self sample included 320 college students who described themselves on an inventory of 587 trait adjectives (Goldberg, 1982), using the response format that was described in Studies 3 and 4 of Goldberg (1981); four middle response options were provided (*average or neutral, it depends on the situation, don't know, term unclear or ambiguous*), all of which were here given a midscale value of 0 on a rating scale that ranged from −3 to +3.

The Liked sample included 316 of the 320 subjects from the Self sample who used the same inventory of 587 terms to describe someone of their sex and approximate age whom they knew well and liked, using the same response format.

The Pooled sample included 205 students in law school and in an upper-division psychology course, roughly one-third of whom were randomly assigned to describe one of three types of peer targets: (a) "Someone whom you know well and *like* as a person" [$n = 76$], (b) "someone whom you know well but *neither like nor dislike*" [$n = 69$], or (c) "someone whom you know well and *dislike* as a person" [$n = 60$]. The targets were further specified in all three conditions to be "of the same sex as you are and about your own age." The subjects used an inventory of 566 trait adjectives (a subset of the 587), with an 8-step rating scale that ranged from *extremely inaccurate* to *extremely accurate;* later, all omitted responses were given a middle value (5) on a transformed 1 to 9 rating scale. Ratings of the three types of targets were pooled in all analyses of this sample.

To eliminate individual differences in subjects' use of the rating scales, the responses of each subject were Z scored on the basis of the 566 common terms in the two inventories. Factor analyses of the 100 potential markers were carried out within each of the three samples, separately for both the original (or "raw") and the standard-scored responses. Thus, there are six data sets available for comparison: the original and standard-scored responses from each of the three samples. These six data sets can be ordered along a continuum of evaluative homogeneity, from the most homogeneous (standard-scored responses derived from self-descriptions) to the least homogeneous (original responses derived from pooled liked, neutral, and disliked targets), with the two data sets derived from descriptions of liked peers falling between the two extremes. Specifically, the six data sets fall in the logical order: Self–Z, Self–raw, Liked–Z, Liked–raw, Pooled–Z, and Pooled–raw.

Previous analyses of these data sets by Peabody and Goldberg (1989) have shown that within the most restricted data set (Z scores from self-descriptions), each of the Big-Five factors tends to be of roughly equal size, and Factors II and IV are well differentiated from each other. In contrast, within the least restricted data set (original responses derived from the pooled sample of targets), Factor II is of enormous size, and Factor IV is reduced to a tiny oblique satellite of Factor II. Because of these large and systematic differences in factor structures associated with the six data sets, they provide an unusually rigorous test of across-sample factor robustness.

Results

The varimax-rotated factor loadings from analyses of the standard-scored responses based on self-descriptions and descriptions of liked peers are presented in Table 3. As in previous tables, asterisks are used to indicate the factor on which each variable had its highest loading. Within the sample of self-descriptions, 100% of the variables had their highest loadings on the factor for which they had been selected as markers, and in the sample of peer descriptions, 97% of the variables passed this test. For analyses of the data sets based on the original response metric, the corresponding values were 96% for the self-descriptions and 94% for the descriptions of liked peers. Within the most "difficult" data sets, those of completely unrestricted range, the corresponding values were 96% (Z) and 93% (raw). For the two samples displayed in Table 4, the factor loadings are as close to the theoretical criterion of simple structure as are likely to be found with real data. For example, in the sample of self-descriptions, only 4 of the 800 loadings on nontargeted factors were larger than .30, and none were as large as .40.

A table presenting the coefficients of factor similarity between the corresponding factors from each of the 15 pairs of analyses is available from the author. These congruence coefficients ranged from .93 to .99 ($M = .95$) for Factor I, from .90 to .99 ($M = .93$) for Factor II, from .92 to .996 ($M = .96$) for Factor III, from .82 to .96 ($M = .90$) for Factor IV, and from .84 to .98 ($M = .92$) for Factor V. The coefficients averaged .93 across all 15 pairs of data sets, attesting to the unusual robustness of these factor structures.

