TOWARD AN ADEQUATE TAXONOMY OF PERSONALITY ATTRIBUTES:

REPLICATED FACTOR STRUCTURE IN PEER NOMINATION PERSONALITY RATINGS [1]

WARREN T. NORMAN

University of Michigan

A rationale and procedures for the development of a taxonomic basis for personality research and theory construction are outlined. Some prior research efforts in this area are briefly reviewed and evaluated. Results of a series of studies are presented which employed primarily groups of university male Ss who varied in length and intimacy of previous associations. Data were obtained using peer nomination rating methods. These analyses yielded clear and consistent evidence for the existence of 5 relatively orthogonal, easily interpreted personality factors. Behavioral correlates of selected patterns of scores on these factors are briefly discussed and suggestions are made concerning subsequent research efforts warranted by these findings.

The arguments and research findings to be discussed in this paper focus on the problems involved in developing an adequate descriptive or observation language for personality as an area of psychological research and theory construction. The basic commitment which underlies the research efforts and programmatic considerations presented is that the construction of more effective theories of the development, structure, and functioning of personality will be facilitated by having available an extensive and well-organized vocabulary by means of which to denote the phenotypic attributes of persons.

It is explicitly *not* assumed that complete theories of personality will simply emerge automatically from such taxonomic efforts. In the opinion of this writer, there is a good deal more to theory construction and refinement than the development of an observation language—even a good one. Neither are the arguments, empirical results, and assessment methods described below presented to support the notion that a complete and adequate, descriptive, taxonomic basis for personality is now a *fait accompli*. Rather they are offered

as a contribution toward such a goal, as a rudimentary basis from which further work can profitably proceed and, by implication, as an assessment of our current status in this enterprise.

LOGIC AND PROCEDURES FOR THE DEVELOPMENT OF A TAXONOMIC BASIS FOR PERSONALITY

Attempts to construct taxonomies of personality characteristics have ordinarily taken as an initial data base some set of perceptible variations in performance and appearance between persons or within individuals over time and varying situations. By far the most general efforts to specify the domain of phenomena on which to base such a system have proceeded from an examination of the natural language. The argument in its essential form has been that perceptible differences between persons in their characteristic appearance or manner of behaving or changes over time and situations of single individuals in these regards become codified as a subset of the descriptive predicates of the natural language in the course of its development. Thus we find Allport and Odbert (1936) reporting the results of their exhaustive search for all "trait names" in a standard English dictionary. From such a compendium (their's comprised some 18,000 terms altogether) presumably, would be drawn the descriptive predicates of

the observation language for any theory of personality.

The second step in the taxonomic effort would consist in selecting from the original set those terms suitable for inclusion in a scientific data language; that is, that subset which possesses unambiguous *denotative* reference to limited classes of relevant observable phenomena. There is no necessary reason why every trait descriptive term that occurs in the natural language should be required. Nor is it likely even that all of those terms, admissible as members of the data language vocabulary, will refer to sufficiently distinct and independent classes of phenomena to warrant including all of them. Thus a third aspect of the taxonomic problem would be to cull from the admissible set of predicates, those which are not needed in order parsimoniously to span the domain of phenomena referred to by the entire collection.

On completion of this step the taxonomic effort could be, and typically has been, terminated and the focus of the research program directed toward other goals. However, it would seem important to carry the endeavor a step or two further. In the first place, if the categorizing of terms has been done on the basis of direct judgmental comparisons among them, some assurance is needed that in *actual denotative use,* terms in a given category are functionally interchangeable, at least to a high degree. Nor should there be terms in two different classes which are functionally synonymous in denotative reference. Second, distinctness of the classes does not imply independence of occurrence of the attributes denoted by the terms in the various categories.

