RAGLAND, Robert Eugene, 1928-
ON SOME RELATIONS BETWEEN CANONICAL CORRELATION, MULTIPLE REGRESSION, AND FACTOR ANALYSIS.

The University of Oklahoma, Ph.D., 1967
Psychology, general

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# ON SOME RELATIONS BETWEEN CANONICAL CORRELATION, MULTIPLE REGRESSION, AND FACTOR ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

ROBERT EUGENE RAGLAND

Norman, Oklahoma

1967

# ON SOME RELATIONS BETWEEN CANONICAL CORRELATION, MULTIPLE REGRESSION, AND FACTOR ANALYSIS

APPROVED BY

William R. Hood

Wm M. Truesdale

John F. McCoy

Dorothy A. Foster

W B Lemmon

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

The author expresses his gratitude to Dr. William Hood, who served as his major professor during the final stages of this dissertation, and Drs. Thomas Ray and William Trousdale, each of whom served as major professor during the earlier stages, and to the other members of the dissertation committee, Drs. W. B. Lemmon, Dorothy Foster, and John McCoy.

Others to whom the author is indebted are: Paul Costiloe, who introduced him to canonical correlation analysis; to Dr. Clinton Miller, who served for a time on the committee and who was especially helpful with the algebraic portions of the manuscript; to Cara Mitchell Knittel, who with Mr. Costiloe developed a computer program for canonical correlation analysis, who puzzled with the author over possible algebraic relationships, and who suggested some fruitful lines of investigation; and to Dr. Mervin Clark, Chief of Medicine and Research, Experimental Therapeutics Unit, University of Oklahoma Medical Center and Central State Griffin Memorial Hospital, who encouraged the author to pursue the present work in the context of a drug research program. The major part of this dissertation was done under the auspices of United States Public Health Grant MY-4260.

TABLE OF CONTENTS

LIST OF TABLES

# ON SOME RELATIONS BETWEEN CANONICAL CORRELATION, MULTIPLE REGRESSION, AND FACTOR ANALYSIS

## CHAPTER I

## INTRODUCTION

### General Introduction

Two important and somewhat overlapping research areas in psychology are prediction and the exploration of relationships among variables. The term "prediction" as used here can refer either to the anticipating of future events or to the dependence of one variable on another in a contemporaneous sense. The relationships among variables may be studied in terms of the variables as observed, or in terms of factors which are presumed to underly them and which may be expressed as linear combinations of them. Only linear prediction and linear relationships will be considered in this dissertation.

The statistical models most often used for prediction are the simple and multiple regression models, which assume that the predictor or predictors have only certain fixed values rather than being free to vary (Johnson & Jackson, 1959). Other linear models which do not have this restriction and which involve a single criterion (Graybill, 1965) will not be dealt with here. The direction of prediction in the regression model is one way, so that if y is predicted from x one cannot also predict x from y.

The models used to explore mutual relationships among variables are the bivariate correlational and the multivariate factor analytic models, which assume that all variables are free to vary and all pairs of relationships bi-directional. The correlation model expresses the relation between x and y in terms of equations enabling one to predict x from y or y from x. The factor analytic model is concerned with relations among several variables, but does not aim explicitly at predicting any one of them from the others.

Both correlation and regression have to do with relating or predicting in terms of observed scores or measures, while factor analysis seeks more fundamental and conceptually simpler relationships which underly those actually observed.

Canonical correlation analysis (Hotelling, 1935, 1936) is a statistical technique which combines some of the features of regression, correlation, and factor analysis. It permits the investigator, within the same analysis and assuming only one model, both to predict and to explore relationships among variables. Several predictors and several criteria, all free to vary, can be analyzed simultaneously, and relationships within and between predictor and criterion sets exhibited and studied. Although canonical correlation analysis has been available to researchers for thirty years, it has been little used, perhaps because of its seeming complexity and because of a lack of general appreciation of just how useful a research tool it can be.

This dissertation will be concerned with canonical correlation both from the mathematical and empirical standpoints. On the mathematical side, the purpose is to explain canonical correlation analysis by relating and comparing it to simple and multiple regression and correlation and to factor analysis. This purpose will be achieved through the exhibiting of certain algebraic relationships and through the use of an illustrative problem involving artificial data with known properties. In the process of explaining canonical correlation analysis in terms of the more familiar statistical techniques, new light will be thrown on those techniques as well.

On the empirical side, canonical correlation analysis will be applied to real data in which the underlying relationships (factor structure) are not clearly known. The understanding gained from the consideration of the algebraic demonstrations and the illustrative problem will then be brought to bear on the interpretation of the results of an analysis of real data.

To avoid misunderstanding at the outset, it must be pointed out that there is a lack of uniformity of terminology in the literature. Canonical correlation analysis has been referred to by at least one author (Koons, 1962) simply as "canonical analysis," while another author (Seal, 1963) uses that label to refer to a somewhat different procedure. Bartlett (1947) and McKeon (1962) use "canonical analysis" to refer to a very generalized approach to multivariate analysis which includes the usual canonical correlation analysis as one aspect. In this dissertation, "canonical correlation analysis" will refer only to the procedure developed by Hotelling for studying the relations between two sets of variables, and "canonical analysis" in the sense meant by Bartlett will not be taken up. There will be no treatment of Rao's (1958) "canonical factor analysis." "Factor analysis" will refer only to the principal components method of analysis (Hotelling, 1933) which employs ones in the diagonal of the correlation matrix and which yields orthogonal factors. There will be no consideration of the issue of rotation of factor axes. The following abbreviations will be used hereafter in the text: CCA = canonical correlation analysis, MRA = multiple regression analysis, FA = factor analysis.

## Mathematical Introduction

Before reviewing the literature on CCA, it is essential for understanding the issues involved first to describe in a general way how it represents an extension of the more familiar correlation and regression techniques and how it relates to FA. In this section, for the sake of clarity and continuity, no proofs will be offered for the descriptive and interpretative comments which are made. Such comments will receive empirical and mathematical support in later sections.

### Simple Correlation and Regression

Where there are only two variables to consider, the relationship between them may be summarized using simple correlation or regression. With either model, the linear expression relating y to x (or for predicting y from x), when both x and y are in standard score form, is:

(1) $\hat{Y}_i = r_{yx}x_i$.

Here $\hat{Y}_i$ is the predicted y score for the *i*th person, and though it is derived from standardized variables, it is not itself in standard form. It has a mean of zero and a variance equal to $r^2_{yx}$.

In this chapter of the dissertation, observed variables such as x and y will be symbolized by lower-case letters to indicate that they are in standard form. Derived variables such as $\hat{Y}_i$, which are linear combinations of x or y, will not ordinarily, as derived, be in standard form and will be symbolized by upper-case letters. When such derived variables are deliberately standardized, this fact will be noted and they will then be referred to in lower-case.

In simple correlation and regression involving x and y scores, $r_{yx}$ is both a regression weight and a correlation coefficient. As a regression weight it reflects the rate of change in $\hat{Y}$ with respect to x, and as a correlation coefficient it reflects the strength of the relationship between y and x. That is, $100r^2_{yx}$ gives the percentage of variance in y accounted for by its relationship to x. The test for determining whether $r_{yx}$ differs significantly from zero is the same whether the correlation or the regression model is assumed.

Under the correlation model, an equation similar to (1) can be written for predicting x from y. Under the regression model, there is no point in predicting x from y, since the x values are fixed.

For the purposes of this dissertation, and perhaps generally, it is convenient to make a change in notation and rewrite equation (1) as:

(2) $P_i = r_{yx}x_i.$

Rather than talking about $\hat{Y}$ as the predicted value of y, it will facilitate the discussion to talk about P as the simple weighted predictor of y. Though there is no $\wedge$ over P, it is understood to be an estimate. Like $\hat{Y}$, P has a mean of zero and a variance of $r^2_{yx}$.

## Multiple Correlation and Regression

To move to a more complex situation, suppose that one has data on four variables, $x_1$, $x_2$, $y_1$, and $y_2$, and that one wishes to develop formulas for predicting the y's from the x's. The simplest approach is to predict each y variable separately from the x variable which correlates most highly with it. If $x_1$ is the predictor which correlates most highly with $y_1$, and $x_2$ correlates highest with $y_2$, the equations are:

(3a) $P_{i1} = r_{y_1x_1}x_{i1}$, where $P_{i1}$ predicts $y_{i1}$, and

(3b) $P_{i2} = r_{y_2x_2}x_{i2}$, where $P_{i2}$ predicts $y_{i2}$.

It will often be the case that the prediction of the y's can be improved by using MRA rather than simple regression or correlation. Whereas the latter make use of <u>simple</u> weighted predictors, MRA makes use of composite weighted predictors. That is, $x_1$ and $x_2$

are first weighted and combined to form a predictor of $y_1$, and then recombined with different weights to predict $y_2$:

(4a) MRA: $P_{i1} = a_{11}x_{i1} + a_{21}x_{i2}$, where $P_{i1}$ predicts $y_{i1}$, and

(4b) MRA: $P_{i2} = a_{12}x_{i1} + a_{22}x_{i2}$, where $P_{i2}$ predicts $y_{i2}$.

The "MRA" prefix is to make clear that these are multiple regression equations. Later equations having this same general form will be prefixed "FA" and "CCA". In this way it is not necessary to use different sets of symbols when talking about the different methods of analysis.