Moreover, each of the five factors was of relatively equal size, especially in the data sets of most restricted range. For example, in the standard-scored self-ratings, the relative factor sizes were 25%, 19%, 21%, 18%, and 17% for Factors I to V, respectively. At the other extreme, in the sample of pooled peers (original responses) the corresponding relative factor sizes were 19%, 32%, 20%, 11%, and 19%. Table 4 presents the internal-consistency reliability coefficients for each of the five marker subsets of 20 items. Across the six data sets, the coefficient alpha reliability estimates ranged from .90 to .92 for Factor I, from .84 to .97 for Factor II, from .88 to .94 for Factor III, from .82 to .88 for Factor IV, and from .82 to .94 for Factor V. Given that these values are so high for 20-item scales, it is important to discover whether they have been purchased at the price of attenuated discriminant validities.

Table 5 presents the relevant findings on this issue, the intercorrelations among all pairs of marker subsets. As reported in Goldberg and Peabody (1989), one major effect associated with evaluation-unrestricted data sets is the increased obliquity of Factor II (Agreeableness) and Factor IV (Emotional Stability). Table 6 shows this effect quite clearly in the analyses of the Pooled sample. Nonetheless, most of the intercorrelations among the marker subsets were quite low, and they were nearly orthogonal in the standard-scored descriptions of oneself and of liked peers. Overall, in the three least restricted data sets, the average interscale correlation was approximately .25, and in the three most restricted data sets, the mean correlation was only approximately .10.

Table 5
Intercorrelations Among the Marker Subsets Within Each Data Set

	Pooled		Liked		Self	
Factor pair	Raw	*Z*	Raw	*Z*	Raw	*Z*
II vs. IV	.58	.53	.33	.25	.10	.13
V vs. II	.51	.46	.41	.18	.18	−.06
V vs. III	.44	.40	.26	.12	.20	.03
III vs. II	.39	.35	.42	.29	.37	.24
III vs. IV	.32	.33	.31	.21	.12	.16
IV vs. V	.28	.28	.09	.04	−.06	.00
I vs. V	.18	.16	.30	.13	.16	.10
I vs. IV	.00	.03	.08	.07	.06	.14
I vs. II	.05	.06	.15	−.03	.18	.09
I vs. III	−.09	−.12	.02	−.10	.04	−.04
M	.27	.25	.24	.12	.13	.08

Congruence With Other Big-Five Factor Markers

This article and that by Goldberg (1990) have provided the rationales for, and the procedures used to develop, at least six different sets of Big-Five factor markers. Two of these sets were based on multiple responses aggregated across the items included within synonym clusters (the original 133 clusters and the 100 revised clusters reported in Goldberg, 1990). Four other sets were based on single responses to stimuli presented either in unipolar or in bipolar formats (the 50 difference scores, the 50 transparent bipolar scales, and the 50 opaque bipolar scales developed in Study 3, and the set of 100 unipolar markers developed in Study 4). How congruent are the latter 100 unipolar markers with the other five sets and with the scales from the NEO and Hogan personality inventories?

Table 6 presents the correlations between the factor scores derived from 98 of the 100 unipolar markers[5] and all of the measures included in Study 3. The 133 synonym clusters, 100 synonym clusters, and 50 difference scores were all scored from standardized responses to the inventory of 479 trait adjectives. Also included in Table 6 are measures of each of the Big-Five domains from Form S of the NEO–PI; in addition to the original NEO–PI domain and facet scales, the "validimax" factor scores recommended by McCrae and Costa (1989c) are also included in the table. Finally, the table includes the scales from the Hogan Personality Inventory as well as the factor scores from a five-factor varimax rotation of the factors derived from the 43 HPI item clusters.[6]

Clearly, factor scores from the 100 unipolar markers were highly congruent with those derived from more representative sets of trait adjectives; the mean convergence correlation across the five corresponding factors were .85 for the 133 clusters and .86 for the 100 clusters described in Goldberg (1990). For the 100 difference scores derived in Study 3 of the present article, the average convergence was very slightly higher (mean $r = .90$). All of these values must be interpreted with caution, however, because the four sets of markers were scored from the same inventory, and they include some common items.

