THE PSYCHOLOGICAL BULLETIN
Vol. 50, No. 6, 1953

# ASSESSING SIMILARITY BETWEEN PROFILES[1]

LEE J. CRONBACH AND GOLDINE C. GLESER
*University of Illinois* *Washington University School of Medicine*

A great many current investigations, particularly in clinical and social psychology, deal with similarity between profiles of test scores. Such studies vary widely with regard to the problems posed and the specific variables used, but they have in common an attempt to deal with several scores or traits simultaneously. Some investigators attempt to identify "types" of people who have similar configurations of scores. Much of so-called inverse factor analysis has this aim. Other studies attempt to differentiate clinical or occupational groups by means of patterns of test scores (e.g., **1**, **28**). In another type of problem, two or more profiles for the same person are compared. The person is assessed more than once on the same set of variables, and the consistency of the profiles is measured. This is one method used to study the validity of clinical procedures (**5**, **24**). Profile comparison also permits exploration of new variables such as self-consistency over time (**31**) and assumed similarity in perception of others (**17**).

At present many techniques are available to the investigator who is concerned with assessing the degree of profile similarity. The method most widely known among psychologists is that of correlating one profile with another, generally termed a *Q* correlation. Burt (**3**) and Stephenson (**32**) have been chiefly responsible for developing this approach.[2] Special indices have also been proposed, such as the coefficients of pattern similarity of Cattell and those of du Mas. A distance measure has been described recently by Osgood and Suci (**26**) and by the present writers (**11**).[3]

A very valuable summary of statistical literature bearing on the use of profiles or patterns to classify individuals into relatively homogeneous groups has been prepared by Hodges (**22**). Other recent reviews which deal in part with this problem are Gaier and Lee's (**20**) and Tyler's (**35**).

The various available methods of measuring profile similarity yield somewhat different results. Proper choice of a measure for a specific investigation requires knowledge of the assumptions, limitations, and information utilized in the several methods of measuring profile similarity. *It appears that the methods most often used have serious limitations. Much superior methods can be proposed.*

We intend in this paper to examine in a general way the problem of comparing sets of scores and to clarify the mathematical logic involved

[1] The study was supported under Contract N6ori-07135 between the Office of Naval Research and the University of Illinois. The first version of this paper was presented to the Midwestern Psychological Association on April 27, 1952, and a more detailed technical report on the material (11) was issued in April, 1952.

[2] Stephenson's current work on *Q* technique (33) departs from the correlational methods reviewed here. We shall not discuss here the logic of his basically new approach using analysis of variance.

[3] The work of Osgood and Suci (26), and our own work, was in large measure independent. While working on our separate problems, however, we exchanged ideas occasionally, and found our interests converging on the *D* measure. We appreciate their cooperation and that of others who have discussed our problem with us.

therein. This permits us to consider the various formulas which have been advanced in the past, and to draw attention to those approaches which seem to have greatest merit.

This paper is primarily concerned with *descriptive* indices applicable to the investigation of questions such as the following:

1. How similar are Persons 1 and 2?
2. How similar is Person 1 to Group *Y*?
3. How homogeneous are the members of Group *Y*?
4. How similar is Group *Y* to Group *Z*?
5. How much more homogeneous is Group *Y* than Group *Z*? Than the combined sample?

Comparable questions may be asked in studies concerned with two or more profiles for the same person.

While it is necessary to describe the degree of similarity between score sets in many of the investigations now being pursued, it is often equally or more important to test hypotheses such as "Group *Y* and Group *Z* can be regarded as samples from the same population" or "Individual 1 is more likely to be a member of Group *Y* than of Group *Z*." Such problems of *inferential* statistics relevant to multivariate analysis have been thoroughly studied by Fisher, Hotelling, and the Calcutta school, and several significance tests are available for normally distributed variables (**29**). We shall not discuss the inferential problems, being concerned solely with descriptive formulas for reporting *degree* of similarity.

## General Methodological Difficulties

While the procedures permitted to the investigator of profile similarity are varied, they involve numerous pitfalls. We shall discuss some of these difficulties as a preliminary to formal analysis of profile comparison methods.

*Similarity as a general quality.* Thinking of persons as "similar" or "dissimilar" is a common oversimplification. This attractive notion, however, does violence to a fundamental principle. If behavior is described in terms of independent dimensions, then persons who are similar in one dimension may be no more similar in some second dimension than persons who are dissimilar in the first dimension. *In other words, similarity is not a general quality. It is possible to discuss similarity only with respect to specified dimensions (or complex characteristics).* This means that the investigator who finds that people are similar in some set of scores cannot assume that they are similar in general. He could begin to discuss general similarity only if his original measurement covered all or a large proportion of the significant dimensions of personality. Thus any problem inquiring whether similar people perform differently from dissimilar people must be stated in terms of the question "Similar in what?" It is most unlikely that similarity in every quality has the same effect.

*Reduction of the configuration by similarity indices.* Many investigators are attracted to profile similarity studies because they believe that in this way they can take into account the entire configuration of scores. However, when we try to treat a set of scores by any of the mathematical methods now being used, we no longer study the entire configuration. Instead, by reducing the configuration or the relationship between two configurations to a single index, we discard much of the information in the score set.

We may illustrate this by referring to Gage's study of insight (**19**). He asked a teacher to predict the responses of a pupil. He scored the

predictions using the responses actually given by the pupil as a key, thus estimating the accuracy of the teacher's perception. It is obvious, however, that a more refined question could be asked regarding the teacher's ability to perceive separate aspects of the pupil. In using a total index, Gage was forced to combine these many separate aspects of insight into an over-all score. It is important that the investigator recognize the limitations of so-called global approaches even though they may be the best for him to use in initial exploration of a particular area.

*Absolute interpretation of index.* Another type of difficulty which frequently complicates interpretation of profile similarity studies is the failure to recognize that the magnitude of the similarity index has no meaning in itself. In conventional psychometrics, we would not give serious attention to the absolute value of a test score. When we compare a person to a key, the number of items on which he and the key agree is a form of correlation. We are aware that we should not interpret this raw score which reflects the difficulty of the items. Instead, we give our attention to the relative standing of the individual in some reference group. Correlations between persons, and other similarity indices, entail precisely the same problem. Too often, accustomed to interpreting correlations as absolute numbers, investigators interpret similarity indices without recognizing that they also depend upon the difficulty or popularity of the items or tests.