The correlation between $P_1$ and $y_1$ is called a multiple correlation and is here symbolized as "mult $R_1$." The weights $a_{11}$ and $a_{21}$, called standard partial regression weights, are chosen so that the error in predicting $y_1$ from $P_1$ is a minimum and thus mult $R_1$ is a maximum. Unlike the situation in simple correlation and regression, these weights are <u>not</u> in general also interpretable as correlation coefficients. In the special case where $x_1$ and $x_2$ are unrelated, the weights are both regression and correlation coefficients:

(4c) If $r_{x_1x_2} = 0$, then $a_{11} = r_{x_1y_1}$, $a_{21} = r_{x_2y_1}$,

$$a_{12} = r_{x_1y_2}, \text{ and } a_{22} = r_{x_2y_2}.$$

Mult $R_2$ is the correlation between $P_2$ and $y_2$, with $a_{12}$ and $a_{22}$ chosen so that mult $R_2$ is a maximum. Although the mult R's are called correlation coefficients, they do not imply a two-way relationship as in the case of simple correlation, since $x_1$ and $x_2$ have only certain fixed values and do not vary as a function of the y's.

$P_1$ and $P_2$ are not in standard score form, but have means of zero and variances equal to mult $R_1^2$ and mult $R_2^2$ respectively (proof in Appendix I). They may be standardized by dividing the weights through by the respective mult R's as follows:

$$\text{(5a)} \quad \text{MRA:} \quad p_{i1} = \frac{a_{11}}{\text{mult } R_1} x_{i1} + \frac{a_{21}}{\text{mult } R_1} x_{i2},$$

$$\text{(5b)} \quad \text{MRA:} \quad p_{i2} = \frac{a_{12}}{\text{mult } R_1} x_{i1} + \frac{a_{22}}{\text{mult } R_1} x_{i2},$$

or more simply, designating the transformed coefficients by v's:

$$\text{(5c)} \quad \text{MRA:} \quad p_{i1} = v_{11}x_{i1} + v_{21}x_{i2},$$

$$\text{(5d)} \quad \text{MRA:} \quad p_{i2} = v_{12}x_{i1} + v_{22}x_{i2}.$$

The v's are partial regression weights which are appropriately applied to standard $x_1$ and $x_2$ scores, just as were the a's. The v's, however, are scaled so that the composite predictors are in standard form (and $p_1$ and $p_2$ are therefore now in lower-case letters). Such a scaling of regression weights is sometimes called "normalizing" (Koons, 1962), though "normalizing" may have other meanings (Anderson, 1958). Whatever it is called, such scaling, being a linear transformation, does not affect the multiple correlation between the composite predictor and the criterion.

In MRA, standardizing the composite does not seem to be too commonly done, but in CCA it has been done by some authors and not by others, with resultant confusion when comparing different studies or different mathematical presentations. Standardizing the composite offers the same advantages as standardizing any other type of score: it simplifies the algebra and it puts all scores on a common scale, making possible comparisons within and between individuals across different types of scores.

The requirement in MRA that the x values be fixed often proves to be a severe restriction in practice, and this aspect of the model is often ignored. The x's are treated as if they were free to vary, with the result that the statistical tests for the mult R's and regression weights are rendered questionable. As will be seen, in CCA the x's are truly free to vary.

## Factor Analysis

Both in the simple and multiple cases, correlation and regression have to do with observed relationships rather than any presumably more basic ones. The factor analytic approach to studying the covariations among $x_1$, $x_2$, $y_1$, and $y_2$ is not designed explicitly to predict any one of them from the others, but rather to show how each variable is composed of a weighted sum of factor variables common to all of them. It may also be employed to produce a set of factor scores for each individual in the sample, though this aspect of FA is not as well-known. These two aspects of FA will now be considered separately.

<u>Observed variables as linear function of factor variables</u>. If

it were known, for example, that each of the x variables reflected both verbal and motor ability on the part of each individual, and if it were known just how much of each type of ability was involved in each of the x's, one could write:

(6a) FA: $x_{i1} = f_{11}p_{i1} + f_{12}p_{i2}$, and

(6b) FA: $x_{i2} = f_{21}p_{i1} + f_{22}p_{i2}$.

That is, the score of the ith person on each of the x variables is a linear combination of his standing on the factor variables $p_1$ and $p_2$. Verbal and motor ability are assumed to be uncorrelated in the population and the p's, as estimates of these abilities, are referred to as orthogonal factor variables or factor scores. Note that the p's are in standard score form.

There is no explicit error term in the FA model here described. If x is a function of a large number of factors, it may be that the last factor or factors will be dismissed as "error." This matter will be considered further in a later chapter.

The f's in (6a) and (6b) are factor loadings. A loading is a product-moment correlation between one of the original (x) variables and one of the factor (p) variables. For example, $f_{11}$ is the correlation between $x_1$ and $p_1$, and $f_{11}^2$ shows what proportion of variance in $x_1$ is accounted for by individual differences in verbal ability. Similarly, $f_{12}$ is the correlation between $x_1$ and $p_2$, and $f_{12}^2$ shows how much of the $x_1$ variance is accounted for by $p_2$. If $f_{11}^2 + f_{12}^2 = 1$, then all the variation in $x_1$ is explicable in terms of the two factors.

The variation in $x_2$ can likewise be explained as a result of its loadings on the two factors. Further, the correlation between $x_1$ and $x_2$, which is an index of their common variation, is equal to the sum of crossproducts of the respective factor loadings:

$$\text{(6c)} \quad \text{FA:} \quad r_{x_1 x_2} = (f_{11})(f_{21}) + (f_{12})(f_{22}).$$

The observed variables could be said in this case to have a factor complexity of order two, since two common factors underly each of them.

<u>Factor variables as linear functions of observed variables</u>. The above equations, (6a) and (6b), are such that the observed variables are expressed as explicit functions of the factor variables. It is possible to invert the expressions so that the p's appear as explicit functions of the x's:

$$\text{(7a)} \quad \text{FA:} \quad p_{i1} = v_{11}x_{i1} + v_{21}x_{i2},$$

$$\text{(7b)} \quad \text{FA:} \quad p_{i2} = v_{12}x_{i1} + v_{22}x_{i2}.$$

The v's are partial regression weights, scaled so that the p's will be in standard form. Note that except for the "FA" prefix, these equations are identical to (5a) and (5b). The "FA" prefix means that while the x's may have the same values in both pairs of equations, the v's will not, and the p's will have quite different meanings. The p's in FA are estimates of pure orthogonal factor measures, while the p's in MRA may be quite complex factorially and are not generally orthogonal to one another. Thus the composite scores in (7a) and (7b) are conceptually simpler than those in (5a) and (5b)

Matching factors from separate analyses. Considering now the y variables, assume for the moment that they too reflect verbal and motor ability and that the loadings are known. (The fact that these assumptions cannot really be made for either the x's or the y's will be taken up later.) Then one could write:

(8a) FA: $y_{i1} = g_{11}q_{i1} + g_{12}q_{i2}$,

(8b) FA: $y_{i2} = g_{21}q_{i1} + g_{22}q_{i2}$.

Here the g's are loadings and the q's are estimates within the y set of pure verbal and pure motor ability. The explicit equations for the factor scores are:

(9a) FA: $q_{i1} = w_{11}y_{i1} + w_{21}y_{i2}$,

(9b) FA: $q_{i2} = w_{12}y_{i1} + w_{22}y_{i2}$,

where the w's are partial regression weights which standardize the q's and the correlation between $q_1$ and $q_2$ is zero. The y's and the x's in these FA equations are not fixed, but are free to vary.

Both $p_1$ from (7a) and $q_1$ from (9a) are here assumed to be estimates of pure verbal ability in the same group of subjects. In the ideal case the correlation between $p_1$ and $q_1$ would be identically one. In actuality it will be less than one, but will furnish some information concerning the extent to which the verbal factor does saturate the two sets of variables (the x set and the y set).

Similar remarks apply to $p_2$ and $q_2$ with respect to the assumed motor factor.

In the example it is assumed that the factors are known, and that the same two factors underly the x and y sets. In practice the factors are not known, but are inferred after the analysis from an examination of the loadings. It is not necessarily the case that the same factors will inhere in both sets, and if they do they may not appear in the same order when the sets are analyzed separately. As will be seen in the review of the literature which follows, matching up pairs offactors from separate analyses is an important problem in FA.

## Canonical Correlation Analysis

CCA is like FA in that it breaks the x and y sets into orthogonal factors, but it does this simultaneously rather than in two separate analyses, and it produces matched pairs of p and q composites. CCA is like MRA in that it maximizes the correlation between the predictor and criterion, but in CCA <u>both</u> the predictor and criterion are composites, while in MRA only the predictor is a composite. CCA directly yields equations such as the following:

$$\text{(10a)} \quad \text{CCA:} \quad p_{i1} = v_{11}x_{i1} + v_{21}x_{i2}, \quad q_{i1} = w_{11}y_{i1} + w_{21}y_{i2},$$
$$r_{p_1q_1} = \text{maximum};$$

$$\text{(10b)} \quad \text{CCA:} \quad p_{i2} = v_{12}x_{i1} + v_{22}x_{i2}, \quad q_{i2} = w_{12}y_{i1} + w_{22}y_{i2},$$
$$r_{p_2q_2} = \text{maximum};$$

$$\text{(10c)} \quad \text{CCA:} \quad r_{p_1p_2} = r_{q_1q_2} = r_{p_1q_2} = r_{p_2q_1} = 0.$$

The "CCA" prefix distinguishes these equations from previous MRA and FA equations of the same general form and using the same symbols. Here the v's and w's are partial regression weights which are scaled so as to standardize the p's and q's.

CCA as developed by Hotelling did not include any consideration of factor loadings, and perhaps as a result a number of users of CCA have interpreted the v and w weights as if they were loadings. It is possible, however, to calculate loadings by solving (10a) and (10b) explicitly for the x's and y's:

(11a) CCA: $x_{i1} = f_{11}p_{i1} + f_{12}p_{i2}$, $y_{i1} = g_{11}q_{i1} + g_{12}q_{i2}$;

(11b) CCA: $x_{i2} = f_{21}p_{i1} + f_{22}p_{i2}$, $y_{i2} = g_{21}q_{i1} + g_{22}q_{i2}$.