When one turns to experimentally independent sets of markers, congruence shows the expected decrement. The con-

Table 4
The Reliabilities of the Five Subsets of 20 Items in Each of the Six Data Sets

	Subset					
Sample	I	II	III	IV	V	*M*
			Mean interitem correlation			
Pooled						
Raw	.37	.60	.46	.26	.43	.42
Z	.34	.55	.45	.24	.40	.40
Liked						
Raw	.34	.35	.34	.19	.26	.29
Z	.31	.26	.31	.18	.18	.25
Self						
Raw	.31	.29	.30	.21	.23	.27
Z	.30	.22	.27	.20	.19	.24
			Coefficient alpha			
Pooled						
Raw	.92	.97	.94	.88	.94	.93
Z	.91	.96	.94	.87	.93	.92
Liked						
Raw	.91	.91	.91	.83	.87	.89
Z	.90	.87	.90	.82	.81	.86
Self						
Raw	.90	.88	.90	.84	.85	.87
Z	.90	.84	.88	.83	.82	.85

[5] Two of the 100 terms, Introverted (Factor I−) and Neat (Factor III+), were not included in the inventory of 479 trait adjectives used in Study 3.

[6] A table presenting the correlations between the factor scores derived from the 100 unipolar terms and each of the 43 HPI clusters is available from the author.

Table 6
Correlations Between the Factor Scores Derived From 98 of the 100 Unipolar Markers and All of the Variables Included in Study 3

Factor	I	II	III	IV	V
Factors scores from 133 clusters					
Factor I	**.87***	**.30**	.01	−.09	.12
Factor II	−.23	**.81***	−.03	**.30**	.04
Factor III	−.04	.06	**.89***	−.08	.06
Factor IV	.17	−.16	.09	**.85***	−.01
Factor V	−.01	−.13	.02	−.02	**.83***
Factor scores from 100 clusters					
Factor I	**.90***	.26	.02	−.06	.09
Factor II	−.18	**.83***	−.09	**.38**	−.01
Factor III	−.05	.12	**.91***	−.03	.03
Factor IV	.14	**−.31**	.08	**.84***	.04
Factor V	−.02	−.06	.04	−.06	**.83***
Factor scores from 50 difference scores					
Factor I	**.88***	.00	.13	.09	.01
Factor II	.04	**.93***	−.01	.01	−.01
Factor III	−.15	.08	**.86***	−.06	.02
Factor IV	−.14	.05	−.02	**.88***	.02
Factor V	.03	.00	.02	−.02	**.93***
Factor scores from 50 transparent bipolar scales					
Factor I	**.79***	.01	−.05	−.04	−.03
Factor II	−.06	**.74***	−.08	.02	−.02
Factor III	−.03	−.05	**.76***	−.06	−.07
Factor IV	.13	.09	.10	**.62***	−.06
Factor V	.00	−.23	−.02	.13	**.65***
Factor scores from 50 opaque bipolar scales					
Factor I	**.80***	.06	−.03	−.07	.01
Factor II	−.08	**.74***	−.09	.10	−.13
Factor III	−.03	−.04	**.73***	−.06	−.06
Factor IV	.14	−.06	.14	**.65***	−.02
Factor V	−.10	−.21	−.04	.06	**.61***
NEO–PI domain scales					
Extraversion	**.69***	.22	.08	.12	.00
Agreeableness	.05	**.56***	−.03	**.30**	.01
Conscientiousness	.13	−.03	**.67***	.04	.09
Neuroticism	−.23	−.03	−.15	**−.69***	−.03
Openness	.05	.07	−.13	.23	**.46***
NEO–PI validimax factors					
Extraversion	**.67***	**.30**	.07	.06	−.02
Agreeableness	−.12	**.55***	**−.31**	.23	.02
Conscientiousness	−.08	−.10	**.57***	−.08	.24
Neuroticism	−.19	.00	−.06	**−.68***	.06
Openness	−.16	−.14	−.05	.26	**.50***