The implication of these last two points is that the development of an adequate taxonomy involves more than an exhaustive cataloguing of designators into distinct homogeneous classes—it also requires *an organization of the classes* based on the mutual relationships or joint occurrences of the phenomena or attributes denoted by the terms in the various sets. This entails the collection of data from representative groups of persons from the language community concerning their use of these terms to describe persons or their actions, and the analysis of these data to determine the degree of relationship or joint usage of terms in the several categories.

SOME PRIOR RESEARCH FINDINGS

The list of trait names which resulted from the search by Allport and Odbert (1936) were used by Cattell (1947, 1957) in his conceptualization of the Language Personality Sphere and in his development of the condensed Personality Sphere set of rating scales. In brief, Cattell limited his consideration first to the subset of 4,504 terms which are contained in Column I of the Allport-Odbert monograph—those characterized by Allport (1937) as,

. . . most clearly "real" traits of personality. They designate generalized and personalized determining tendencies—consistent and stable modes of an individual's adjustment to his environment. . . . These terms do not imply merely temporary and specific behavior as do the terms in Column II . . . they are more neutral and less censorial than those in Column III; and they are less metaphorical and remote in their applicability to personality than those in Column IV [p. 366].

Having thus initially curtailed his consideration to those terms denoting relatively "consistent and stable" aspects of personality a further reduction (à la Phase 3 above) was accomplished by Cattell (1957, pp. 71–72). Based on the notion of "just distinguishable" differences in judgments of semantic meaning, some 171 terms were selected to represent "synonym groups" presumably spanning the "personality sphere." Ratings obtained on these "trait elements" were then correlated and another reduction in the set was accomplished by cluster analysis methods. Each of the 36 clusters so identified (and subsequently augmented to number 42–46) was converted into bipolar descriptions and this collection was dubbed the *standard reduced personality sphere* set of rating scales.

Thus we see that by a combined use of semantic judgments and correlational evidence the initial set of trait descriptive terms was rather severely condensed by Cattell and converted into a comparatively small number of rating scales. These scales were then intended for use in subsequent investigations to determine the organizational or structural features of personality—the final phase of a taxonomic

program of the sort discussed above. These scales have in fact been employed by Cattell and by others in a number of such studies. The results of factor anlayses of the obtained data have not been unequivocal, however. Cattell (1957) contends that,

... at the level of behavior ratings, an almost complete structuring has been obtained . . . . The results of four consecutive and interrelated studies yield twelve very stable and two or three less definite primary personality factors [p. 73].

By way of contrast, Tupes and Christal (1961) report the results of analyzing data collected on these scales from eight separate samples (including reanalyses of two of Cattell's own matrices):

These samples differed in length of acquaintanceship from three days to a year or more; in kind of acquaintanceship from assessment programs in a military training course to a fraternity house situation; in type of subject from airmen with only a high school education to male and female undergraduate students to first-year graduate students; and in type of rater from very naive persons to clinical psychologists and psychiatrists with years of experience in the evaluation of personality.

In all solutions except one there appeared to be *five relatively strong and recurrent factors and nothing more of any consequence* [italics added]. In one solution, based on data from undergraduate women, the fifth factor split into two highly related factors [pp. 13–14].

Thus there appears to be a considerable disparity between the results presented by Cattell and those reported by Tupes and Christal—even for the two sets of data which both analyzed.

One possible explanation lies in the rotational methods employed; an oblique simple structure criterion by Cattell but an orthogonal approximation to simple structure by Tupes and Christal. This would not account, of course, for the differences in the numbers of factors extracted. In the male college sample analyzed by both, Cattell (1947) extracted 13 centroid factors whereas Tupes and Christal could find no basis for extracting more than 8 factors from these same data.

This writer has several reasons for preferring the Tupes and Christal version to that of Cattell. First, Cattell (1947) presents a "representative set" of reliability coefficients for 12 of the 36 scales he employed in this study. Eleven of these 12 are *lower (sic)* in value than the corresponding communality values he reports for the unrotated, centroid factors from these data—over half of them appreciably lower. If one grants the appropriateness of the method of reliability estimation used and considers the relation between common factor variance and reliable variance in the factor analytic model, Cattell's results present at least a quandary.