The f's are correlations between the x's and the p's, and the g's are correlations between the y's and the q's. It is also possible to obtain the correlations of the x's with the q's and the y's with the p's: these are loadings between sets. Both kinds of loadings can be useful in evaluating the results of a CCA.

In CCA, the correlation between $p_j$ and $q_j$ is called a canonical correlation, and the composites $p_j$ and $q_j$ are sometimes referred to as canonical variates. (The notation $p_{ij}$ refers to the score of the ith individual on the jth composite predictor, while $p_j$ is a general reference to the jth composite predictor.) In this dissertation, the term "canonical factor" will sometimes be used as a synonym for canonical variate. The term "factor variable" or "factor score" will be used to refer to a composite score such as $p_j$ or $q_j$ derived from either an FA or a CCA.

# CHAPTER II

## REVIEW OF THE LITERATURE

Hotelling, who developed CCA, also made two specific proposals for its application which are of relevance to psychological research. The first (Hotelling, 1935) was in the area of predicting academic achievement. He suggested that instead of using a single criterion of academic success, such as cumulative grade point average, several of the more common criteria be used, allowing the analysis itself to weight and combine these into a "most predictable criterion." This proposal seems rarely to have been followed: almost all researchers continue to rely on multiple regression and the single criterion. Often the single criterion is not truly unitary. Grade point average, for example, is a weighted sum (all weights being equal) of grades in a number of courses; CCA, by assigning different weights to grades in different course areas, could enhance the predictability of academic success.

One investigator who did follow Hotelling's suggestion was Jones (1964), who used CCA to predict talented behavior in students. Test data gathered on 450 students in the seventh grade was used to predict their performance as high school seniors. Talented behavior was defined in terms of grade average, aptitude and achievement test scores, teacher and peer nominations for various kinds of talents, and awards received for talented achievements. These raw criterion measures were factor-analyzed to remove redundancy, yielding a set of 21 uncorrelated factor variables. The raw predictors, drawn from test scores and classroom performance, were also factor-analyzed to yield seven uncorrelated factor variables. (While the factored predictors did not correlate with one another and the factored criteria did not correlate with one another, the factored predictors did correlate with the factored criteria. That is, the factor variables were orthogonal within sets but not across sets.) The seven factored predictors and the 21 factored criteria were then subjected to a CCA, yielding five significant canonical correlations for the first five matched pairs of canonical variates. These correlations ranged from .78 to .29.

Jones did not make use of canonical loadings in interpreting the meaning of the canonical factors, but he was explicitly aware of the fact that the canonical weights are partial regression coefficients which can not <u>in general</u> be used for interpretation. He was able to use these weights for interpretation in his study, however, because of having factor-analyzed the predictor and criterion sets before using them in the CCA. Jones' discussion of the rationale was as follows:

> Ward (1962) has shown that beta weights cannot be interpreted in the multiple linear regression case because the weights represent only one of several possible solutions to the regression equation and because the predictor variables typically have some linear dependencies. Ward states, however, that the beta weights can be interpreted in the special case of orthogonal variables . . . . Because the variables in this study are all factor variables . . . they meet Ward's criterion of orthogonality (1964, p. 37).

It will be proved in Appendix II that the reason canonical weights can be interpreted in this special case is that they are in this instance identical with the canonical loadings within sets. Thus Jones has shown one way in which a meaningful CCA can be carried out. In the general case, however, weights and loadings are not identical. If for some reason an investigator cannot or does not wish to use factored variables as input to a CCA, he can still interpret the results by recourse to the canonical factor loadings, as will be shown later.

Hotelling's second proposal for the application of CCA was as follows:

> A use sometimes made for factor analysis in the past is in testing for the relations between two sets of variates. One study of the relations of character with mentality, for example, used seven mental tests and seven estimates of character traits by acquaintances. A standard factor analysis technique was applied to the mental measurements, and independently to the character scores, so as to get seven mental factors and seven character factors. A name was applied to each of the fourteen factors thus found, and a plausible matching of character factor with mental factor was arranged so as to get seven pairs. The correlations within the pairs thus calculated were then computed, and judged in each case insignificant.
>
> This kind of use of factor analysis should clearly be superseded by an examination of canonical correlations between the two sets of variates. The largest canonical correlation will be larger than all the correlations between the two sets actually found, and has a better chance of being found significant and thus demonstrating the existence of a relation between the two sets if one actually exists. The other canonical correlations, in decreasing order of size, and the corresponding canonical variates, will help elucidate the nature of any relations that may exist between character and mentality. The separate factor analyses of the two sets were really quite useless for this purpose (1957, p. 74).

Ten years before Hotelling made the statement just quoted, Burt (1947) published a study in which he applied both CCA and separate factor analyses to the same data. The data he used were artificial, since he wished to show under what conditions the two types of analysis would lead to the same result. One set of (fictitious) variables was supposed to consist of students' scores on four tests, while the other set (also fictitious) consisted of estimates by four different teachers of the academic ability of each of the students. Thus there were eight variables in all, four in each set.

Burt factor-analyzed each set separately to get four "test" factors and four "teacher" factors. He then wished to use each test factor to predict one of the teacher factors. Rather than pairing off the factors on the basis of plausible resemblances in the way that Hotelling criticized, Burt resorted to still another FA, lumping all eight variables together and analyzing them as if they constituted one set. This overall analysis yielded four general factors pretaining both to test and teachers' estimates; in fact, the loadings on these four general factors were <u>identical</u> to the loadings obtained in the separate analyses. Burt had arranged the data so that this would be the case. The overall analysis enabled Burt to match factors from the separate analyses on the basis of knowledge concerning the existence of general factors underlying both sets.

Burt also carried out a CCA of the two sets. Since CCA as Burt was using it produced weights but not loadings, he had to find some basis for comparing the canonical and factor analyses. He solved this problem by calculating factor <u>weights</u> (like those in equation 7a) for comparison with the canonical weights, and he found the respective sets of weights to be proportional. He concluded that the same factors were extracted in the CCA and in the separate FAs (and in the overall FA, for that matter).

Acknowledging that his artificial data presented something of a special case, Burt argued that his approach would be useful with real data. To the extent that the results of an overall FA and separate FAs agree, so also would the results of a CCA and separate FAs agree. He suggested that an investigator employ both CCA and an overall FA in attempting to understand a given problem. Presumably the CCA would provide maximum predictability for the pairs of canonical variates, while the FA would provide a key to the interpretation of the CCA. It will be shown in this dissertation that through the interpretation of canonical loadings CCA can be used <u>both</u> to elucidate factor structure and to describe relationships between sets of variables. A separate FA is not required.

In this same article Burt made a statement which seems to have been misunderstood by later investigators. He said, "Factors we may consider to be identifiable in terms of their weights" (1947, p. 104). Burt was not here referring to <u>interpreting</u> factors; he was noting that if the same variables receive the same relative weights by two different factoring procedures, then the same factor must be involved in both cases. Factors may be identifiable in terms of their weights, but it is a long established practice (Fruchter, 1947; Harman, 1960) to interpret them in terms of their loadings.

Quite recently, Das (1965) pursu[illegible]rt's suggestion that both an overall FA and a CCA be carried out on the same set of empirical (rather than artificial) data. Before making the comparative analyses, Das argued than only one particular <u>kind</u> of FA, namely principal components analysis (PCA), was really comparable with CCA. According to Das, the fundamental PCA equation is as follows:

(12) PCA: $p_{i1} = v_{11}x_{i1} + v_{21}x_{i2} + \cdots + v_{n1}x_{in},$

while the fundamental CCA equations are:

(13) CCA: $p_{i1} = v_{11}x_{i1} + v_{21}x_{i2} + \cdots + v_{n1}x_{in};$

$$q_{i1} = w_{11}y_{i1} + w_{21}y_{i2} + \cdots + w_{n1}y_{in} \, .$$

In both of these, the factor variables are expressed as functions of the original variables. Both the PCA and CCA solutions yield regression weights which enable one to compute factor scores for each individual in the sample.

Other types of FA assume the following fundamental equation (assuming that there are as many factors as original variables):

(14) FA: $x_{i1} = f_{11}p_{i1} + f_{12}p_{i2} + \cdots + f_{1n}p_{in}.$

Here the FA solution directly yields the loadings rather than the weights. Das makes much of the fact that in (12) and (13) the factor variables (p's and q's) are dependent whereas in (14) they are independent. While this may be true in a formal sense, it does not seem to have much practical significance. The weights in (12), and separately in (13) can be used to calculate loadings, and the loadings in (14) can be used to calculate weights. Burt, in his 1947 article cited by Das, computed both weights and loadings from what was essentially a model such as (14). As will be shown, <u>both</u> are useful in <u>both</u> CCA and FA. The weights enable one to compute factor scores for each person on each factor, and the loadings enable one to make possibly fruitful hypotheses about what the factors mean.

There is another way in which PCA and CCA are alike that does seem to have some practical significance. Both require that <u>ones</u> be placed in the diagonal of the correlation matrix. Other types of factor analysis permit "communality" estimates rather than ones in the diagonal. The issue of communalities will not be treated here (see Harman, 1960), except to note that the choice of communality estimates is to some extent a subjective matter which can affect the values of the weights and loadings. For this reason, PCA is the only type of FA considered in this dissertation.

Das did not compare CCA and PCA in his empirical demonstration. Rather he used principal axis FA, which is computationally the same as PCA except that communality estimates are used, stating that he wished to see how the differences in model would affect the outcome. His data consisted of the scores of 223 Indian college students on five "experimental non-verbal reasoning tests" and 12 "reference tests for reasoning." The principal axis FA was done on all 17 tests, while for CCA the tests were divided into sets of five and 12 as indicated.