Factor	I	II	III	IV	V
NEO–PI facet scales					
Assertiveness	**.58***	−.15	.28	.13	.20
Activity	**.47***	−.03	.27	−.09	.05
Gregariousness	**.47***	.24	−.18	−.04	−.11
Positive Emotions	**.46***	**.37**	.03	.24	.01
Warmth	**.42**	**.44***	.05	.10	.01
Excitement seeking	**.30***	.03	−.09	.12	−.16
Anxiety	−.15	.00	−.01	**−.63***	.08
Depression	−.29	−.02	−.14	**−.58***	−.05
Self-Consciousness	**−.35**	.04	−.07	**−.54***	−.16
Hostility	.00	−.26	.04	**−.52***	.04
Vulnerability	−.19	.09	**−.37**	**−.47***	−.06
Impulsiveness	−.03	.03	−.22	**−.37***	.00
Aesthetics	−.09	.14	−.07	.02	**.45***
Ideas	−.02	−.14	−.05	.29	**.36***
Feelings	.18	.16	.09	−.03	**.33***
Values	−.05	−.06	−.06	.12	**.26***
Fantasy	.05	.10	−.18*	.18	.17
Actions	.14	.05	−.26	**.32***	.15
HPI regular scales					
Sociability	**.56***	.00	−.20	−.07	.03
Ambition	**.41***	−.22	.11	.03	.26
Likability	.22	**.52***	−.04	**.31**	−.03
Prudence	−.16	.19	**.56***	−.11	−.06
Adjustment	**.32**	.01	.16	**.62***	.07
Intellectance	−.06	−.21	−.15	.16	**.39***
Validity	.15	.23	**.30***	.11	.17
Resiliency	.12	.01	**.59***	.16	.00
Managerial Potential	.28	.15	.22	**.49***	.14
Service Orientation	**.31**	.28	.21	**.42***	.04
Clerical Aptitude	**.34**	.00	.26	**.37***	.08
Reliability	−.25	**.30**	**.38***	**.33**	−.06
Sales Potential	.12*	.08	−.04	−.10	.07
Factor scores from the 43 HPI clusters					
Factor 2	**.67***	−.01	.08	.00	.23
Factor 3	−.24	.04	**.49***	−.07	.09
Factor 5	.02	.00	**.39***	.05	−.01
Factor 1	.10	.12	.10	**.67***	−.02
Factor 4	−.10	**−.33**	−.09	.18	**.41***

Note. Values equal to or larger than |.30| are listed in boldface type. Sample sizes: 133 clusters, 100 clusters, and 50 difference scores: $N = 192$; transparent and opaque bipolar scales: $N = 175$; variables from the NEO Personality Inventory (NEO-PI): $N = 171$; variables from the Hogan Personality Inventory (HPI): $N = 181$.
* Highest correlate of each variable.

vergence correlations with the bipolar scales averaged .71 in both the transparent and opaque formats, somewhat higher than with the NEO–PI domain scales and validimax factors, which averaged .60; as would be expected, the convergence with the NEO–PI scales was lower for Factor V (Intellect versus Openness) than for the other four factors. Finally, convergence with the HPI scales was substantially lower (averaging .53 for the five most concordant scales); indeed, factor analyses of the 43 HPI clusters revealed no HPI factor that was primarily associated with Factor II (Agreeableness).

General Discussion

One major purpose of this article has been to provide a set of Big-Five factor markers that can now replace those developed over 25 years ago by Norman (1963). Moreover, this new marker set can also be considered as an alternative to the scales in the NEO and the Hogan personality inventories. In selecting among alternative sets of Big-Five markers, investigators must decide between markers based on a reasonably representative sampling of variables and those that provide roughly equal coverage of each of the Big-Five domains. In addition, for each of these two options they will face the inevitable compromises between (a) increasing reliability by using larger marker sets and (b) decreasing the amount of subject testing time by using smaller sets.