Second, despite the fact that oblique rotations were used by Cattell, the defining variables or salients for his rotated factors have loadings which are remarkably small in magnitude when compared either with their unrotated communalities or with the loadings obtained by Tupes and Christal for these scales on their orthogonal factors. While interpretation is certainly facilitated by high hyperplane counts, it is also easier when the salients load both strongly and uniquely on a single factor. In Cattell's analysis no variable loads as high as .60 on any factor and over half of the variables which are salient on one factor are salient on two or more.

It has been argued however, that factorial invariance is an important result of using a complete simple structure criterion for rotation. The four studies reported by Cattell (1957) are in fact construed by him as producing evidence of invariance for the structure he presents. But the Tupes and Christal analyses certainly present no less convincing evidence of invariance. One of the major objectives of the studies and analyses to be presented below, in fact, was to determine *numerical estimates of the degree of factor similarity* for factor solutions obtained using a set of scales drawn from the reduced personality sphere set with diverse samples of subjects.

A second objective was to determine whether the use of completely analytical factor extraction and rotation procedures would yield a factor structure similar to those obtained by Tupes and Christal who used simplified calculation and visual rotation methods.[2] A third objective was to determine

---

[2] For one of their samples, Tupes and Christal also rotated the matrix using an IBM 650 normal varimax program. The two solutions obtained were extremely similar.

whether even further reduction in the number of scales included in the analyses would alter markedly the factor structures obtained.

METHOD AND PROCEDURE

In the present studies, the peer nomination rating scales selected for data collection were drawn from the reduced personality sphere set on the basis of the results of several analyses presented by Tupes and Christal (1958). The four scales with the highest median factor loadings for each of the five factors identified in these earlier analyses were selected. Abbreviated descriptions of these scales together with suggested names for the factors they represent are listed in Table 1. The actual trait descriptions employed in the data collection were those presented by Cattell (1947) in his *Psychometrika* article and also may be found in the Appendix of a previous report by Norman (1961). The polar descriptions for these scales were typed, one scale to a page, and assembled in a booklet form. The scales representing a given factor were systematically alternated with scales for other factors throughout the booklet to minimize the effects of serial response sets.

Four samples of male college students from the University of Michigan were selected for study as follows:

A. 84 senior ROTC students
B. 82 fraternity and co-op residence men (largely seniors)
C. 215 fraternity men (largely sophomores and juniors)
D. 241 dormitory residence hall men (largely freshmen, sophomores, and juniors)

In each of the four samples, rating groups numbering 6–16 men per group were formed in each residence or ROTC unit so as to maximize interrater familiarity and to minimize status differences owing to academic class standing. All subjects in each of the samples were volunteers who agreed to participate in their respective groups in a large scale psychological testing study. Each subject was paid at the rate of $1.00 an hour to insure his cooperation and participation throughout all phases of the testing program.

Each rating group was administered the set of 20 peer nomination scales. The administrative instructions required each person in a rating group to nominate one-third of the *other* members of his group (that is, excluding self) on Pole "A" and one-third on Pole "B" of each scale. Scores for each person in the group on each scale were determined by summing nominations received by him on Pole A, subtracting the number of nominations received on Pole B, adjusting the score for group size, and