One often cited study by Fosberg shows the fallacy of this type of interpretation (18). Fosberg hoped to demonstrate that the Rorschach test is proof against faking. He therefore asked individuals to take the test in the normal manner, and then to take it attempting to make the best possible impression. He correlated the two psychograms for a given individual and interpreted the resulting high correlations as showing that the Rorschach was proof against faking. Now it is true that the psychogram under "fake good" conditions could be predicted from that under normal conditions. But the "fake good" psychogram could have been predicted quite well from the psychogram of some other person chosen at random. Between any two Rorschach records taken at random, there will tend to be a high correlation just because certain scores (e.g., *D*, *F*) will usually be large, and other scores (e.g., *m*, *cF*) will usually be small.

It is evident that any estimate of the similarity of particular profiles must be evaluated relative to the similarity of people in general on the measures in question. A high index of similarity between two persons might indicate that they are unusually alike, or might indicate that they possess in common only the characteristics most humans have. For example, Gage (19) considered insight to be indicated by a marked similarity between prediction and actual response. He found that a large part of the correlation between predicted response and actual response was accounted for by the teacher's ability to predict the responses of pupils in general. When the teacher was asked to predict the *average* response of pupils, the correlation with the actual responses of an individual pupil was frequently as high as when the teacher attempted to predict that particular pupil's response.

*Noncomparability of scale units.* Combining many traits into any sort of composite index, whether it be a *D* measure, a *Q* correlation, a discriminant function, or any of the

other methods presently used, involves assumptions regarding the scale of measurement which usually cannot be defended (7, 8). If, for example, one score measures intelligence and a second one reflects anxiety level, any single index based on this profile involves an assumption that one unit of intelligence is equivalent to some number of units on the anxiety scale. Such an assumption is perhaps necessary if it permits investigations which would otherwise be impossible. It may also be possible to justify the units assigned to the respective scales by a mathematical treatment which selects the weights to maximize some prediction. This is an empirical solution, however, and does not contribute directly to development of theory.

## A Generalized Concept of Pattern Similarity

We now introduce a model for the concept of similarity between persons which provides a basis for systematic discussion of the assumptions underlying most of the common measures of profile similarity.

A profile or pattern pertaining to a person consists of a set of scores. We shall use the following notation:

$j$ = any of the variates $a$, $b$, $c$, $\cdots$ which are $k$ in number;
$i$ = any one of the persons 1, 2, $\cdots$ $N$;
$x_{ji}$ = the score of person $i$ on variate $j$.

Considering only two persons, we have the set of $x_{j1}(x_{a1}, x_{b1}, \cdots x_{k1})$ for person 1, and the set of $x_{j2}$ for person 2. Without placing any restriction upon our data, we may regard the $x_{j1}$ as the coordinates of a point $P_1$ in $k$-dimensional space. The $x_{j2}$ define a point $P_2$. The more similar the measures of two individuals the closer will their points lie in the $k$-dimensional space, and, conversely, the further apart the points the more dissimilar are the corresponding measurements. Accordingly we define the *dissimilarity* of two individuals as the linear distance between their respective points.

If we represent the variables by orthogonal axes, the distance $D$ between any two points may be easily obtained by use of the generalized Pythagorean rule,

$$D_{12}^2 = \sum_{j=1}^{k} (x_{j1} - x_{j2})^2. \qquad [1]$$

$D^2$ can be used directly as a measure of similarity. In most cases, however, it is preferable to obtain $D$, since the larger differences between persons are much exaggerated in squaring. $D$ is less skewed than $D^2$ but is not normally distributed.

Formula [1] is a general expression for the dissimilarity between two profiles. It may be applied to practically any type of score set; viz., responses to a series of items, raw scores on a set of tests, profiles of deviation scores, ratings of a group of stimuli on a subjective scale, or responses in a Stephenson forced-sort procedure. While formula [1] results in a measure of dissimilarity no matter what types of scores are used, the interpretation of the results depends on the nature of the scores.

One basic decision made by the investigator is whether to work with the original score set or to convert it by centering about the person's mean or by standardizing within the person. He may make these conversions before using formula [1]. Many of the current formulas automatically introduce some such treatment of scores. Converted scores in general alter the domain within which similarity is measured and consequently alter the results.

*Elevation and scatter within profiles.* A set of $k$ scores, whether expressed in raw or standard measure, has $k$ degrees of freedom and may be considered as a configuration in $k$ space. When the profile is expressed as a set of deviations about the person's mean or when the profile is standardized within the person, the number of degrees of freedom is reduced. This has important consequences. In order to discuss them we introduce the terms *elevation*, *scatter*, and *shape*. *Elevation* is the mean of all scores for a given person. *Scatter* is the square root of the sum of squares of the individual's deviation scores about his own mean; that is, it is the standard deviation within the profile, multiplied by $\sqrt{k}$. *Shape* is the residual information in the score set after equating profiles for both elevation and scatter. We can clarify these terms by introducing numerical illustrations. Suppose that we have five traits $a$, $b$, $c$, $d$, $e$, and persons $A$, $B$, and $C$.

| | $a$ | $b$ | $c$ | $d$ | $e$ |
|---|---|---|---|---|---|
| $A$ | 2 | −2 | 0 | 3 | 2 |
| $B$ | 0 | −4 | −2 | 1 | 0 |
| $C$ | 3 | −1 | 3 | −1 | −4 |

According to formula [1], $D_{AB}^2$ is 20. $D_{AC}^2 = D_{BC}^2 = 63$.

Elevation is determined by averaging the scores for each individual. For the example above, the elevations are as follows: $A$, 1; $B$, −1; $C$, 0. Removing elevation, the individual profiles become:

| | $a$ | $b$ | $c$ | $d$ | $e$ |
|---|---|---|---|---|---|
| $A$ | 1 | −3 | −1 | 2 | 1 |
| $B$ | 1 | −3 | −1 | 2 | 1 |
| $C$ | 3 | −1 | 3 | −1 | −4 |

Now the distance between $A$ and $B$ is 0. Those persons who are different when their total profiles are taken into account are indistinguishable on the basis of their profiles of deviation scores. $D_{AC}^2$ and $D_{BC}^2$ now equal 58.

The operation of eliminating differences in elevation from the profiles is referred to by Thomson (**34**) and others as centering about persons. Geometrically, it is equivalent to projecting all persons into a $k-1$ space orthogonal to the line defined by the equations $a = b = c \cdots$. Comparison of deviation scores is involved in testing certain hypotheses regarding scatter in mental tests (**2**). Burt eliminates elevation when he obtains a matrix of covariances between profiles for use in factoring persons into types (**3**). If we use $D'$ as a symbol for distance between profiles after projection into $k-1$ space, we have the following equation:

$$D'_{12}{}^2 = D_{12}^2 - k\Delta^2 El_{12} \qquad [2]$$

Here $\Delta El$ represents the difference in elevation between the two persons. It is evident that the difference between persons has two components, one due to elevation, and one due to the remaining information in the profile. Treatment of deviation scores discards information about differences in elevation.