Moreover, investigators must decide on the method to be used for scoring their subjects' responses. In each of the four studies described in this article, factor scores (derived from varimax rotations of five principal components) were used for this purpose. The use of factor scores ensures that all pairs of factors are orthogonal, thus maximizing their discriminative validity; in addition, it permits the investigator to examine the factor structure in the new sample so as to compare the pattern of factor loadings with theoretical expectations (e.g., the pattern displayed here in Table 3).

On the other hand, it is not prudent to carry out a factor analysis using small samples. Moreover, it is always necessary to examine the new factor structure so as to discover the Big-Five equivalent of each factor and to determine its direction of scoring. That is, in a particular analysis the Surgency factor might correspond to the fourth rotated factor (rather than to the first), and high scores on that factor might be associated with introversion (rather than with extraversion).

Investigators who are working with small samples, or who do not have access to adequate computational facilities, may wish to score the marker dimensions directly, rather than to calculate factor scores. One advantage of this procedure is uniformity of scores across studies; empirically derived factor structures will inevitably vary at least slightly from sample to sample. On the other hand, because the dimensions will no longer be orthogonal, there will be some loss in discriminant validity, a disadvantage shared with the scores from the scales on the NEO–PI and the HPI.[7]

As has been argued elsewhere (e.g., Peabody & Goldberg, 1989), experimental studies of the determinants of factor size and location should normally include representative samples of variables. For example, the 57 bipolar scales from Peabody (1987), which were selected to be representative of the lexicon of English trait adjectives, may be useful in some contexts as a brief set of factor markers. However, in any representative set of variables, brevity has its price; these 57 scales do not provide highly reliable markers of Factors IV and V. In contrast, the 100 synonym clusters reported in Goldberg (1990), which are also based on representative sampling of variables, provide highly reliable markers but at the cost of administering 339 unipolar adjectives.

For those investigators who desire markers that do not differ appreciably in their reliabilities across the five domains, various alternative sets have been described in this article. Study 1 included a comparison between bipolar scales and two types of variables derived from antonym pairs administered in unipolar format, 100 single trait terms and 50 difference scores. The findings from that study, which were replicated in Study 3, show that the factor structures derived from these two types of unipolar variables are virtually identical. Moreover, in both studies the unipolar variables provided more univocal representations of the Big-Five structure than did the bipolar scales.

However, because bipolar scales are preferred by some investigators, one of the purposes of Studies 2 and 3 was to compare a new transparent format for administering bipolar scales with the quasirandomized format typically used. In general, more univocal patterns of factor loadings were obtained from the marker sets administered in the transparent format than from those administered in the traditional one. Moreover, because of their transparent nature, these bipolar markers are quite useful in educational settings as a demonstration of the Big-Five structure. For other purposes, however, most investigators may prefer the 100 unipolar variables developed in Study 4; one format for administering this marker set is included here in Appendix A. The findings from Study 4 suggest that factor scores based on subject-standardized responses to these 100 items provide quite univocal markers of each of the Big-Five domains. Moreover, the factors derived from this small set of factor markers have been shown to be substantially similar to those derived from representative samples of variables. It is to be hoped that the availability of this easily administered set of factor markers will now encourage investigators of diverse theoretical viewpoints to communicate in a common psychometric tongue.

---

[7] For example, the correlations among the five scales of the NEO–PI (after reflecting the scores on the Neuroticism scale) averaged .24 in both Study 2 and Study 3.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47,* (1, Whole No. 211).

American Psychological Association. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Bond, M. H. (1979). Dimensions used in perceiving peers: Cross-cultural comparisons of Hong Kong, Japanese, American and Filipino university students. *International Journal of Psychology, 14,* 47–56.