For comparative purposes, Das reported the loadings for the first five principal factors and the weights for all five canonical variates. He acknowledged that weights and loadings have different statistical meanings, and he showed that variables with higher loadings on a given factor in the FA did not always have higher weights on the corresponding factor of the CCA. He did note a rough correspondence in the relative sizes of weights and loadings for the first factors only, which led him to conclude that "where there is a general factor running through the combined set of measurements, it will appear as the first canonical variate when the measurements are divided into two sets" (p. 64). He added that this conclusion is consistent with Burt's expectations.

Das' final conclusions were:

> It may be inferred from these illustrative results that where generation of ideas regarding the nature of a domain or set of measures is required, factor analysis is likely to retain its pre-dominant position in the analysis of psychological data. ... For purposes of prediction and the testing of statistical hypotheses concerning the relations between two sets of variables, canonical analysis appears to be the more appropriate statistical method (p. 66).

In other words, Das seems to imply, with FA one can interpret but not make statistical tests, while in CCA one can make statistical tests but not interpret. It is the main purpose of this dissertation to show, however, that with CCA, through the use of canonical loadings, both interpretation and the making of statistical tests are possible.

There are relatively few other studies reported in the literature in which CCA has been used with real data. Cooley and Lohnes (1962) present the results of a CCA in which relationships are sought between a number of measures of early home environment and several measures of present orientation toward people. They present only weights in their results, and appear to interpret them as if they were loadings. King, Bowman, and Moreland (1961), seeking factors common to biochemical levels and intelligence, take a similar interpretative approach, citing Burt's (1947) statement that factors are identifiable by their weights as justification for interpreting in terms of weights. Wittenborn (1963) followed up a CCA with a series of MRAs, taking one criterion variable at a time, in an attempt to clarify the nature of the relationships between the predictor and criterion sets. Such a two-step procedure is unnecessary, since the use of canonical loadings will do a clearer job of showing how the two sets are related.

The relative neglect of CCA, as well as the failure to exploit it fully in those cases where it has been used, may be due in part to the extreme difficulty, for non-mathematicians, of Hotelling's original articles. Almost a dozen years after these appeared, Thomson (1947) attempted a "popular" presentation. Though very much easier to follow, this article still did not suggest that any more information could be gotten out of an analysis than one or more canonical correlation coefficients. Thomson indicated briefly, without comment, how CCA is an extension of MRA to the case of several criterion variables. He calculated one kind of canonical loadings, but only as a computational step in getting the weights; he did not identify the loadings as such, nor indicate that they might be helpful in elucidating factor structure. Thomson's linking of CCA to MRA, however, was an important step forward in furthering the understanding of CCA, and his demonstration will be recapitulated, with more commentary than he provided, later in this dissertation.

In a recent book on multivariate statistics, Anderson (1958) devotes a chapter to CCA. Although more straightforward than Hotelling's original articles, Anderson's presentation is still heavy going for non-mathematicians. His development is based on deviation scores rather than standard scores, which introduces some extra complexity into the discussion. Anderson provides one worked example, involving two predictor and two criterion variables. He interprets the results of the weights, and makes no mention of loadings.

Two recent books dealing with the use of electronic computers in the behavioral sciences contain chapters on CCA. Koons (1962) took the small correlation matrix that Thomson had analyzed earlier and showed in more detail how to analyze it. Like Thomson, he used canonical loadings only as a step in the calculation of the weights. Toward the end of the article he calls for the joint application of CCA and FA to the same data as a way of increasing the understanding of both methods. Koons' presentation would be a useful "cookbook" for potential users of CCA, except that there are some algebraic and computational errors that could be confusing.

The other computer-oriented chapter on CCA is by Cooley and

Lohnes (1962). They show algebraically how canonical weights and correlations are obtained. They provide a worked example, but as already noted, the results are interpreted in terms of weights. They also provide a computer program for CCA written in IBM FORTRAN. Although a knowledge of matrix algebra is assumed on the part of the reader, this chapter by Cooley and Lohnes is probably the most easily understandable introduction to the subject now available. It will not, however, enable the researcher fully to exploit the possibilities inherent in CCA.

Meredith (1964) described a method of applying CCA to the "true score" component of observed scores, thus correcting the canonical correlation coefficient for attentuation. He coincidentally took up the "more or less unrelated" problem of interpretation:

> The usual method of analysis is to compute the canonical correlations and the associated matrices of regression weights (transformation matrices) for determining the canonical variates from the original measures. If the variables within each set are moderately intercorrelated the possibility of interpreting the canonical variates by inspection of the appropriate regression weights is practically nil. However, the correlations between the canonical variates and the original measures can be very enlightening (p. 55).

These "enlightening" correlations are what are here called loadings. Meredith did not refer to them as such. He showed how to calculate both loadings within and loadings between sets, but did not point out that there is a relation between them. It is proved in Appendix II that for each factor, the loadings between sets equal the loadings within sets multiplied by the associated canonical correlation coefficient. It is also proved that the sum of squares of loadings between sets for a given variable equals the squared multiple correlation between that variable and all the variables of the other set.

Meredith provided an example consisting of a CCA of the Wechsler Intelligence Scale for Children, with the verbal subtests making up one battery and the performance subtests making up the other. Subjects were 100 boys and 100 girls. He did the analysis both in terms of the observed scores and the hypothetical true scores, and perhaps because his attention was focused on this aspect of the analysis, he did not interpret the canonical loadings, although they were given in the results section. Inspection of his first factor loadings suggests that there is a general verbal factor which strongly saturates all the verbal subtests, particularly Digit Span and Comprehension, while it saturates primarily only three of the performance subtests--Picture Arrangement, Picture Completion, and Coding. The first canonical correlation between batteries is .68, which becomes a .97 when corrected for attenuation.

In one section of a mathematically-oriented monograph relating CCA to FA, discriminant function analysis, and scaling theory, McKeon (1962) indicates that CCA may be used for matching factors from separate test batteries. He proves that the matrix of what are here called loadings within sets can be obtained from a matrix of suitably scaled canonical weights, though he uses the terms "factor structure" and "factor matrix" rather than "loadings." Using data published by another author, McKeon offers as an example the CCA of a battery of occupational preference tests and a battery of personality measures. The matrix of canonical weights, the "factor matrix," and the canonical correlations are given without comment or discussion.

McKeon has in essence solved the interpretation problem in CCA, though perhaps more by implication than by explication. A later paper by Mukherjee (1966), which cites McKeon in the introduction, presents a series of CCAs of learning data in which only the weights are given--"factor matrices" or "loadings" are not mentioned. Mukherjee uses the highest weighted variable as the basis for interpreting a factor, and discounts the variable with the lowest weight, though it will later be shown that a variable may have a low weight and still have a high loading.

Mukherjee also makes the statement that the criterion variable with the highest weight can be predicted as well as the entire composite criterion of which it is a part. That is, he states in effect that if $y_1$ has the highest weight in $q_1$, for example, then $r_{p_1 y_1} = r_{p_1 q_1}$ = the first canonical correlation. Actually, $r_{p_1 y_1}$ is what is here called a loading between sets, and it would equal the first canonical correlation only if the weights for $y_2$, $y_3$, and so on, were all identically zero.

McKeon, in another section of his monograph, takes up the question of finding an overall index of association between batteries. Each canonical correlation is a measure of association between matched pairs of factors extracted from the two batteries, and McKeon notes that the root mean square canonical correlation has been proposed as an overall measure of association between the two batteries each taken as a whole. He offers an alternative which is a function only of the first squared canonical correlation coefficient. In this dissertation an index (called $H^2$ for Hotelling) will be proposed which takes into account not only the size and significance of all the canonical correlations, but also the factor structure of the two batteries.

To summarize the review of the literature, there seems to have been no investigator, whether of an applied or a theoretical bent, who has made an attempt to clarify thoroughly the relationships between MRA, FA, and CCA, or who has fully exploited the interpretative possibilities inherent in the examination of the canonical loadings within and between sets. Weights and loadings have been confused by many authors, while others have used loadings without recognizing that they have some of the same properties as ordinary factor loadings, plus some unique properties that can also be helpful in interpretation. There are also differences in notation and terminology in the literature which, though not presented in detail here, are a possible source of perplexity to the investigator who would like to find out what CCA is all about.

In what follows, an artificial problem and a problem involving real data will be analyzed extensively, with sufficient commentary given at every step to make clear what that step involves. References to differences in notation and terminology will be made at appropriate places in the various analyses. It will be shown how MRA and CCA are related, and how CCA, through the use of loadings, can elucidate factor structure. An overall index of relationship between batteries which takes account of factor structure will be proposed. Algebraic proofs, and a complete summary in matrix notation of all important definitions and relationships, will be presented separately for MRA and CCA in Appendices I and II respectively.

## CHAPTER III

## ILLUSTRATIVE ANALYSES OF ARTIFICIAL DATA

The artificial data for this illustrative problem is shown in terms of "true" factor scores in Table 1. These factors have no

TABLE 1.--"True" factor scores for the illustrative problem

| Subject | Variables | | | |
|---|---|---|---|---|
| | $p_1$ | $p_2$ | $q_1$ | $q_2$ |
| 1 | -1.11 | .62 | -1.12 | - .58 |
| 2 | - .27 | .71 | - .26 | 1.76 |
| 3 | .29 | .40 | .31 | - .55 |
| 4 | - .46 | -1.72 | - .46 | - .48 |
| 5 | 1.55 | - .02 | 1.54 | - .15 |
| Mean | 0.00 | 0.00 | 0.00 | 0.00 |
| Variance | 1.00 | 1.00 | 1.00 | 1.00 |

particular meaning--this problem is intended to show in detail how to do the complete MRA and CCA. Interpretation will be taken up in the next chapter. All calculations for this problem were carried out to six decimal places; these were rounded to two places for brevity of presentation. Note that the factor measures in Table 1 are in standard score form.