TABLE 1

ABBREVIATED DESCRIPTIONS OF THE PEER NOMINATION RATING SCALES AND THEIR A PRIORI FACTOR DESIGNATIONS

Factor name[a]	Abbreviated scale labels[b]		
	Number	Pole A	Pole B
I. Extroversion or Surgency	1	Talkative-Silent	
	2	Frank, Open-Secretive	
	3	Adventurous-Cautious	
	4	Sociable-Reclusive	
II. Agreeableness	5	Goodnatured-Irritable	
	6	Not Jealous-Jealous	
	7	Mild, Gentle-Headstrong	
	8	Cooperative-Negativistic	
III. Conscientiousness	9	Fussy, Tidy-Careless	
	10	Responsible-Undependable	
	11	Scrupulous-Unscrupulous	
	12	Persevering-Quitting, Fickle	
IV. Emotional Stability	13	Poised-Nervous, Tense	
	14	Calm-Anxious	
	15	Composed-Excitable	
	16	Not Hypochondriacal-Hypochondriacal	
V. Culture	17	Artistically Sensitive-Artistically Insensitive	
	18	Intellectual-Unreflective, Narrow	
	19	Polished, Refined-Crude, Boorish	
	20	Imaginative-Simple, Direct	

Note.—Based on the findings of Tupes and Christal (1958).
[a] Pole A.
[b] For the actual scale labels and instructions to raters employed in data collection see Norman (1961, Appendix).

adding a constant to eliminate negative values. The actual formula used in all computations of rating scale scores was,

$$RSS = 10 + \frac{10}{N - 1}(X_A - X_B)$$

where N = number of raters in the group, X_A = number of "A" nominations received by the subject from all raters in the group on the given scale, and X_B = number of "B" nominations received on the scale.

There are several points worth noting concerning the effects of these procedures for data collection and scoring (which are essentially the same as those used in prior studies by Tupes and Christal and only slightly different from those used by Cattell). The first is that they preclude any differences between rating groups in average scores on the rating scales. However, the variances and higher moments of the group distributions are not so constrained. Second, if scores are obtained from rating groups of markedly different sizes and then (as is customary) are combined in a single distribution, those scores obtained from the larger groups, being based on judgments from more raters, will possess higher reliabilities. The effect of this is that two persons with the same "true score," one of whom is a member of a small rating group and the other of a large one, can be expected to have systematically different observed scores with the person from the smaller group being more extreme. The details of this argument, some empirical data confirming the existence of this effect, and a proposed solution to the problem are presented by Willingham (1959a, 1959b). However, since the magnitude of this bias is relatively small for differences in group size such as those which obtain in the present studies, Willingham's correction was not employed. And finally the point should be made that although *group mean profiles* across scales are necessarily congruent and flat, a profile of scores *for any given individual* is in no such way constrained, either as to its elevation, scatter, or shape components.

For each of the four samples, rating scale scores were intercorrelated and separate factor analyses were carried out analytically using a principal axes method of factoring and a normalized varimax rotation procedure. Initial estimates of communalities were obtained from the highest correlations in the columns and communalities were iterated until they stabilized. The computer program was set to extract 100% of the common factor variance but no specification of the number of factors to be extracted was given. In addition, Samples C (Fraternity group) and D (Residence Hall group) were combined and were then randomly partitioned by rating group to form a pair of double cross-validation samples which were carefully matched in terms of type of residence and distribution of academic class standing. The rating scale scores for this regrouping of Samples C and D

were also analyzed according to the specifications described above.

Factor relationship indices were computed between all pairs of factors resulting from the six analyses using a method recently developed by Henry F. Kaiser.[3] In addition, rough approximations of factor scores were computed by simply summing the scores obtained on the four rating scales representing a given factor for each subject. These "factor scores" were then intercorrelated separately within each of the six analysis samples.

RESULTS

The factor tables for Samples C and D are presented in Table 2, together with the communality values for each scale in each of these two analyses.[4] At the bottom of the table are recorded also the percentage of common factor variance attributable to each factor in each analysis and an index of factor purity. This latter index was computed by taking the ratio of the sum of the squared loadings for the a priori factor salients divided by the latent root for that factor, the ratio multiplied by 100.