Bond, M. H. (1983). Linking person perception dimensions to behavioral intention dimensions: The Chinese connection. *Journal of Cross-Cultural Psychology, 14,* 41–63.

Bond, M. H., & Forgas, J. P. (1984). Linking person perception to behavior intention across cultures: The role of cultural collectivism. *Journal of Cross-Cultural Psychology, 15,* 337–352.

Bond, M. H., Nakazato, H., & Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese person perception. *Journal of Cross-Cultural Psychology, 6,* 346–357.

Borkenau, P. (1988). The multiple classification of acts and the Big Five factors of personality. *Journal of Research in Personality, 22,* 337–352.

Botwin, M. D., & Buss, D. M. (1989). The structure of act report data: Is the five factor model of personality recaptured? *Journal of Personality and Social Psychology, 56,* 988–1001.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* Yonkers-on-Hudson, NY: World Book.

Costa, P. T., Jr., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI factor scales with measures of the five factor model of personality. *Journal of Personality Assessment, 50,* 640–650.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1987). Personality assessment in psychosomatic medicine: Value of a trait taxonomy. In G. A. Fava & T. N. Wise (Eds.), *Advances in psychosomatic medicine: Vol. 17. Research paradigms in psychosomatic medicine* (pp. 71–82). Basel: Karger.

Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology, 55,* 258–265.

D'Andrade, R. G. (1965). Trait psychology and componential analysis. *American Anthropologist, 57,* 215–228.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 41, pp. 417–440). Palo Alto, CA: Annual Reviews.

Goldberg, L. R. (1981). Unconfounding situational attributions from uncertain, neutral, and ambiguous ones: A psychometric analysis of descriptions of oneself and various types of others. *Journal of Personality and Social Psychology, 41,* 517–552.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1, pp. 203–234). Hillsdale, NJ: Erlbaum.

Goldberg, L. R. (1990). An alternative "Description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48,* 82–98.

Guthrie, G. M., & Bennett, A. B., Jr. (1971). Cultural differences in implicit personality theory. *International Journal of Psychology, 6,* 305–312.

Hakel, M. D. (1969). Significance of implicit personality theories for personality research and theory. *Proceedings of the 77th annual American Psychological Association Convention* (pp. 403–404).

Hamsher, J. H. (1969). Disguise of purpose in psychological testing. *Proceedings of the 77th annual American Psychological Association convention* (pp. 141–142).

Hofstee, W. K. B. (1969). Method effects in judging the desirability of traits. *Educational and Psychological Measurement, 29,* 583–604.

Hogan, R. (1986). *Hogan Personality Inventory manual.* Minneapolis, MN: National Computer Systems.

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* (pp. 66–100). New York: Guilford Press.

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

Kestenbaum, J. M., & Hammersla, J. (1976). Filler items and social desirability in Rotter's Locus of Control Scale. *Journal of Personality Assessment, 40,* 162–168.

Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review, 90,* 185–214.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology, 55,* 312–320.

Lieberman, L. R., & Walters, W. M., Jr. (1968). Effect of background items on responses to personality inventory items. *Journal of Consulting and Clinical Psychology, 32,* 230–232.

McCrae, R. R., & Costa, P. T., Jr. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6,* 587–597.

McCrae, R. R., & Costa, P. T., Jr. (1986). Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist, 41,* 1001–1003.

McCrae, R. R., & Costa, P. T., Jr. (1989a). More reasons to adopt the five-factor model. *American Psychologist, 44,* 451–452.

McCrae, R. R., & Costa, P. T., Jr. (1989b). 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., & Costa, P. T., Jr. (1989c). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research, 24,* 107–124.

McCrae, R. R., & Costa, P. T., Jr. (1989d). The structure of interpersonal traits: Wiggins' circumplex and the five-factor model. *Journal of Personality and Social Psychology, 56,* 586–595.

Nakazato, H., Bond, M. H., & Shiraishi, D. (1976). Dimensions of personality perception: An examination of Norman's hypothesis. *Japanese Journal of Psychology, 47,* 139–148.