The ideal factor scores in Table 1 are transformed into "observed" scores using the following relations:

(15a) $x_{i1} = (.81)p_{i1} + (-.60)p_{i2}$, $y_{i1} = (.87)q_{i1} + (.49)q_{i2}$;

(15b) $x_{i2} = (.28)p_{i1} + (-.96)p_{i2}$, $y_{i2} = (-.96)q_{i1} + (.29)q_{i2}$.

(The i subscript on x, y, p, and q will hereafter be dropped, but implicitly there.) The coefficients of the p's and q's are factor loadings. The "observed scores" thus obtained from (15a) and (15b) for each of the five subjects are given in Table 2. It is from these observed scores that the MRA and CCA proceed.

TABLE 2.--"Observed" scores for the illustrative problem

| Subject | Variables | | | |
|---|---|---|---|---|
| | $x_1$ | $x_2$ | $y_1$ | $y_2$ |
| 1 | -1.26 | - .91 | -1.26 | .91 |
| 2 | - .63 | - .76 | .63 | .76 |
| 3 | 0.00 | - .30 | 0.00 | - .45 |
| 4 | .63 | 1.52 | - .63 | .30 |
| 5 | 1.26 | .45 | 1.26 | -1.52 |
| Mean | 0.00 | 0.00 | 0.00 | 0.00 |
| Variance | 1.00 | 1.00 | 1.00 | 1.00 |

## Vector and Matrix Notation

The four columns of Table 2 are standard score vectors, symbolized by $x_1$, $x_2$, $y_1$, and $y_2$. For example,

$$(16) \qquad \underline{x}_2 = \begin{bmatrix} -.91 \\ -.76 \\ -.30 \\ 1.52 \\ .45 \end{bmatrix} .$$

The sum of squares for any variable is obtained by premultiplying the standard score vector by its transpose: sum of squares = $\underline{x}_j' \underline{x}_j$. The sum of crossproducts for any two variables is of the form $\underline{x}_i' \underline{x}_j$, $i \neq j$. For n degrees of freedom, the variance of any variable, say $\underline{y}_2$, is $(1/n)\underline{y}_2' \underline{y}_2 = 1.00$, and the correlation between any two variables, say $x_1$ and $y_2$, is $(1/n)\ \underline{x}_1' \ \underline{y}_2$, or alternatively $(1/n)\ \underline{y}_2' \underline{x}_1$.

The two columns of x scores make up the predictor matrix $\underline{X}$:

$$(17) \qquad \underline{X} = \begin{bmatrix} \underline{x}_1 & \underline{x}_2 \end{bmatrix} = \begin{bmatrix} -1.26 & -.91 \\ -.63 & -.76 \\ 0.00 & -.30 \\ .63 & 1.52 \\ 1.26 & .45 \end{bmatrix} .$$

The matrix product $(1/n)\ \underline{X}'\underline{X}$ gives the matrix of intercorrelations for the x's and is symbolized by $\underline{R}_{11}$ :

$$(18) \quad \underline{R}_{11} = (1/n)\ \underline{X}'\underline{X} = \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} \\ r_{x_2x_1} & r_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 1.00 & .79 \\ .79 & 1.00 \end{bmatrix} .$$

$\underline{Y}$ is the two-column matrix of y scores from Table 2, and the product $(1/n)\ \underline{Y}'\underline{Y}$ gives the intercorrelations of the y's ($\underline{R}_{22}$). The product (1/n) $\underline{X}'\underline{Y}$ gives the correlations of the x's with the y's ($\underline{R}_{12}$), and (1/n) $\underline{Y}'\underline{X}$ the correlations of the y's with the x's ($\underline{R}_{21}$). $\underline{R}_{12}$ and $\underline{R}_{21}$ contain the same entries but with rows and columns interchanged, so that one is said to be the "transpose" of the other.

All four columns of Table 2 make up the data matrix $\underline{D}$:

$$(19) \qquad \underline{D} = \begin{bmatrix} \underline{X} & \underline{Y} \end{bmatrix} = \begin{bmatrix} \underline{x}_1 & \underline{x}_2 & \underline{y}_1 & \underline{y}_2 \end{bmatrix} .$$

The matrix of intercorrelations of all the variables in Table 2 may be obtained directly using $\underline{D}$:

$$(20) \qquad \underline{R} = (1/n)\,\underline{D}'\underline{D} = \begin{bmatrix} (1/n)\,\underline{X}'\underline{X} & (1/n)\,\underline{X}'\underline{Y} \\ (1/n)\,\underline{Y}'\underline{X} & (1/n)\,\underline{Y}'\underline{Y} \end{bmatrix} = \begin{bmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{21} & \underline{R}_{22} \end{bmatrix} .$$

$\underline{R}$ is a supermatrix (a matrix whose elements are matrices). Numerically,

$$(21) \qquad \underline{R} = \begin{bmatrix} 1.00 & .79 & .60 & -.84 \\ .79 & 1.00 & .07 & -.37 \\ & & & \\ .60 & .07 & 1.00 & -.70 \\ -.84 & -.37 & -.70 & 1.00 \end{bmatrix} .$$

## Multiple Regression With a Single Criterion

The usual MRA approach to predicting the y's from the x's would take each y separately. To predict $y_1$, for example, the following correlation matrix would be used:

$$(22)\qquad \begin{bmatrix} \underline{R}_{11} & \underline{r}_{12} \\ \underline{r}_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1.00 & .79 & .60 \\ .79 & 1.00 & .07 \\ .60 & .07 & 1.00 \end{bmatrix}.$$

The correlations involving $y_2$ are left out, and the matrices $\underline{R}_{12}$ and $\underline{R}_{21}$ are replaced by the vectors $\underline{r}_{12}$ and $\underline{r}_{21}$, while the matrix $\underline{R}_{22}$ is replaced by the single number $r_{22}$ which represents the correlation of $y_1$ with itself. From (22) the vector of standard partial regression weights, $\underline{a}$, for predicting $y_1$ from the x's, is found by:

$$(23)\qquad \text{MRA:}\ \underline{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \underline{R}_{11}^{-1}\,\underline{r}_{12} = \begin{bmatrix} 2.67 & -2.11 \\ -2.11 & 2.67 \end{bmatrix} \begin{bmatrix} .60 \\ .07 \end{bmatrix} = \begin{bmatrix} 1.45 \\ -1.08 \end{bmatrix},$$

where $\underline{R}_{11}^{-1}$ is the inverse of $\underline{R}_{11}$. (The inverse of a matrix is analogous to the reciprocal of an ordinary number.) The prediction equation for $y_1$ is therefore:

$$(24)\qquad \text{MRA:}\ P_1 = (1.45)x_1 + (-1.08)x_2, \qquad r_{p_1 y_1} = \text{maximum} = \text{mult R}.$$

$P_1$ is not in standard score form. Mult $R^2$ is found by:

$$(25)\qquad \text{MRA:}\ \text{mult } R^2 = \underline{r}_{21}\,\underline{R}_{11}^{-1}\,\underline{r}_{12} = .79.$$

## Simultaneous Regression Analyses Involving More Than One Criterion Variable

In the same manner as above, separate equations could be developed for predicting $y_2$. However, it is possible to do both MRAs at once using the full correlation matrix shown in (20) and (21). The matrix of regression weights for both $y_1$ and $y_2$ is given by:

$$(26) \qquad \text{MRA:} \quad \underline{A} = \underline{R}_{11}^{-1}\,\underline{R}_{12} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1.45 & -1.45 \\ -1.08 & .78 \end{bmatrix},$$

which yields a <u>pair</u> of prediction equations:

$$(27a) \qquad \text{MRA:} \quad P_1 = (\ 1.45)x_1 + (-1.08)x_2,$$

$$(27b) \qquad \text{MRA:} \quad P_2 = (-1.45)x_1 + (\ \ .78)x_2.$$

The two squared mult R's between $P_1$ and $y_1$ and between $P_2$ and $y_2$ respectively are found on the diagonal of the matrix $\underline{R}_{2.1}$, where

$$(28) \quad \text{MRA:} \quad \underline{R}_{2.1} = \underline{R}_{21}\,\underline{R}_{11}^{-1}\,\underline{R}_{12} = \begin{bmatrix} .79 & -.82 \\ -.82 & .93 \end{bmatrix}.$$

That is, mult $R_1^2 = .79$ and mult $R_2^2 = .93$. The off-diagonal elements of $\underline{R}_{2.1}$ have no simple meaning, but are useful in subsequent calculations.

### Standardizing the Composite Predictors

$P_1$ and $P_2$ are not in standard form; they have zero means and variances equal to their respective mult $R^2$'s. They may be standardized by dividing the $P_1$ weights through by mult $R_1$ and the $P_2$

weights through by mult $R_2$. This will give a new matrix of weights, $\underline{V}$:

$$\text{(29a)} \quad \text{MRA:} \quad \underline{V} = \underline{A}\ \underline{Z}, \text{ where } \underline{Z} = \begin{bmatrix} (\text{mult } R_1)^{-1} & 0 \\ 0 & (\text{mult } R_2)^{-1} \end{bmatrix}.$$

$$\text{(29b)} \quad \text{MRA:} \quad V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} 1.63 & -1.51 \\ -1.21 & .81 \end{bmatrix}.$$

The prediction equations become:

$$\text{(30a)} \quad \text{MRA:} \quad p_1 = v_{11}x_1 + v_{21}x_2 = (1.63)x_1 + (-1.21)x_2, \text{ and}$$

$$\text{(30b)} \quad \text{MRA:} \quad p_2 = v_{12}x_1 + v_{22}x_2 = (-1.51)x_1 + (\ .81)x_2.$$

The multiple correlations are the same whether the a weights or v weights are used.