When differences in scatter between profiles are eliminated, the measure of similarity is reduced to a consideration of shape alone. This is accomplished by dividing each deviation score by the individual's scatter, thus standardizing the profile. Geometrically, this operation amounts to projecting every score set in $k-1$ space onto a $k-2$ hypersphere. The center of the hypersphere is at the point representing in $k-1$ space a completely flat profile.

If for each of the three persons in the above example we divide his deviation profile by his scatter, we obtain the following new profiles:

| | *a* | *b* | *c* | *d* | *e* |
|---|---|---|---|---|---|
| *A* | 1/4 | −3/4 | −1/4 | 1/2 | 1/4 |
| *B* | 1/4 | −3/4 | −1/4 | 1/2 | 1/4 |
| *C* | 1/2 | −1/6 | 1/2 | −1/6 | −2/3 |

Now $D_{AC}^2 = D_{BC}^2 = 2.25$. $D_{AB}^2 = 0$ as before.

Letting $D''$ be our symbol for the $D$ measure obtained from two standardized score sets,

$$D''^2 = \frac{D'^2 - \Delta^2 S}{S_1 S_2} = \frac{D^2 - k\Delta^2 El - \Delta^2 S}{S_1 S_2}. \quad [3]$$

Here $S$ is the scatter of an individual and $\Delta S$ is the difference in scatter. It is clear from this equation that by standardizing the profile we eliminate from consideration one further type of difference between the persons.

Elevation and scatter have commonly been eliminated in past studies of similarity between persons. It is easily shown that

$$D''^2 = 2(1 - Q) \quad [4]$$

where $Q$ is the product-moment correlation between scores. It will be recalled that in product-moment correlation, one subtracts the product of means from the cross-product terms, and divides by the standard deviations (which are proportional to the measures of scatter). In other words, all correlations between profiles are essentially measures of distance in $k-2$ space.

Equations [3] and [4] make clear that $D$ in $k$ space will, in general, not give the same result as $Q$ for a given pair of score sets, nor can $D$ be inferred from factor loadings derived from $Q$ correlations. Osgood and Suci (**26**) demonstrated close correspondence between the two sorts of measures, but only for an unusual set of data where $\Delta El$ and $\Delta S$ are small. Warrington (**36**) determined the extent to which information is discarded in various treatments by building hypothetical data from a mathematical model. For his analysis, he employed five factors, represented with varied loadings in 60 items. Each of his hypothetical persons was assigned scores, distances between persons were determined, and these were correlated with distance measures based on the factor scores. For perfectly reliable items the similarity measures correlated .92 with the criterion. This "validity" dropped to .85 when elevation was removed from the similarity index but not from the criterion. For items of moderate reliability, the validity dropped from .81 to .55 when elevation was removed.

## Relation of Other Formulas to the *D* Measure

Table 1 lists the formulas most frequently used in psychological investigations of profile similarity, together with some of their more prominent characteristics.

*Treatments in* k *space.* The $D$ measure presented in formula [1] considers all $k$ dimensions in the original data. This measure has recently been discussed by Osgood and Suci (**26**), but a quite similar formula appeared in the literature much earlier, as Pearson's "coefficient of racial likeness" (CRL) (**27**), which was developed to measure the similarity between two groups or the similarity of an individual to a group. In its original form, CRL was essentially the same as $D^2$ save that all variates were expressed in standard measure and a multiplier involving the number of cases per group was included.

The Pearson index proved unsatisfactory in the anthropological research for which it was developed. Some of the criticisms arise out of its insensitivity to differences of number of cases from group to group. These

criticisms are irrelevant to our present purpose. Morand (see Rao, **30**) notes that in some anthropological research the index has given unreasonable results for groups which were regarded as quite dissimilar, intuitively or theoretically. From the context, we judge that this difficulty is a consequence of the high weight CRL assigns to general factors among the variates. This problem may arise in measures of similarity whenever variates are intercorrelated. We shall discuss the problem of correlation in more detail later.

Cattell (6) has proposed an index $r_p$ which is like $D$ in many respects. He introduces a transformation, however, which makes the obtained index range from 1 to $-1$. In our notation,

$$r_p = \frac{K\Sigma\sigma_j^2 - kD^2}{K\Sigma\sigma_j^2 + kD^2} \qquad [5]$$

where $K$ represents twice the median $\chi^2$ corresponding to the given number of variates. $D$ or $r_p$ would give the same results so far as the *ordering* of dissimilarity is concerned.

Cattell arrived at this index because he believes that similarity should be measured by an index which is comparable to a correlation. This assumption seems to us neither necessary nor desirable. If persons fall into a multivariate normal distribution there should be very many similar pairs, and relatively few pairs who are far from each other. Furthermore, if we are dealing with variates having an unlimited range then no matter how far apart person 1 is from person 2, there is no theoretical reason why there cannot be a person 3 such that $P_1P_3 > P_1P_2$. Therefore we see no reason why the measure of separation should have a limit. "Complete dissimilarity of persons" is an undefinable concept.

Webster (**37**) proposed that intraclass $r$ might have advantages for measuring similarity in $k$ space.

$$r_{in} = 1 - \frac{D_{12}^2}{S_1^2 + S_2^2 + \frac{1}{2}k\Delta^2 El} \cdot \qquad [6]$$

The denominator in [6] is the sum of squares of scores of both persons about the grand mean of their scores. The larger this denominator the closer will $r_{in}$ approach $+1$ for pairs having the same $D$. To illustrate, con-