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

Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population.* Department of Psychology, University of Michigan, Ann Arbor.

Norman, W. T. (1969). "To see oursels as ithers see us!": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research, 4,* 417–443.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.

Osberg, T. M. (1985). Order effects in the administration of personality measures: The case of the self-consciousness scale. *Journal of Personality Assessment, 49,* 536–540.

Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology, 1,* 44–49.

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

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

Perkins, J. E., & Goldberg, L. R. (1964). Contextual effects on the MMPI. *Journal of Consulting Psychology, 28,* 133–140.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Schriesheim, C. A. (1981a). The effect of grouping or randomizing items on leniency response bias. *Educational and Psychological Measurement, 41,* 401–411.

Schriesheim, C. A. (1981b). Leniency effects on convergent and discriminant validity for grouped questionnaire items: A further investigation. *Educational and Psychological Measurement, 41,* 1093–1099.

Schriesheim, C. A., & Denisi, A. S. (1980). Item presentation as an influence on questionnaire validity: A field experiment. *Educational and Psychological Measurement, 40,* 175–182.

Schriesheim, C. A., Solomon, E., & Kopelman, R. E. (1989). Grouped versus randomized format: An investigation of scale convergent and discriminant validity using LISREL confirmatory factor analysis. *Applied Psychological Measurement, 13,* 19–32.

Smith, S. S., & Kihlstrom, J. F. (1987). When is a schema not a schema? The "big five" traits as cognitive structures. *Social Cognition, 5,* 26–57.

Solomon, E., & Kopelman, R. E. (1984). Questionnaire format and scale reliability: An examination of three modes of item presentation. *Psychological Reports, 54,* 447–452.

Strahan, R. F., & Wilson, D. W. (1976). Buffer items, insight, and the Marlowe-Crowne Social Desirability scale. *Journal of Personality Assessment, 40,* 320–323.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical Report ASD-TR-61-97. Lackland Air Force Base, TX: U.S. Air Force.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology, 57,* 120–128.

Weiss, D. S., & Mendelsohn, G. A. (1986). An empirical demonstration of the implausibility of the semantic similarity explanation of how trait ratings are made and what they mean. *Journal of Personality and Social Psychology, 50,* 595–601.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37,* 395–412.

Wiggins, J. S. (1980). Circumplex models of interpersonal behavior. In L. Wheeler (Eds.), *Review of personality and social psychology* (Vol. 1, pp. 265–294). Beverly Hills, CA: Sage.

Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 305–316.

Wiggins, J. S., & Trapnell, P. D. (in press). Personality structure: The return of the Big Five. In S. R. Briggs, R. Hogan, & W. H. Jones (Eds.), *Handbook of personality psychology.* Orlando, FL: Academic Press.

Wiggins, N. H., & Blackburn, M. C. (1976). Implicit theories of personality: An individual differences approach. *Multivariate Behavioral Research, 11,* 267–285.

Appendix A

The 100 Unipolar Markers Developed in Study 4

How Accurately Can You Describe Yourself?

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly your same age.

Before each trait, please write a number indicating how accurately that trait describes you, using the following rating scale:

Inaccurate					Accurate			
Extremely	Very	Quite	Slightly	Neither	Slightly	Quite	Very	Extremely
1	2	3	4	5	6	7	8	9