Equations (30a) and 30b) yield a pair of composite predictor scores for <u>each</u> of the five subjects. These five pairs may be written as $\underline{P}$:

$$\text{(31)} \quad \text{MRA:} \quad \underline{P} = \underline{X}\ \underline{V} = \begin{bmatrix} \underline{p}_1 & \underline{p}_2 \end{bmatrix} = \begin{bmatrix} -.96 & 1.17 \\ -.11 & .34 \\ .37 & -.24 \\ -.80 & .27 \\ 1.51 & -1.54 \end{bmatrix}.$$

This $\underline{P}$ matrix is shown below in tabular form in the right half of Table 3, for comparison with the "true" predictor scores from Table 1.

TABLE 3.--"True" factor predictor scores compared with composite predictors from multiple regression

| Subject | "True" Scores $p_1$ | $p_2$ | Regression Scores $p_1$ | $p_2$ |
|---|---|---|---|---|
| 1 | -1.11 | .62 | - .96 | 1.17 |
| 2 | - .27 | .71 | - .11 | .34 |
| 3 | .29 | .40 | .37 | - .24 |
| 4 | - .46 | -1.72 | - .80 | .27 |
| 5 | 1.55 | - .02 | 1.51 | -1.54 |

While there is some resemblance in Table 3 between the "true" $p_1$ scores and the MRA $p_1$ scores, there is little resemblance between the two sets of $p_2$ scores. In general, MRA will not produce weighted combinations which approximate underlying factor structure (Cattell, 1952, pp. 18-20). Further, the "true" $p_1$ and $p_2$ are uncorrelated, while the $p_1$ and $p_2$ from MRA are, and in general would be, correlated.

## Redundancy in Multiple Regression

The primary interest in MRA is focused on the mult R's, but it can be instructive to look at all the intercorrelations among $p_1$, $p_2$, $y_1$, and $y_2$. These are shown in Table 4. It is shown in Appendix I that the $\underline{R}_{2.1}$ matrix from equation (28) is useful in calculating the correlations in Table 4.

TABLE 4.--Upper triangular correlation matrix of predictors weighted by multiple regression and simple unweighted criteria (multiple correlations are underlined)

| | $p_1$ | $p_2$ | $y_1$ | $y_2$ |
|---|---|---|---|---|
| $p_1$ | 1.00 | -.95 | <u>.89</u> | -.92 |
| $p_2$ | | 1.00 | -.85 | <u>.96</u> |
| $y_1$ | | | 1.00 | -.70 |
| $y_2$ | | | | 1.00 |

It is clear from Table 4 that there is a good deal of redundancy within this system of variables. For example, $p_1$ correlates -.95 with $p_2$, which means that what one predicts, the other will predict almost as well. Further, although $p_1$ has been weighted so that it is the maximum predictor of $y_1$, it actually correlates more highly with $y_2$ than with $y_1$. There are no clear-cut relationships in Table 4. Everything is correlated highly with everything else, so that it is hard to give an unequivocal interpretation to <u>any</u> relationship. By contrast, it will be seen that in CCA relationships are more clear-cut. Prediction using MRA may be said to be "blindly empirical" in that accuracy of prediction is maximized with little understanding gained of relationships among predictors and criteria.

## Canonical Correlation as a Generalization of Multiple Regression

Whereas MRA weights and combines only the predictor variables, CCA finds weights for both the predictors and the criteria. It will now be shown, following and elaborating on Thomson (1947), how in an algebraic sense CCA is a generalization of MRA to the case of multiple criteria.

Let $y_1$ and $y_2$ from the present example be weighted and combined into a single criterion, $Q_j$:

$$\text{(32)} \quad \text{CCA, MRA:} \quad Q_j = b_{1j}y_1 + b_{2j}y_2,$$

where $Q_j$ is not necessarily in standard form and the b's are not yet specified in value. The reason for the j subscript on Q will become apparent. $Q_j$ is now a single criterion to be predicted from the x's. The matrix equation for the standard partial regression weights to be applied to the x's is:

$$\text{(33)} \quad \text{CCA, MRA:} \quad \underline{a}_j = \underline{R}_{11}^{-1}\underline{r}_{12},$$

which is identical in form to equation (23). The vector $\underline{r}_{12}$ has a different meaning in the two equations, however. In (23) it represents the correlations of $y_1$ with the x's, while in (33) it stands for the correlations of $Q_j$ with the x's. The squared multiple correlation between $Q_j$ and the weighted sum of x's is:

$$\text{(34)} \quad \text{CCA, MRA:} \quad \text{mult } R_j^2 = \underline{r}_{21}\,\underline{R}_{11}^{-1}\,\underline{r}_{12},$$

which has the same form as equation (25), but with $Q_j$ rather than $y_1$ as the single criterion.

In (33) and (34), $\underline{r}_{12} = \begin{bmatrix} r_{x_1Q_j} \\ r_{x_2Q_j} \end{bmatrix}$. It is a function of the as yet unspecified b's and the correlations of $x_1$ and $x_2$ with $y_1$ and $y_2$:

$$\text{(35a)} \quad \text{CCA, MRA:} \quad r_{x_1Q_j} = \frac{b_{1j}r_{x_1y_1} + b_{2j}r_{x_1y_2}}{\sqrt{b_{1j}^2 + 2b_{1j}b_{2j}r_{y_1y_2} + b_{2j}^2}}$$

$$\text{(35b)} \quad \text{CCA, MRA:} \quad r_{x_2Q_j} = \frac{b_{1j}r_{x_2y_1} + b_{2j}r_{x_2y_2}}{\sqrt{b_{1j}^2 + 2b_{1j}b_{2j}r_{y_1y_2} + b_{2j}^2}}$$

Both of these complicated expressions can be expressed compactly in one matrix equation:

$$\text{(35c)} \quad \text{CCA, MRA:} \quad \underline{r}_{12} = \frac{\underline{R}_{12}\ \underline{b}_j}{\sqrt{\underline{b}_j'\ \underline{R}_{22}\ \underline{b}_j}}, \text{ where } \underline{b}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \end{bmatrix}, \text{ and}$$

where $\underline{R}_{12}$ and $\underline{R}_{21}$ are as given in equations (20) and (21). The expression in the denominator of (35c), which is the matrix equivalent of the denominator in (35a) and (35b), is a single number, or scalar, when multiplied out.

The transpose of $\underline{r}_{12}$, which is $\underline{r}_{21}$, is expressed as:

$$\text{(35d)} \quad \text{CCA, MRA: } \underline{r}_{21} = \frac{\underline{b}'_j \, \underline{R}_{21}}{\sqrt{\underline{b}'_j \, \underline{R}_{22} \, \underline{b}_j}} \quad .$$

Sbustituting the expression for $\underline{r}_{12}$ from (35c) into (33) gives an equation for the vector of a weights in terms of the b's:

$$\text{(36)} \quad \text{CCA, MRA: } \underline{a}_j = \underline{R}_{11}^{-1} \, \underline{r}_{12} = \frac{\underline{R}_{11}^{-1} \, \underline{R}_{12} \, \underline{b}_j}{\sqrt{\underline{b}'_j \, \underline{R}_{22} \, \underline{b}_j}} \quad .$$

Substituting both (35c) and (35d) into (34) gives an expression for the squared multiple correlation:

$$\text{(37)} \quad \text{CCA, MRA: } \quad \text{mult } R_j^2 = \underline{r}_{21} \, \underline{R}_{11}^{-1} \, \underline{r}_{12} = \frac{\underline{b}'_j \, \underline{R}_{21} \, \underline{R}_{11}^{-1} \, \underline{R}_{12} \, \underline{b}_j}{\underline{b}'_j \, \underline{R}_{22} \, \underline{b}_j} \quad .$$

Equations (36) and (37) hold for all real nonzero values of $b_{1j}$ and $b_{2j}$. It may be asked, however, if one particular pair of values for the b's will result in the highest possible multiple correlation. The answer is "yes," and these values may be found by differentiating (37) with respect to $\underline{b}_j$ and setting the result equal to zero. When simplified, one form of the resulting matrix equation is:

$$\text{(38)} \quad \text{CCA, MRA: } \quad (\underline{R}_{22}^{-1} \, \underline{R}_{21} \, \underline{R}_{11}^{-1} \, \underline{R}_{12} - \lambda_j \, \underline{I})\underline{b}_j = \underline{0},$$

where $\lambda_j$ = mult $R_j^2$ as given in equation (37), and $\underline{I}$ is an identity matrix. (An identity matrix has ones in the diagonal and zeros elsewhere.) $\lambda_j$ is called a latent root, or an eigen value, of the matrix formed by the product $\underline{R}_{22}^{-1}\underline{R}_{21}\underline{R}_{11}^{-1}\underline{R}_{12}$, and $\underline{b}_j$ is called a latent or eigen vector. Equation (38) is called a characteristic equation. It is sometimes written as:

$$(39)\quad \text{CCA, MRA:}\quad (\underline{R}_{21}\underline{R}_{11}^{-1}\underline{R}_{12} - \lambda_j\underline{R}_{22})\underline{b}_j = \underline{0}.$$

Equation (39) can be obtained from (38) by premultiplying through by $\underline{R}_{22}$. In either equation, the product $\underline{R}_{21}\underline{R}_{11}^{-1}\underline{R}_{12}$ is familiar; it is the matrix $\underline{R}_{2.1}$ from equation (28).