It can be seen that the communality values are generally quite high. However, the most striking feature of the results in Table 2 is the very close approximation to orthogonal simple structure achieved in each analysis by the varimax rotation procedure. This is most easily observed by running down the column of factor loadings and observing the difference between the loadings of the a priori salients (in bold face type) on that factor and the loadings of all other scales in that column. The sole exception to the general pattern in any of the six analyses occurred for the loading of Scale 19 on Factor III in the

TABLE 2

FACTOR TABLES BASED ON THE PEER NOMINATION RATING SCALE DATA FROM TWO OF THE
SIX ANALYSIS SAMPLES

Scale	Sample C[a]						Sample D[b]					
	Factor					h^2	Factor					h^2
	I	II	III	IV	V		I	II	III	IV	V	
1	**90**	02	−02	04	−00	81	**90**	−07	06	−04	03	82
2	**78**	−08	07	−03	07	62	**79**	03	−15	03	05	65
3	**79**	15	−20	32	01	79	**79**	−03	−16	30	−00	73
4	**86**	01	−18	−01	−02	77	**90**	−03	01	09	−09	83
5	17	**80**	17	12	07	72	16	**85**	−04	15	11	78
6	−10	**64**	20	49	07	70	−20	**72**	−05	40	06	73
7	−20	**80**	27	19	10	80	−37	**79**	15	07	13	80
8	33	**74**	28	13	11	76	28	**73**	01	14	12	65
9	−33	−08	**66**	−35	20	70	−45	02	**43**	−31	34	60
10	−03	32	**86**	08	18	89	−25	57	**42**	04	41	74
11	−30	44	**68**	−02	20	79	−53	55	**37**	−10	26	80
12	−05	28	**74**	12	27	72	−31	39	**33**	12	45	57
13	01	56	15	**61**	05	70	01	45	11	**58**	17	58
14	06	21	−10	**82**	−07	73	14	17	−04	**81**	−03	70
15	13	06	16	**71**	24	60	16	02	02	**71**	27	60
16	21	27	−00	**65**	−09	55	10	30	−05	**65**	−08	54
17	−04	08	39	−10	**75**	73	−20	08	27	−08	**60**	48
18	−04	05	47	04	**74**	78	−01	14	10	08	**84**	73
19	15	25	53	16	**46**	60	13	32	41	10	**56**	61
20	12	19	03	10	**68**	52	04	17	−20	09	**70**	56
Percentage of variance[c]	23	22	22	18	15		29	27	7	18	19	
Factor Purity Index[d]	85	70	69	77	86		73	65	60	81	72	

Note.—Decimal points omitted for factor loadings and communalities. All values rounded to two significant digits. Loadings for a priori salients in bold face type. Principal axes solutions with normalized varimax rotations.
[a] Fraternity groups ($N = 215$).
[b] Residence Hall groups ($N = 241$).
[c] (Latent root ÷ sum of the latent roots) × 100.
[d] (Sum of squared loadings for a priori salients ÷ latent root) × 100.

factor matrix for Sample D. In that one instance, a non-a-priori salient has a loading exceeding two of the a priori salients. In this one instance, apparently, the obliquity between Factors III and V was sufficiently great to make the orthogonal varimax solution a rather bad fit to the actual data. The factor purity indices in the last line of the table also indicate that the a priori salients for each factor are in fact accounting for most of the variance attributable to that factor.

The percentage of variance entries at the bottom of the factor tables indicates that in each analysis the factors are accounting for roughly equivalent proportions of the total common factor variance—the most pronounced exception in any of the six analyses

again being Factor III in Sample D. By no means is there any indication of a large general factor, for instance, one attributable to general evaluative tendencies or social desirability, present anywhere in these results. This is not to say, of course, that these dimensions of personality are "value-free"; only that the methods of data collection and analysis used have rather successfully eliminated the influence of this pervasive and well-known class of response sets. Important as these sorts of response determiners may be as objects of study in their own right, the intent here was to get at other differential aspects of personality as clearly and as independently of "desirability" or "evaluation" effects as possible. The forced-choice nomination method of data collection and the nor-