________ Active	________ Extraverted	________ Negligent	________ Trustful
________ Agreeable	________ Fearful	________ Nervous	________ Unadventurous
________ Anxious	________ Fretful	________ Organized	________ Uncharitable
________ Artistic	________ Generous	________ Philosophical	________ Uncooperative
________ Assertive	________ Haphazard	________ Pleasant	________ Uncreative
________ Bashful	________ Harsh	________ Practical	________ Undemanding
________ Bold	________ Helpful	________ Prompt	________ Undependable
________ Bright	________ High-strung	________ Quiet	________ Unemotional
________ Careful	________ Imaginative	________ Relaxed	________ Unenvious
________ Careless	________ Imperceptive	________ Reserved	________ Unexcitable
________ Cold	________ Imperturbable	________ Rude	________ Unimaginative
________ Complex	________ Impractical	________ Self-pitying	________ Uninquisitive
________ Conscientious	________ Inconsistent	________ Selfish	________ Unintellectual
________ Considerate	________ Inefficient	________ Shallow	________ Unintelligent
________ Cooperative	________ Inhibited	________ Shy	________ Unkind
________ Creative	________ Innovative	________ Simple	________ Unreflective
________ Daring	________ Insecure	________ Sloppy	________ Unrestrained
________ Deep	________ Intellectual	________ Steady	________ Unsophisticated
________ Demanding	________ Introspective	________ Sympathetic	________ Unsympathetic
________ Disorganized	________ Introverted	________ Systematic	________ Unsystematic
________ Distrustful	________ Irritable	________ Talkative	________ Untalkative
________ Efficient	________ Jealous	________ Temperamental	________ Verbal
________ Emotional	________ Kind	________ Thorough	________ Vigorous
________ Energetic	________ Moody	________ Timid	________ Warm
________ Envious	________ Neat	________ Touchy	________ Withdrawn

(*Appendix B follows on next page*)

Appendix B

Example of the Format of the Transparent Bipolar Inventory

	Very		Moderately		Neither		Moderately		Very	
					Introversion–Extroversion					
introverted	1	2	3	4	5	6	7	8	9	extraverted
unenergetic	1	2	3	4	5	6	7	8	9	energetic
silent	1	2	3	4	5	6	7	8	9	talkative
timid	1	2	3	4	5	6	7	8	9	bold
inactive	1	2	3	4	5	6	7	8	9	active
unassertive	1	2	3	4	5	6	7	8	9	assertive
unadventurous	1	2	3	4	5	6	7	8	9	adventurous
					Pleasantness or Agreeableness					
cold	1	2	3	4	5	6	7	8	9	warm
unkind	1	2	3	4	5	6	7	8	9	kind
uncooperative	1	2	3	4	5	6	7	8	9	cooperative
selfish	1	2	3	4	5	6	7	8	9	unselfish
disagreeable	1	2	3	4	5	6	7	8	9	agreeable
distrustful	1	2	3	4	5	6	7	8	9	trustful
stingy	1	2	3	4	5	6	7	8	9	generous
					Conscientiousness or Dependability					
disorganized	1	2	3	4	5	6	7	8	9	organized
irresponsible	1	2	3	4	5	6	7	8	9	responsible
negligent	1	2	3	4	5	6	7	8	9	conscientious
impractical	1	2	3	4	5	6	7	8	9	practical
careless	1	2	3	4	5	6	7	8	9	thorough
lazy	1	2	3	4	5	6	7	8	9	hardworking
extravagant	1	2	3	4	5	6	7	8	9	thrifty
					Emotional Stability					
angry	1	2	3	4	5	6	7	8	9	calm
tense	1	2	3	4	5	6	7	8	9	relaxed
nervous	1	2	3	4	5	6	7	8	9	at ease
envious	1	2	3	4	5	6	7	8	9	not envious
unstable	1	2	3	4	5	6	7	8	9	stable
discontented	1	2	3	4	5	6	7	8	9	contented
emotional	1	2	3	4	5	6	7	8	9	unemotional
					Intellect or Sophistication					
unintelligent	1	2	3	4	5	6	7	8	9	intelligent
unanalytical	1	2	3	4	5	6	7	8	9	analytical
unreflective	1	2	3	4	5	6	7	8	9	reflective
uninquisitive	1	2	3	4	5	6	7	8	9	curious
unimaginative	1	2	3	4	5	6	7	8	9	imaginative
uncreative	1	2	3	4	5	6	7	8	9	creative
unsophisticated	1	2	3	4	5	6	7	8	9	sophisticated

Received February 12, 1991
Revision received May 13, 1991
Accepted May 13, 1991 ■