The characteristic equation (38) is solved by setting the determinant of the matrix within the parenthesis equal to zero and solving for $\lambda_j$:

$$(40)\quad \text{CCA, MRA:}\quad \left|\underline{R}_{22}^{-1}\underline{R}_{21}\underline{R}_{11}^{-1}\underline{R}_{12} - \lambda_j\underline{I}\right| = 0.$$

Equation (40) can now be applied to the illustrative problem that has already been analyzed by MRA. The numerical value of $\underline{R}_{21}\underline{R}_{11}^{-1}\underline{R}_{21}$ is known from (28), and premultiplying this by $\underline{R}_{22}^{-1}$ gives:

$$(41a)\quad \text{CCA, MRA:}\quad \begin{bmatrix} 1.94 & 1.35 \\ 1.35 & 1.94 \end{bmatrix}\begin{bmatrix} .79 & -.82 \\ -.82 & .93 \end{bmatrix} = \begin{bmatrix} .44 & -.33 \\ -.51 & .70 \end{bmatrix}.$$

Substituting these values into (40) gives:

$$(41b)\quad \text{CCA, MRA:}\quad \left|\begin{bmatrix} .44 & -.33 \\ -.51 & .70 \end{bmatrix} - \begin{bmatrix} \lambda_j & 0 \\ 0 & \lambda_j \end{bmatrix}\right| = 0 \quad .$$

Expanding the determinant and simplifying:

(41c) CCA, MRA: $(.44 - \lambda_j)\ (.70 - \lambda_j) - (-.33)(-.51) = 0,$

(41d) CCA, MRA: $\lambda_j^2 - 1.14\lambda_j + .14 = 0,$

(41e) CCA, MRA: $\lambda_1 = 1.00 = \text{canon } R_1^2$ ;

(41f) CCA, MRA: $\lambda_2 = .14 = \text{canon } R_2^2$ .

Corresponding to the largest value of $\lambda$ , $\lambda_1 = 1.00$, is the vector $\underline{b}_1$, which is found by substituting $\lambda_1$ back into equation (38):

(42a) CCA, MRA: $$\left[\begin{bmatrix} .44 & -.33 \\ -.51 & .70 \end{bmatrix} - \begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \end{bmatrix}\right]\begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \underline{0}$$

(42b) CCA, MRA: $$\begin{bmatrix} -.56 & -.33 \\ -.51 & -.30 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \underline{0}, \text{ or}$$

(42c) CCA, MRA: $(-.56)b_{11} + (-.33)b_{21} = 0,$

(42d) CCA, MRA: $(-.51)b_{11} + (-.30)b_{21} = 0.$

Either (42c) or (42d), when solved for $b_{21}$, gives:

(42d) CCA, MRA: $b_{21} = (-1.71)b_{11}.$

Thus, any pair of $b_{11}$ and $b_{21}$ values which stand in the proportion (1.00) : (-1.71) will satisfy the characteristic equation. Setting $b_{11}$ arbitrarily equal to 1.00,

$$\text{(42f)} \quad \text{CCA, MRA:} \quad \underline{b}_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} 1.00 \\ -1.71 \end{bmatrix} \text{, and}$$

$$\text{(42g)} \quad \text{CCA, MRA:} \quad Q_1 = b_{11}y_1 + b_{21}y_2 = y_1 - 1.71y_2.$$

The problem of finding b weights which give the highest squared multiple correlation is now solved. Substituting the values of $\underline{b}_1$ from (42f) into (37) will give the highest mult $R_1^2 = 1.00$. Of course, such a computational step is unnecessary, since the highest mult $R_1^2$ has already been found as the largest root of (38):

$$\text{(43)} \quad \text{highest mult } R_1^2 = \text{canon } R_1^2 = \lambda_1 \doteq 1.00.$$

The vector of regression weights, $\underline{a}_1$, for predicting $Q_1$ are found by substituting the $\underline{b}_1$ values into (36). The results of the analysis thus far, then, yield an unstandardized composite predictor, $P_1$, which is maximally correlated with an unstandardized composite criterion $Q_1$:

$$\text{(44)} \quad \text{CCA:} \quad P_1 = a_{11}x_1 + a_{21}x_2, \quad Q_1 = b_{11}y_1 + b_{21}y_2, \; r_{P_1Q_1} = \text{max}.$$

The use of the double prefix "CCA, MRA" will be dropped hereafter; it has been used long enough to make the point that MRA is a special case of CCA.

The problem of standardizing $P_1$ and $Q_1$ will be deferred until $\lambda_2$, $P_2$, and $Q_2$ have been discussed. If $\lambda_2 = .14$ is substituted into (38), a second solution for the b weights is found:

$$\text{(45)} \quad \text{CCA:} \quad \underline{b}_2 = \begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix} = \begin{bmatrix} 1.00 \\ .91 \end{bmatrix},$$

which leads to a second set of a weights and thus a second pair of composite scores:

$$\text{(46)} \quad P_2 = a_{12}x_1 + a_{22}x_2, \quad Q_2 = b_{12}y_1 + b_{22}y_2,$$

subject to the restriction that $r_{P_1P_2} = r_{Q_1Q_2} = r_{P_1Q_2} = 0$. That is, $P_2$ and $Q_2$ constitute a second pair of composites which is orthogonal to the first pair.

The results of the analysis thus far can be collected and expressed in matrix notation as follows:

$$\text{(47a)} \quad \text{CCA:} \quad \underline{A} = \begin{bmatrix} \underline{a}_1 & \underline{a}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ \underline{B} = \begin{bmatrix} \underline{b}_1 & \underline{b}_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

$$\text{(47b)} \quad \text{CCA:} \quad \text{unstandardized } \underline{P} = \underline{X}\ \underline{A}, \quad \text{unstandardized } \underline{Q} = \underline{Y}\ \underline{B}.$$

The variance-covariance matrix for the composite predictors is:

$$\text{(48a)} \quad \text{CCA:} \quad \text{unstandardized } (1/n)\ \underline{P}'\underline{P} - (1/n)\underline{A}'\underline{X}'\underline{X}\ \underline{A} = \underline{A}'\underline{R}_{11}\ \underline{A}$$

$$= \begin{bmatrix} \underline{a}_1' \, \underline{R}_{11} \, \underline{a}_1 & 0 \\ 0 & \underline{a}_2' \, \underline{R}_{11} \, \underline{a}_2 \end{bmatrix} = \underline{Z}_1 .$$

The variances are on the diagonal, and the covariances (off-diagonal elements) are zero, since $P_1$ is orthogonal to $P_2$. Similarly, for the Q's:

$$\text{(48b)} \quad \text{CCA:} \quad \text{unstandardized } (1/n)\, \underline{Q}'\underline{Q} = \begin{bmatrix} \underline{b}_1' \, \underline{R}_{22} \underline{b}_1 & 0 \\ 0 & \underline{b}_2' \, \underline{R}_{22} \underline{b}_2 \end{bmatrix} = \underline{Z}_2 .$$

Standardizing the Factor Scores

$P_1$, $P_2$, $Q_1$, and $Q_2$ can now be standardized by dividing through the weights of each by its own standard deviation. This is done in terms of matrices by:

$$\text{(48c)} \quad \text{CCA:} \quad \underline{V} = \underline{A}\, \underline{Z}_1^{-\frac{1}{2}} , \text{ and } \underline{W} = \underline{B}\, \underline{Z}_2^{-\frac{1}{2}}$$

and the standardized matrix equations, symbolically and numerically, are:

$$\text{(48d)} \quad \text{CCA:} \quad \underline{P} = \underline{X}\,\underline{V} = \begin{bmatrix} -1.26 & -.91 \\ -.63 & -.76 \\ 0.00 & -.30 \\ .63 & 1.52 \\ 1.26 & .45 \end{bmatrix} \begin{bmatrix} 1.57 & .46 \\ -.96 & -1.33 \end{bmatrix} = \begin{bmatrix} -1.11 & .62 \\ -.27 & .71 \\ .29 & .40 \\ -.46 & -1.72 \\ 1.55 & -.02 \end{bmatrix}$$

$$\text{(48e)} \quad \text{CCA:} \quad \underline{Q} = \underline{Y}\,\underline{W} = \begin{bmatrix} -1.26 & .91 \\ .63 & .76 \\ 0.00 & -.45 \\ -.63 & .30 \\ 1.26 & -1.52 \end{bmatrix} \begin{bmatrix} .40 & 1.33 \\ -.68 & 1.21 \end{bmatrix} = \begin{bmatrix} -1.12 & -.58 \\ -.26 & 1.76 \\ .31 & -.55 \\ -.46 & -.48 \\ 1.54 & -.15 \end{bmatrix}$$

Note that the X and Y values in (48d) and (48e) are taken from Table 2, while the composite scores, P and Q, turn out to be identical with the "true" factor scores in Table 1. In this artificial problem, the CCA has reproduced the factor scores exactly from an analysis of the observed x and y scores.

## Alternative Scaling of Factor Scores

Some authors prefer to set the variances of the composites equal to the respective canon $R^2$ values: var $P_j$ = var $Q_j$ = $\lambda_j$. This can be done by replacing $\underline{V}$ with $\underline{V}^* = \underline{V}\ \underline{\lambda}^{\frac{1}{2}}$ and $\underline{W}$ with $\underline{W}^* = \underline{W}\ \underline{\lambda}^{\frac{1}{2}}$ where $\underline{\lambda}^{\frac{1}{2}} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} & 0 \\ 0 & \lambda_2^{\frac{1}{2}} \end{bmatrix}$. Koons (1962) calls $\underline{V}^*$ and $\underline{W}^*$ the "absolute" regression weights and $\underline{V}$ and $\underline{W}$ the "normalized" regression weights. The use of the term "normalized" could be confusing since ordinarily a normalized vector is one in which the sum of squares of the elements equals one. The sum of squares of a vector of $\underline{V}$ and $\underline{W}$ weights does not equal one, nor does the sum of squares of the composite scores weighted by $\underline{V}$ and $\underline{W}$ equal one. $\underline{V}$ and $\underline{W}$ are scaled so the variance, not the sum of squares, of the composite scores is one.