TABLE 1

SIMILARITY FORMULAS AND THEIR CHARACTERISTICS

| *Symbol and Proponent* | *Procedure* | *Type of Comparison* | *Remarks* |
|---|---|---|---|
| $D$ (Osgood-Suci, **26**; Cronbach-Gleser, **11**) | Distance measure | $k$ (also $k-1$, $k-2$) | A general formula |
| CRL (Pearson, **27**) | Distance measure for standardized variates | $k$ | |
| $r_p$ (Cattell, **6**) | Transformed distance measure for standardized variates | $k$ (also $k-1$) | Converts $D$ to a scale from 1 to $-1$ |
| $Q$ (Stephenson, **32**) | Product-moment correlation across variates | $k-2$ | Symbol $Q$ used here instead of $r$ for clarity |
| Rho (Spearman) | Correlation across scores ranked within a profile | $k-2$ | |
| Tau (Kendall, **25**) | Based on rank arrangements | $k-2$ | Highly correlated with rho |
| $r_{ps}$ (du Mas, **13**) | Based on tally of similarity of slope along profiles | $k-2$ | Estimate of tau based on partial data |

sider person $X$ with standard scores 1.0, 1.0, on two variates, and person $Y$ with standard scores 1.1, 1.1. For this pair, $D$ is $\sqrt{0.02}$. $S$ for each person is zero, the denominator is small, and $r_{in}$ is $-1$. In other words, this pair of persons is reported by intraclass $r$ to have maximum dissimilarity, whereas the $D$ measure reports them to be close together. The definition upon which the $D$ measure is based appears to present a more satisfactory conception of similarity than the definition embodied in the intraclass measure.

*Treatments in $k-2$ space.* Several formulas have the effect of measuring similarity in $k-2$ space. We have already noted that a $Q$ correlation based on raw scores gives the same result as obtaining $D$ from scores standardized within profiles. Correlation is thus a special case of the $D$ measure.

Measures of similarity are at times based on scores ranked from highest to lowest within the profile. The correlation of two such sets of ranks yields rho, which is thus like $Q$ in many of its properties. Sometimes rank correlations are used in the belief that assumptions regarding the test score metric are thereby avoided. This is not the case. When scores are ranked, the separation between two adjacent ranks is fixed over the whole range, forcing all profiles into the same rectangular distribution. This forcing may be justified in certain studies, but it does involve a definite assumption.

Kendall's tau is a rank correlation based on the *direction* of differences between all possible pairs of variates. Tau is very closely related to rho but is somewhat more laborious to compute. In some statistical work, it is an advantage that the sampling distribution for tau is known.

Du Mas has suggested the coefficient $r_{ps}$. Kelly and Fiske drew our attention to the fact that $r_{ps}$ is a sort of approximation to tau, in which pairs of adjacent variates only are considered. Results from du Mas' method therefore depend upon the order in which variates are listed in a profile. Different results would be obtained if some other order were used. $r_{ps}$ is biased when the arrangement of traits is not strictly random. Furthermore, it uses relatively little of the information in the profile, and is therefore inexact. $r_{ps}$ does not appear to have advantages over rho or $D''$.

*Should differences in elevation be disregarded?* A basic question is whether similarity between score sets is more meaningfully investigated by allowing differences in elevation to affect the result.

Cattell (6) and du Mas (14) have argued that differences in level between profiles are generally important and should be included in the index. For many studies, it is surely desirable not to regard two people as similar if their profiles have the same shape but differ in elevation. In the Wechsler test, for example, the elevation, being the sum of the scores, is a measure of over-all ability. The interpretation of the profile shape is dependent upon elevation. The fact that Vocabulary is higher than Digit Span means something qualitatively different for a college graduate with an IQ of 120 from what it means for a 10-year-old with an IQ of 100. To reduce the data by leaving elevation out of account may cause people to appear similar who are quite different in the domain the investigator desires to study.

On the other hand, there may be studies in which the elevation component is of no interest. If, for example, data are obtained from a personality questionnaire in which a person re-

sponds *yes* or *no* to each item, and the total score in each category is the number of questions marked *yes*, the differences in elevation between persons will be due partly to a response set (**9**). The investigator may decide that this "*yes*-saying tendency" is irrelevant to his problem, and if so, he will want to eliminate that component from his data. If he makes such a decision, reduction of the data to $k-1$ space is appropriate.

The elevation component in a profile represents the sum (or average) of all scores, and depends on the direction of scoring of the variates. A trait could be scored as "submission," for instance, instead of "dominance"; any such reversal alters the composition of the elevation score. If there is no particular reason for scoring each variate in one direction rather than the other—and this is generally the case unless variates are systematically correlated—then the elevation component is determined arbitrarily by these scoring decisions. It is highly undesirable to eliminate the elevation component when it is thus arbitrarily defined.

If a general factor is present in the variates it is often possible to choose a direction for scoring each variate which yields consistently positive intercorrelations among variates. The elevation factor, when this set of scores is used, will be heavily loaded with the first principal component of the scores, i.e., the general factor (**23**). This first factor may be an important one to consider in judging the similarity of profiles.

*In general, it appears undesirable to eliminate elevation* unless the investigator can interpret it definitely as representing individual differences in a quality which he does not wish to take into account in his similarity measure. If he is uncertain as to which is the more appropriate direction for scoring each of the variates, then the investigator should use the measure $D$ in $k$ space. Ebel (**15**), working on the problem of similarity of score sets as it is encountered in studying the reliability of rating, makes a similar recommendation. In his problem, the mean level of ratings assigned by each rater is comparable to our elevation. He lists practical considerations which make it wise at some times, and unwise at others, to consider differences in level in assessing the agreement of raters.

*Should differences in scatter be disregarded?* Any treatment which equalizes scatter of profiles before computing the difference measure is equivalent to projecting points onto the surface of a hypersphere within the $k-1$ space. This has the effect of increasing the jaggedness of profiles which are relatively flat, or, we might say, of reducing the jaggedness of profiles having a large amount of scatter. This introduces a serious difficulty. Figure 1 illustrates the fact that in projection onto the sphere differences between persons near the center are much magnified. The small $D'_{12}$ becomes a large $D''_{12}$. $D''_{34}$, however, is little greater than $D'_{34}$. Points 1 and 2, near the center of the sphere, represent persons with flat profiles. Persons who would be judged quite similar in $k$ or $k-1$ spaces are sometimes reported as markedly dissimilar in the $k-2$ measure.