TABLE 3

KAISER COEFFICIENTS OF FACTOR SIMILARITY BASED
ON THE ANALYSES OF THE PEER NOMINATION
RATING SCALE DATA FROM TWO OF THE
SIX ANALYSIS SAMPLES

Sample D[a]	Sample C[b]				
	Factor				
	I	II	III	IV	V
Factor					
I	[990]	−023	−136	−009	−029
II	056	[964]	256	006	−049
III	110	−264	[910]	030	−300
IV	004	002	−037	[999]	−014
V	068	−034	295	024	[952]

Note.—Decimal points omitted. All values rounded to three significant digits.
[a] Residence Hall groups ($N = 241$).
[b] Fraternity groups ($N = 215$).

malized varimax rotation procedures were chosen, in fact, partly because they are known to militate against the influence of such general factors, though they certainly do not preclude finding such effects if they are operative. It is thus reassuring to this writer at least, to observe that no factor was found which even closely resembled such a general evaluative response set effect in any of these analyses.

The factor similarity indices between analysis Samples C and D reported in Table 3 (and those between all other pairs of samples as well; cf. Footnote 4, p. 578) also clearly support this impression of relative independence and clear definition of the five personality factors. These values are strictly interpretable as correlation coefficients between factors after the common factor projections of the test vectors (*not* the factors!) have been matched as closely as possible between the two studies.

Of the 75 corresponding factor values obtained (by relating factor structures between all pairs of samples), none was less than .888 in value. Of the 300 other indices (between noncorresponding factors) the greatest magnitude obtained was a value of .304. This method of relating factors is comparatively new and an extensive empirical basis for interpretating these indices has not yet accumulated. This writer, however has seen about a dozen other matrices of this sort computed

between diverse samples using a variety of rating and self-report personality measures. In no case has one been seen in which the correspondence was so close as those obtained in the studies reported here. If one simply uses as a basis for interpretation the fact that these indices are linear correlations, the difference between the lowest (of 75) corresponding factor values, .888, and the magnitude of the largest (of 300) noncorresponding factor values, .304, certainly indicates a high degree of correspondence between the factor structure, across these samples.

One final indication of the degree of relationship among these factors is presented in Table 4. Here the "factor scores" obtained by simply summing each individual's rating scale scores for the four a priori salients on each of the five factors have been correlated between each pair of factors across persons within each of the two analysis samples, C and D (cf. Footnote 4, 578). In the matrix for the fraternity seniors (Sample B), only the correlation between Factor II and Factor IV was of appreciable magnitude (.46). In the remaining five matrices various combinations of factors did relate somewhat appreciably, notably II with III, II with IV, and III with V. These results, together with the find-

TABLE 4

"FACTOR SCORE" INTERCORRELATIONS BASED ON THE
PEER NOMINATION RATING SCALE DATA FROM
TWO OF THE SIX ANALYSIS SAMPLES

Factor	Factor				
	I	II	III	IV	V
Sample C[a]					
I					
II	08				
III	−27	48			
IV	20	55	10		
V	02	37	63	18	
Sample D[b]					
I					
II	−02				
III	−49	49			
IV	22	44	05		
V	−02	31	59	19	

Note.—Decimal points omitted. All values rounded to two significant digits. See text for method used to compute "factor scores."
[a] Fraternity groups ($N = 215$).
[b] Residence Hall groups ($N = 241$).

ings from the factor analyses where occasionally a scale was found to load moderately on a factor for which it was not an a priori salient, do seem to indicate that some degree of obliquity may exist among certain pairs of the dimensions of personality for which these factor measures have been developed.

DISCUSSION

Despite the reservations just mentioned, however, it is clear that a relatively orthogonal and highly stable structure of personal characteristics has been identified and that reasonably good measures of these characteristics are available in situations where the peer nomination method of data collection can be reasonably employed. It is also reassuring to note that the results from our one sample where the subjects had lived together for periods of 1–3 years in a very intimate interpersonal context (the Fraternity Senior study) the factor structure was somewhat "cleaner" and the factor scores considerably less highly intercorrelated. By way of contrast the ROTC Senior study employed persons who had been together only in formal classroom and military drill situations, but who had not, in general, shared common residence experiences. In a preliminary analysis of the data from this latter study where only 98% of the common factor variance had been called for, a fifth factor failed to emerge. An inspection of the factor table indicated that Factors II and IV had failed to separate. The subjects in this study complained more than usual that they felt uncomfortable about filling out the rating forms and felt unable to make the discriminations among some of their peers called for by the rating procedure. Inspection of the interscale correlations and the rating scale scores for given individuals within this sample, in fact, did reflect the presence of some halo effect. Nevertheless, sufficiently differential discriminations apparently were made by the respondents despite these difficulties to produce a reasonably clear and interpretable five factor structure when analyzed according to the revised specifications.

Some complaints of the sort mentioned above are of course to be expected, even from groups of intimate associates, when they are required to use a forced-nomination rating procedure. But, the question can be raised as to whether the results obtained here are truly a reflection of discriminations made by the respondents of the attributes of their peers. Perhaps the structures obtained might more simply be attributed to just the administrative and analysis procedure used. If there had been no degree of agreement among raters, however, or if the nominations (being forced) were actually assigned at random, then the variance to be expected for such a distribution can be, and was, calculated. In every sample, the obtained variances on every scale were larger than would be expected for such a random nominating process. Nor did the interscale correlations in any sample cluster homogeneously about zero as would be expected under such conditions. If, alternatively, the ratings had been made on the basis of some single general attribute (e.g., "prestige") then the correlations among scales would have been homogeneous but nonzero and would have yielded only a single factor instead of the clearly articulated and replicated five factor structure actually obtained. It is of course still possible that the attributes of the persons *being rated* were not what the raters were using as a basis for their nominations. But if not, it is hard to imagine what sorts of alternative mechanisms could have produced the degree of differentiated consensus necessary to generate these results.

One fact does remain, however—that even groups with rather limited histories of interpersonal association (such as those composing the ROTC sample or the sample of Peace Corps Volunteers to be discussed below) produce a highly similar factor structure. This may well indicate that the "level" or "depth" of the contacts required to produce such results may not be very great. (One associate has remarked that these factors sound very much like the sorts of things she sees in letters of recommendation!) But this hardly seems a defect of a method and set of measures for assessing and determining the organization of *phenotypic* attributes of persons. To the contrary what seems somewhat surprising (and encouraging) is that relatively naive subjects come to such consensual and differentiated views of one another on so many separate dimensions on the basis of

such limited amounts and variety of inter-personal contact.

Relative to extending the class of subject populations in which these five factors recur, these analyses go little beyond what has been demonstrated previously by Tupes and Christal (1958, 1961) in their series of studies. However, results obtained from one additional sample recently tested does enlarge somewhat the scope of these findings. A sample of 55 Peace Corps Volunteers composed of 16 women and 39 men who were administered the peer ratings toward the end of their 2-month training period, yielded the same, clear five factor pattern described above. Although ratings were collected in groups composed only of like-sex subjects, for the analysis these groups were combined to yield a suitable (if minimal) sample size. Considering that the period of association for these persons had been fairly brief, that some members in each rating group had had only scant contact with the others, and that these persons felt rather strongly that they would not be able to rate each other meaningfully on these scales, the fact that the analysis did yield the same five, clearly defined factors augers well for the general utility of this method of assessment and for the invariance of this factor structure.