Another aspect of the same problem is illustrated in Fig. 2. Any profile contains some error of measurement so that the location of the individual in $k-1$ space is only approximate. We indicate the possible positions in $k-1$ space of each individual over many trials by a cloud of points within the circle. The possible positions a person might take in $k-2$ space are then indicated by

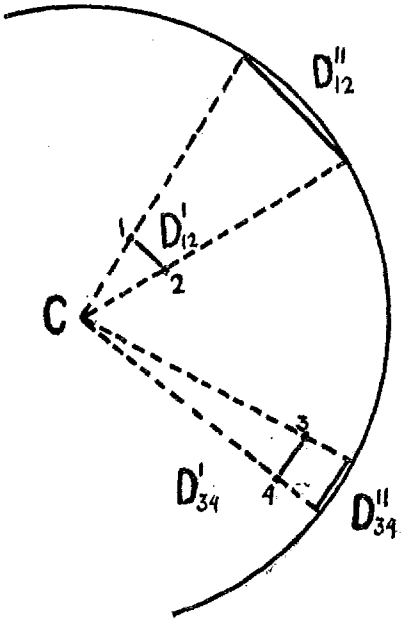


FIG. 1. MAGNIFICATION OF DISTANCES IN PROJECTION ONTO SPHERE

the distribution of points on the edge of the circle. It is clear that the greater the error, the greater the dispersion in $k-1$ and $k-2$ spaces. For a person who has a moderate amount of error and whose scatter is low, the projection in $k-2$ space has almost no meaning. On different trials he might fall anywhere in the $k-2$ space, and it is a matter of chance which persons he is similar to in a particular set of data. Either a high or low value of $Q$ can arise by chance.

Results from analysis of profiles in $k-2$ space are dependable only when scatter is large relative to the error dispersion for the individual. So long as some profiles may be expected to be flat or nearly so, treatments of these profiles in $k-2$ space will be very much influenced by random error. This difficulty is greatest when most of the variance in the $k$ scores comprising the profile is accounted for by a small number of factors. As more factors are represented in the variate set it is less likely that flat profiles will be obtained.

We must question whether the study of profiles in $k-2$ space, or more specifically whether correlation between profiles, is a justifiable line of investigation. This procedure has the disadvantage of removing the elevation factor and in addition tends to magnify error variance. In general, therefore, we would regard treatments in $k-2$ space as inferior to treatment of the data by $D$ or $D'$.

Such success as Stephenson and his followers have obtained despite these difficulties may be explained by precautions Stephenson has introduced into his design. For one thing, Stephenson has always employed a large number of variates, each one being an item describing some per-

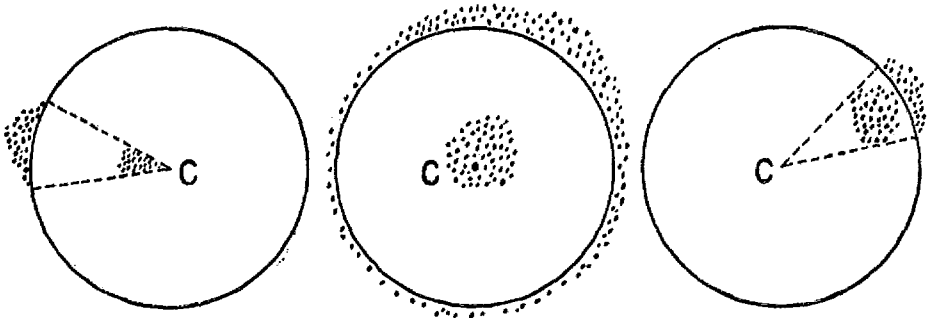


FIG. 2. EFFECT OF ERROR AND SCATTER ON THE PROJECTION ONTO A SPHERE

sonality trait. If the item intercorrelations are not generally positive, the first component removed as an elevation factor is a relatively small proportion of the total variance or information in the profiles. The part removed may be an important portion, but the $k-1$ profile still contains a great deal of useful information. The large number of variates also makes flat profiles in $k-1$ space less frequent.

In Stephenson's "balanced design questionnaire" each item is accompanied by another statement which has approximately the opposite meaning. By this device, Stephenson essentially assures that the sum over all items (i.e., elevation) departs from zero only by chance, and thus no information is eliminated from the data during the statistical elimination of the elevation component.

The magnification of error in projection to $k-2$ space will be slight if few persons have flat profiles. This can be assured by introducing items which have unequal means for the group. Then the centroid of the group will be far from the center of the sphere on which persons are projected. The difficulty with this solution is that, as the centroid of the group moves farther from the center of the sphere, persons are less differentiated in $k-2$ space, and error accounts for a larger proportion of the dispersion.

It is not surprising that most profile studies today utilize comparisons in $k-2$ space, since the problems have been conceived in terms of correlation as used to study relationships between tests. It is questionable, however, whether that model is a particularly good one. In determining the similarity between two *tests*, it is reasonable to eliminate the mean and variance from consideration. As Thomson (**34**) and Burt (**4**) have pointed out, the test mean represents its general level of difficulty for the population, while the variance is a function of the units used. Differences between tests in these values are usually quite arbitrary, depending on the choice and number of items. When we are mainly interested in the underlying relationship between tests these differences are of no importance and are neglected in the correlation formula. In dealing with similarity of *individuals*, however, it is necessary to consider rather carefully what logic is involved when individuals are equated for level and scatter.

Measures in $k-2$ space can give useful information only if both the dispersion of persons in $k-1$ space and the scatter for nearly all persons are large relative to the error dispersion. Data in $k-1$ space are required to determine whether these conditions are met. Then one can determine whether profiles in $k-1$ space are reliable, and whether there are many flat profiles. The investigator can, if he wishes, eliminate the people with flat profiles from the study. The forced-sort does not collect data on scatter, and one has no basis for judging which profiles are reliably located.

It seems quite important for those studying similarity to investigate reliability directly by obtaining two estimates for each profile. Reliability of $k-2$ space measures has ordinarily not been examined in past investigations of similarity.

In those studies where $k-2$ space measures have been used in the past, properly interpreted positive results need not be discounted. The faults to which we have drawn attention operate to obscure true relations and to make the measurement technique insensitive. This would make nonsignificant results likely in some instances where a better technique

would find more relationship. It would tend to make particular $Q$ correlations or differences between such correlations undependable and inconsistent.

In summary, our consideration of all possibilities leads us to the opinion that the most generally advisable procedure for comparing profiles is to employ $D$ in $k$ space, except where it is known that the elevation factor is saturated with a variable which it is desired not to consider.

## Contribution of Each Variate to the Similarity Measure

*The Mahalanobis distance.* A formula which we have not discussed to this point is the generalized distance measure of Mahalanobis (see Rao, 30). The Mahalanobis distance is found from the formula:

$$\mathbf{D}^2 = \sum_j \sum_{j'} \alpha^{jj'} \Delta x_j \Delta x_{j'} \qquad [7]$$

where $\alpha^{jj'}$ is the $jj'$ element of the inverse of the covariance matrix between variates within groups. We use **D** to distinguish this measure from our $D$. The Mahalanobis measure was designed for the purpose of measuring the distance between groups, rather than between individuals, but the formula can also be interpreted as related to the difference between individuals. If this is attempted, the intercorrelations of the variates for an appropriate reference group must be known.