With regard to other behavioral correlates of these dimensions, Tupes (1957) has demonstrated the predictive validity of a certain pattern of these characteristics against criterion ratings of military officer effectiveness. These results were obtained on a large sample of Air Force Officer Candidate students highly selected initially on the basis of measured abilities and military aptitudes and despite the attenuation caused by a considerable dropout rate during the training period. More recently, Isaacson, McKeachie, and Milholland [5] have found that peer nomination measures on these characteristics account for substantial proportions of the variance in student ratings of teaching effectiveness. The subjects in these studies were graduate student teaching fellows in the introductory psychology course at the University of Michigan. Rather distinct patterns of scores on these

five factors seem to be indicative of success in these two contexts and yet a third appears optimal to account for prognostic judgments of success in the field by the Peace Corps Volunteers.

A major objective of the current research program of the author is to develop self-report or individual performance measures of these five personality factors (and for whatever additional factors might be identified subsequently). Some degree of success in this effort has been achieved to date primarily by use of self-report, forced-choice inventory methods of data collection and a newly proposed method for scoring key construction. A preliminary report of this work is currently in press (Norman, 1963).

Before concluding, it is well to note a couple of points made by Tupes and Christal (1961) concerning their findings based on the peer nomination rating data:

In many ways it seems remarkable that such stability should be found in an area which to date has granted anything but consistent results. Undoubtedly the consistency has been hidden by the inconsistency of factorial techniques and philosophies, the lack of replication using identical variables, and disagreement among analysts as to factor titles. None of the factors identified in this study are new. They have been identified many times in previous analyses, although they have not always been called by the same names . . . .

It is unlikely that the five factors identified are the *only* fundamental personality factors. There are quite likely other fundamental concepts involved among the Allport-Odbert adjectives on which the variables used in the present study were based [p. 12].

It seems to this writer on the basis of the results from previous studies and those presented here that it is time to return to the total pool of trait names in the natural language—there to search for additional personality indicators not easily subsumed under one or another of these five recurrent factors. If one were then to attempt the sort of extensive categorization of these terms described above; if he were then to collect and analyze data based on scales constructed from such terms including well defined markers from each of the five previously isolated factors in each analysis; if samples from diverse subject populations were routinely employed; and if the length, intimacy, and variety of con-

tacts subjects had had with one another prior to data collection were systematically varied —then perhaps an adequate taxonomy of personal attributes and a useful "observation language" for personality as an area of research and theory building might well result.

REFERENCES

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

ALLPORT, G. W., & ODBERT, H. S. Trait-names: A psycho-lexical study. *Psychol. Monogr.,* 1936, 47(1, Whole No. 211).

CATTELL, R. B. Confirmation and clarification of primary personality factors. *Psychometrika,* 1947, 12, 197–220.

CATTELL, R. B. *Personality and motivation structure and measurement.* Yonkers-on-Hudson: World Book, 1957.

NORMAN, W. T. Development of self-report tests to measure personality factors identified from peer nominations. *USAF ASD tech. Note,* 1961, No. 61-44.

NORMAN, W. T. Personality measurement, faking, and detection: An assessment method for use in personnel selection. *J. appl. Psychol.,* 1963, 47, in press.

TUPES, E. C. Relationships between behavior trait ratings by peers and later officer performance of USAF Officer Candidate School graduates. *USAF Personnel Train. Res. Cent. res. Rep.,* 1957, No. 57-125.

TUPES, E. C., & CHRISTAL, R. E. Stability of personality trait rating factors obtained under diverse conditions. *USAF WADC tech. Note,* 1958, No. 58-61.

TUPES, E. C., & CHRISTAL, R. E. Recurrent personality factors based on trait ratings. *USAF ASD tech. Rep.,* 1961, No. 61-97.

WILLINGHAM, W. W. Estimating the internal consistency of mutual peer nominations. *Psychol. Rep.,* 1959, 5, 163–167. (a)

WILLINGHAM, W. W. Methods for deriving standard scores for peer nominations with subgroups of unequal size. *USN Sch. Aviat. Med. res. Rep.,* 1959, Proj. No. NM 14 02 11, Sub. 1, No. 27. (b)