The **D** measure is a measure of similarity in which the orthogonal components of the original set of variates are assigned *equal weight.* In other words, the complex formula presented above yields the same results as would be obtained if one factored the correlation matrix into $k$ orthogonal factors, computed the person's scores on these components, and then applied the $D$ formula to measure similarity. For variates which are standardized and uncorrelated **D** is identical to $D$.

**D** has several interesting properties. It has a known distribution function and thus forms a basis for testing the significance of a difference between groups. Moreover, **D** is closely related to Fisher's discriminant function, and particularly to the proportion of individuals classified into the wrong group by the most efficient possible discriminant function (30, p. 180). It is not, however, especially suited to the descriptive problem which we are discussing.

In any set of correlated variates, some variance is due to general qualities or factors represented in several variates, some due to meaningful factors found only in a single variate, and some due to error of measurement. In a principal-components analysis, $k$ factors will be determined but the last factors may be almost entirely due to error of measurement. The Mahalanobis measure weights unreliable and unimportant factors equally with the first few components in the variates. That is, it assumes that any $k$ variates represent $k$ equally important factors. This is undesirable in a descriptive index, since differences between individuals on factors which are not well represented in the test battery will be unstable from one trial to another, and hence **D** for individuals will be unstable. When the formula is applied to differences between groups, no such problem arises, for groups will show negligible differences on factors which consist largely of error.

*Weights in the D measure.* The interpretation of the $D$ measure is facilitated if we consider what weight it assigns to the orthogonal components underlying the variates. Some investigators have proposed that uncorrelated scores be employed in any

study of similarity. We find, however, that a meaningful interpretation can be made when $D$ is applied to correlated variates.

First, we may note that when the variates used in formula [1] are uncorrelated, they contribute to $D^2$ in proportion to their variances. Hence the investigator who standardizes his variates is assigning equal weights to them, and any difference in variances assigns greater weight to some of the tests than to others. When variates are correlated, $D^2$ is dependent not only on the relative variances of the variates used, but also on the configuration of the variates in the factor space.

In order to obtain some insight as to the weighting of factors resulting from the use of formula [1] on correlated variates, let us consider first the case in which all variates are standardized. Then $D^2$ computed from such standardized scores is identical with that obtained if one were to determine the principal axes of the test configuration, compute each individual's score on each of these components, and then weight these component scores by the square root of the latent root for that component before computing $D^2$.[4] The principal-axis solution is a method of factor analysis which removes as much of the variance as possible in each successive factor. The latent root corresponding to each factor reflects the proportion of variance that is accounted for by that factor. *Thus, $D^2$ weights factors according to their representation in the test configuration.*

When an investigator employs a group of correlated variates, the factors represented most frequently among his measures are often especially important to the problem under investigation. If the $D$ measure were applied to a Wechsler profile, for example, the general factor running through the variates would have higher weight than any more specific element found in only one or two subtests, and this might be wholly desirable.

The relatively large weight assigned to the first principal component must be considered in interpreting results even of data gathered by means of the $Q$ sort where elevation per se has been eliminated. Rogers' work will serve as a convenient example of this possible difficulty. He had a patient describe herself and her ideal by $Q$ sort before and after

[4] The following demonstration of this relationship is based on C. Harris' suggestion (21) that properties of $D$ can be studied by describing the measure in matrix notation. Let us define the matrix $S$ as the array of standardized scores of persons, where columns pertain to individuals and rows to tests. $S = FX$ where $F$ is the matrix of factor loadings of the tests obtained by the principal-axis method and $X$ represents the matrix of subjects' standard scores on the factors. Then if $F$ is nonsingular one can obtain the $X$ matrix from $X = F^{-1}S$.

Suppose, however, we weight the factor scores by the square root of the appropriate latent roots. Let $L$ signify the diagonal matrix of latent roots. Then

$$L^{1/2}X = L^{1/2}F^{-1}S$$

and

$$(X'L^{1/2})(L^{1/2}X) = (S'F'^{-1}L^{1/2})(L^{1/2}F^{-1}S)$$
$$= S'F'^{-1}LF^{-1}S.$$

Since $F'F = \mathrm{L}$, for a principal components solution,

$$(X'L^{1/2})(L^{1/2}X) = S'F'^{-1}F'FF^{-1}S = S'S.$$

Now Harris has shown that $D$ is obtained from $S'S$ by adding any two diagonal entries representing two persons and subtracting the corresponding off-diagonal entries. Performing this operation on the matrix $X'L^{1/2}L^{1/2}X$ gives the same result as an operation on $S'S$ itself. Therefore $D$ from factor scores weighted by square roots of latent roots is identical to $D$ from standard scores on tests. If Harris' operation were performed on the matrix $X'X$, the result would be the Mahalanobis measure **D**.

therapy (**31**). He found that the pre- and posttherapy selves were not highly similar, that the two ideals were closely related, and that the $Q$ correlation between self and ideal was increased after therapy. This might be interpreted as a change in the structure or configuration of personality. If, however, many of the items express a general "adjustment" factor, then there is a strong common bipolar factor running through the items. This factor will have large weight in the $Q$ correlation. We therefore cannot be sure whether the results in Rogers' study are due to *configurational* changes in the personality of his subject, or due merely to her increased willingness to describe herself as well-adjusted.

The recognition that the $D$ measure allows greater weight to factors which are represented more strongly in the score set emphasizes the importance of choosing the original variate set with care (**16**). In studies where the variates are assembled as a random collection of items, there is considerable danger that the weights assigned to the various psychological components will not be fully appropriate.

In our discussion to this point we have assumed that variates are standardized. In Wechsler profiles, for instance, this is accomplished by the use of a standard score scale for each subtest. In the majority of investigations of profile similarity, similarity has been determined from raw scores on tests or items. The contribution of each principal component to $D^2$, when unstandardized variates are used, is proportional to the corresponding latent root of the covariance matrix between variates. This means that the contribution of any component to the $D$ measure depends upon the number of variates in which it appears, its loading in those variates, and the variance of the tests in which it appears.

In many studies the first principal component will have a weight substantially greater than that for the remaining components. While the investigator may be willing to let the weights on the lesser components fall out by chance, he may have a specific reason for desiring to reduce the weight given to the outstanding first component. In a study of the similarity of persons in the domain of adjustment, for instance, he may wish to group people more nearly according to the character of their complaints than according to their degree of adjustment. This degree of adjustment is likely to loom large as a factor in a set of adjustment measures, however. We therefore suggest the possibility of computing an elevation score for each person, and determining a new measure $D_w$:

$$D_w^2 = D^2 - k(1 - w)\Delta^2 El. \quad [8]$$

Here, the weight $w$ can range from zero to 1, with the extreme values yielding $D'^2$ and $D^2$, respectively.

Before leaving this subject, we should note that the weights of variates in $D'$ are proportional to the contributions of the principal components to the variate set after the elevation factor is eliminated. The elevation factor is usually very nearly the same as the first principal component if variates are positively intercorrelated. The transformation of data to eliminate scatter, which is involved in treatments in $k-2$ space, produces substantial alterations in the intercorrelation of variates. For this reason, the factors which account for most of the variance in $k$ and $k-1$ space may not be the same as the principal components in $k-2$ space.

Our recommendation on the basis of all the foregoing considerations is

that the investigator may properly use $D$ or $D_w$, whether variates are correlated or not. He should give careful thought to the question of whether or not to standardize variates. In many studies of similarity it is probably desirable to perform a factor analysis on the matrix of correlations or covariances among tests before studying similarity of persons. This permits the investigator to select his set of traits or their weighting on a more intelligent basis than he could without the factor analysis.

*Cluster scoring.* It may often be desirable to employ many items to measure a much smaller number of traits. This is the plan used in assembling items for many tests (e.g., Kuder, Guilford-Martin). Consideration should therefore be given to special problems arising for such a set of items. A particularly important question is whether the *items* should be treated as variates in the $D$ measure, or whether scores on *clusters* of items (i.e., subtests) should be used.

When assembling groups of items to measure particular traits it is difficult for the investigator to make sure that these traits will have the desired weight in the $D$ measure based on item scores. The principal components of the items will not be the same as the intended traits. Each trait will be a complex and unknown combination of the principal components. Its weight will depend highly on the choice of items and their particular factor structure.

The investigator has several possible procedures which may help him to approach the desired weights. Stephenson has suggested constructing items which systematically sample the domain of traits under consideration (**32**). If this sampling were perfect, he would insure uniform coverage of the domain so that the traits would be uniformly weighted. This approach is likely to succeed only if the item writer has more knowledge of the factorial structure of personality items than is presently available. Another solution is to perform a factor analysis on the set of items, then rotate to the desired factor solution and obtain trait scores on which to compute similarity measures. This, however, is generally impractical.

In some cases a more practical solution is to combine items into groups or clusters and obtain subtest scores for each person. Such cluster scoring is feasible only when there is a logical or statistical basis for combining items. Cluster scoring may be based on a priori grouping of items, but these groupings should be analyzed for internal consistency. From the matrix of intercorrelations of the pool of items, it would be possible to assign items to relatively homogeneous subtests (**12**).

$D$ based on cluster scores weights the underlying components of the items differently from $D$ based on the original items. In the cluster distance, the element common to the several items is given greater weight than it has when the distances on the separate items are combined. The sum of a group of items gives relatively great weight to factors present in more than one item (**10, 23**). If specific factors each present in only one item are not especially important, cluster scoring reduces their combined weight in order to give greater weight to the common element running through a whole group of items. To give a specific illustration, a score on hypochondriasis or health adjustment based on a number of items will give great weight to a general tendency to claim somatic symptoms. It will give less weight

than the item scores to specific symptoms such as a tendency to have colds or to have headaches.

In the same manner that cluster scoring reduces the weight given to specifics, it also reduces the weight given to differences between persons arising from error of measurement. Hence cluster scores, and similarity measures based on them, will be more reliable than scores based on the items. Warrington (36), with his hypothetical data, has confirmed this greater dependability of cluster-scored profiles. For one particular criterion, for instance, using $Q$-sort data, he found these validity coefficients:

$D$ measure based on items as variates, perfect item reliability .70
$D$ measure based on clusters as variates, perfect item reliability .74
$D$ measure based on items as variates, moderate item reliability .18
$D$ measure based on clusters as variates, moderate item reliability .66

It is apparent that cluster scoring overcomes much of the loss of information due to item unreliability. Stephenson is now essentially using cluster scoring in his analysis of variance based on the $Q$ sort (33).

Cluster scoring has an interesting effect on data gathered by means of a $Q$ sort. In this case even though individuals cannot differ in scatter over the total set of items, their subtest profiles can differ widely in scatter. Thus it is possible for some persons to have flat cluster profiles and others to have a high degree of scatter. This results because cluster scores utilize considerably fewer degrees of freedom than are implied in the item profile.

## Some Short-Cut Formulas

In the course of our investigation, we have discovered the possibility of developing short-cut formulas for studying groups of persons. These are not entirely satisfactory, because they are based on the average of $D^2$ over a set of pairs. In general, $D$ provides a better metric than $D^2$ for studying similarity, since large distances are much magnified in squaring. The following formulas may nonetheless be useful as a first rapid way of answering questions about groups. The formulas also provide insight into the nature of distance measures, since factors which increase mean $D^2$ will also in general increase mean $D$ and median $D$. The formulas are particularly useful as a tool for checking computations.

In any group, the mean distance between persons over all pairs of persons in the sample is

$$\overline{D_{ii'}{}^2} = 2\,\frac{N}{N-1}\sum_j V_j. \qquad [9]$$

$V_j$ is the variance, equivalent to $\sigma_j{}^2$. This is an expression for the homogeneity of a group or its dispersion. If we take one-half the mean $D^2$ within the group, we obtain the mean dispersion (distance squared) from the centroid of the sample.

The average $D^2$ of an individual $i$ from other members of this Group $Y$, is obtained from

$$\overline{D_{ii'}{}^2} = \frac{N}{N-1}\,(O_Y P_i{}^2 + \sum V_j). \qquad [10]$$

Here $i'$ varies over all other persons in Group $Y$, $O_Y$ is the centroid of the sample, and $O_Y P_i$ is the distance from $i$ to this centroid. $O$ has the coordinates $\bar{x}_j$, the average for $j$ in Group $Y$. If $i$ is not a member of $Y$, the coefficient $N/N-1$ is dropped to get the average $D^2$ from $i$ to all members of $Y$.

The average $D^2$ between members of two groups, that is, the average when each member of one group is paired with every member of the other is

$\overline{D_{ii'}^2} = \overline{O_Y P_i^2} + \overline{O_Z P_{i'}^2} + O_Y O_Z^2 (i = 1, 2, \cdots N_Y; i' = 1, 2, \cdots N_Z)$. [11]

Here we see the average cross-similarity as made up of three components: squared distance between group means, dispersion within the first group, and dispersion within the second group.

## Conclusions

Studies of similarity between sets of scores have used a large number of techniques for assessing similarity. The most satisfactory model appears to be to conceive of the tests as coordinates, and each person's score set as a point in the test space. Then distances between points, computed by the $D$ measure, are an index of similarity between score sets. This measure is a general one, to which other common techniques such as $Q$ correlation can be related. These other techniques frequently disregard or distort some of the information in the data, in ways which may be undesirable in a particular study.

The investigator of similarity must give particular attention to his choice of variates. The similarity measure depends on the content of the variate sets, on the scales used for measuring the variates, on the choice among possible similarity indices, and upon the decision whether to score separate variates or clusters of variates (i.e., subtests). The similarity index gives especially large weight to the first principal component among the scores or items, and therefore may be relatively insensitive to the shape or configuration of profiles. On the other hand, techniques which leave the elevation of the profile out of account are usually undesirable. A formula for a weighted similarity index is offered to reduce any overemphasis on the first component.

Many commonly used operations, including the $Q$ sort and product-moment correlation between persons, ignore differences in scatter between profiles. It is not generally desirable to do this, especially because if any profiles are relatively flat, the similarity indices involving them will be highly unreliable. The loss of information about differences in scatter may also be undesirable on theoretical grounds.

It is most important that any investigator understand the assumptions and limitations of whatever technique he employs to study similarity. Different treatments will yield different conclusions. In many studies, the most appropriate technique will be to apply the formula for $D$ or $D_w$ to profiles based on clusters of items.

Profile research is necessarily faced with many difficulties. In spite of these, it is our hope that the adoption of techniques which include as much information as the data provide, and which do not introduce additional errors of their own, will permit studies of similarity to advance psychological knowledge.

## References

1. Barnette, W. L. Occupational aptitude patterns of selected groups of counseled veterans. *Psychol. Monogr.*, 1951, 65, No. 5 (Whole No. 322).
2. Block, J., Levine, L., & McNemar, Q. Testing for the existence of psychometric patterns. *J. abnorm. soc. Psychol.*, 1951, 46, 356–359.
3. Burt, C. L. Correlations between persons. *Brit. J. Psychol.*, 1937, 28, 59–96.
4. Burt, C. L. *The factors of the mind.* London: Univer. of London Press, 1940.
5. Caldwell, Betty McD., Ulett, G. A., Mensh, I. N., & Granick, S. Levels of data in Rorschach interpretation. *J. clin. Psychol.*, 1952, 8, 374–379.

6. CATTELL, R. B. $r_p$ and other coefficients of pattern similarity. *Psychometrika*, 1949, **14**, 279–298.
7. CATTELL, R. B. On the disuse and misuse of P, Q, $Q_s$, and O techniques in clinical psychology. *J. clin. Psychol.*, 1951, **7**, 203–214.
8. CATTELL, R. B. The three basic factor-analytic research designs—their interrelations and derivatives. *Psychol. Bull.*, 1952, **49**, 499–520.
9. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
10. CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.
11. CRONBACH, L. J., & GLESER, GOLDINE. Similarity between persons and related problems of profile analysis. Urbana: Univer. of Illinois, 1952. Tech. Report No. 2, under contract N6ori-07135 with the Bureau of Naval Research (Mimeo.) American Documentation Institute, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., Document No. 3921, $2.75, microfilm; $7.50, photostats.
12. DUBOIS, P. H., LOEVINGER, JANE, & GLESER, GOLDINE C. The construction of homogeneous keys for a biographical inventory. Human Resources Research Center, *Research Bulletin*, 1952, 52–18.
13. DU MAS, F. M. A quick method for analyzing the similarity of profiles. *J. clin. Psychol.*, 1946, **2**, 80–83.
14. DU MAS, F. M. On the interpretation of personality profiles. *J. clin. Psychol.*, 1947, **3**, 57–65.
15. EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 407–424.
16. EYSENCK, H. J., Personality. In C. P. Stone (Ed.), *Annual Review of Psychology*. Vol. 3. Stanford: Annual Reviews, 1952. Pp. 151–174.
17. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.
18. FOSBERG, I. A. An experimental study of the reliability of the Rorschach technique. *Rorschach Res. Exch.*, 1941, **5**, 72–84.
19. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
20. GAIER, E. L., & LEE, MARILYN C. Pattern analysis: the configural approach to predictive measurement. *Psychol. Bull.*, 1953, **50**, 140–148.
21. HARRIS, C. W. Note on profile similarity. Unpublished manuscript.
22. HODGES, J. L., JR. Discriminatory analysis: I. Survey of discriminatory analysis. USAF School of Aviation Medicine, Randolph Field, Texas. 1950.
23. HOLZINGER, K. J. Factoring test scores and implications for the method of averages. *Psychometrika*, 1944, **9**, 257–262.
24. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.
25. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
26. OSGOOD, C. E., & SUCI, G. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
27. PEARSON, K. On the coefficient of racial likeness. *Biometrika*, 1928, **18**, 105–117.
28. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.
29. RAO, C. R. Tests of significance in multivariate analysis. *Biometrika*, 1948, **35**, 58–79.
30. RAO, C. R. The utilization of multiple measurements in problems of biological classification. *J. roy. stat. Soc., Sec. B.*, 1948, **10**, 159–203.
31. ROGERS, C. R. The case of Mrs. Oak—a research analysis, *Studies in client-centered psychotherapy*. Psychological Service Center Press, Washington, D.C., 1952, 47–165.
32. STEPHENSON, W. A statistical approach to typology; the study of trait-universes. *J. clin. Psychol.*, 1950, **6**, 26–38.
33. STEPHENSON, W. Some observations on *Q* technique. *Psychol. Bull.*, 1952, **49**, 483–498.
34. THOMSON, G. *The factorial analysis of human ability*. (4th Ed.) London: Univer. of London Press, 1950.
35. TYLER, F. T. Some examples of multivariate analysis in educational and psychological research. *Psychometrika*, 1952, **3**, 289–296.
36. WARRINGTON, W. G. The efficiency of the Q-sort and other test designs for measuring the similarity between persons. Unpublished doctor's dissertation, Univer. of Illinois, 1952.
37. WEBSTER, H., A note on profile similarity. *Psychol. Bull.*, 1952, **49**, 538–539.

*Received February 7, 1953.*