## Canonical Loadings Within Sets

The matrices $\underline{V}$ and $\underline{W}$ are regression weights obtained in the usual textbook methods of analysis. For interpretative purposes loadings within sets may be calculated directly utilizing the definition of a loading as a correlation between an observed score and a factor score. The loadings of the observed predictors (x's) on the composite predictors (p's) is given by:

$$\text{(49a)}\quad \text{CCA:}\quad \underline{F} = (1/n)\ \underline{X}'\underline{P} = (1/n)\ \underline{X}'\underline{X}\ \underline{V} = \underline{R}_{11}\underline{V} = \begin{bmatrix} .81 & -.60 \\ .28 & -.96 \end{bmatrix}.$$

These loadings make it possible to express the observed scores as a function of the factor scores. Interchanging the right and left expressions of (49a):

(49b) CCA: $(1/n)\ \underline{X}'\ \underline{P} = \underline{F}$,

(49c) CCA: $(1/n)\ \underline{P}'\ \underline{X} = \underline{F}'$,

(49d) CCA: $(1/n)\ \underline{P}\ \underline{P}'\ \underline{X} = \underline{P}\ \underline{F}'$, and since $\underline{P} = \underline{X}\ \underline{V}$,

(49e) CCA: $(1/n)\ \underline{X}\ \underline{V}\ \underline{V}'\ \underline{X}'\ \underline{X} = \underline{P}\ \underline{F}'$,

(49f) CCA: $\underline{X}\ \underline{V}\ \underline{V}'\ \underline{R}_{11} = \underline{P}\ \underline{F}'$, and from Appendix 2, $\underline{V}\ \underline{V}' = \underline{R}_{11}^{-1}$,

hence

(49g) CCA: $\underline{X} = \underline{P}\ \underline{F}'$, or in ordinary notation:

(49h) CCA: $x_1 = f_{11}p_1 + f_{12}p_2 = (.81)p_1 + (-.60)p_2$,

(49i) CCA: $x_2 = f_{21}p_1 + f_{22}p_2 = (.28)p_1 + (-.96)p_2$.

The numerical values on these last two equations are <u>identical</u> with those for $x_1$ and $x_2$ in (15a) and (15b).

The loadings of the observed criteria (y's) on the composite criteria (q's) are:

$$\text{(50a)}\quad \text{CCA:}\ \underline{G} = (1/n)\ \underline{Y}'\underline{Q} = (1/n)\ \underline{Y}'\underline{Y}\ \underline{W} = \underline{R}_{22}\ \underline{W} = \begin{bmatrix} .87 & .49 \\ -.96 & .29 \end{bmatrix},\ \text{and}$$

(50b) CCA: $\underline{Y} = \underline{Q}\ \underline{G}'$, or

(50c) CCA: $y_1 = g_{11}q_1 + g_{12}q_2 = (\ .87)q_1 + (.49)q_2,$

(50d) CCA: $y_2 = g_{21}q_1 + g_{22}q_2 = (-.96)q_1 + (.29)q_2,$

exactly as in the equations for the y's in the right half of equations (15a) and (15b).

The use of loadings within sets for interpretation will be taken up more thoroughly in connection with an empirical problem. One property of such loadings may be pointed out here, however:

(51) CCA: $\underline{F}\ \underline{F}' = \underline{R}_{11}$, and $\underline{G}\ \underline{G}' = \underline{R}_{22}$.

That is, the sum of crossproducts of any two rows of $\underline{F}$ (or $\underline{G}$) gives the correlation between the observed variables whose loadings are given on those two rows, as was previously shown in equation (6c).

Canonical Loadings Between Sets

The loadings of the x's on the q's are found by:

(52a) CCA: $\underline{M} = (1/n)\underline{X}'\ \underline{Q} = (1/n)\underline{X}'\underline{Y}\ \underline{W} = \underline{R}_{12}\underline{W},$

and the loadings of the y's on the p's by:

(52b) CCA: $\underline{N} = (1/n)\underline{Y}'\underline{P} - (1/n)\underline{Y}'\underline{X}\ \underline{V} = \underline{R}_{21}\ \underline{V}.$

Actual calculation of these loadings is facilitated by the relations:

(52c) CCA: $\underline{M} = \underline{F}\ \underline{\Lambda}^{(1/2)}$ and $\underline{N} = \underline{G}\ \underline{\Lambda}^{(1/2)}$ (proof in Appendix II).

Since the y variables have been designated as criteria, only the $\underline{N}$ loadings will be discussed here (although the x's could just as well be designated as criteria and $\underline{M}$ discussed in the same way).

The numerical values for $\underline{N}$ are:

$$(53) \quad \text{CCA:} \quad \underline{N} = \begin{bmatrix} .87 & .18 \\ -.96 & .11 \end{bmatrix}.$$

The values in the first row of $\underline{N}$ are the correlations of $y_1$ with $p_1$ and with $p_2$: $r_{y_1p_1} = .87$ and $r_{y_1p_2} = .18$. The sum of squares of the loadings between sets, $(.87)^2 + (.18)^2 = .79$, is the squared multiple correlation between $y_1$ and the x's. Similar remarks apply to $y_2$ with respect to the second row of the matrix $\underline{N}$ in (53). It is proved in Appendix II that these relationships are true in general.

In the next chapter, loadings within and between sets will be discussed more fully in connection with a CCA of empirical data.

## Significance of Canonical Correlations

Bartlett (1947) has provided an approximate significance test for the canon R's which is summarized in Cooley and Lohnes (1962). If N = the total number of subjects, $m_1$ = the number of x variables, $m_2$ = the number of y variables, where $m_2 \leq m_1$, and $\lambda_j$ = canon $R_j^2$, then:

$$(54a) \qquad \Lambda = \prod_{j=1}^{m_2} (1 - \lambda_j), \text{ and}$$

$$(54b) \qquad \text{Chi-square} = -\left[N - .5(m_1 + m_2 + 1)\right] \log_e \Lambda$$

with $m_1m_2$ degrees of freedom. A significant chi-square leads to rejection of the null hypothesis of no relation between the two sets. If chi-square is significant, the first root may be removed from (54a) so that

$$\text{(55a)} \qquad \Lambda' = \prod_{j=2}^{m_2} (1 - \lambda_j), \text{ and}$$

$$\text{(55b)} \quad \text{Chi-square} = -\left[N - .5(m_1 + m_2 + 1)\right] \log_e \Lambda'$$

with $(m_1 - 1)(m_2 - 1)$ degrees of freedom. This process can be repeated until an insignificant chi-square is obtained or until all the roots have been removed except the last. The tests for the canon R's from the current problem are given in Table 5. Because

TABLE 5.-- Significance tests for canonical correlations from artificial data problem

| Number of Roots Removed | Largest Remaining Root (Canon $R^2$) | 1 - canon $R^2$ | Lambda | Chi-square | df | p |
|---|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | $\infty$ | 4 | <.01 |
| 1 | .14 | .86 | .86 | .38 | 1 | n.s. |

the largest canon R is 1.00 when rounded to two decimals, the resultant $\Lambda' = 0.00$ and chi-square is infinite. Carrying more decimals would result in a very large chi-square value which is clearly significant. The second root is not significant.

## Intercorrelations of Canonical Variates

To complete the comparison and contrast of MRA and CCA, Table 6 is presented for consideration as an analogue of Table 4. In Table 6, the intercorrelations between the composite predictors and composite criteria are shown. The canonical correlation coefficients are underlined in Table 6, just as the multiple correlation coefficients were underlined in Table 4. In Table 6 it can be seen that there are no relationships among the variables

TABLE 6.--Correlation matrix for canonical variates (for comparison with Table 4)

| | $p_1$ | $p_2$ | $q_1$ | $q_2$ |
|---|---|---|---|---|
| $p_1$ | 1.00 | .00 | <u>1.00</u> | .00 |
| $p_2$ | | 1.00 | .00 | <u>.37</u> |
| $q_1$ | | | 1.00 | .00 |
| $q_2$ | | | | 1.00 |

except for the specific relationships between the two matched predictor-criterion pairs. The relationships in Table 6 are clear-cut, in contrast to Table 4 where everything is fairly highly correlated with everything else. The results of a CCA should in theory at least, be conceptually much simpler than the results of several MRAs involving the same set of predictors and a series of single criterion variables.

## The Case in Which $m_2 \neq m_1$

In the present problem and in the one which follows the x and y sets contain equal numbers of variables. That is, $m_2 = m_1$. If there are more predictors than criteria, then $m_2 < m_1$, which makes no practical difference in the computations as outlined here. The characteristic equation (38) will still yield the appropriate number of latent roots and vectors. The case of $m_2 < m_1$ does make a difference in some of the algebraic proofs, and for this reason it is discussed in Appendix II.

If the number of criteria exceed the number of predictors, then $m_2 > m_1$, and in this case the characteristic equation (38) would lead to extra computational labor. It is therefore advisable to disregard the labels "predictor" and "criteria" and simply define the y set as being smaller than or equal to the x set, so that $m_2 \leq m_1$ always holds.

In other words, if there are more predictors than criteria, define x as the predictor set and y as the criterion set. If there are more criteria than predictors, define x as the criterion set and y as the predictor set. $\underline{R}_{11}$ will then always be the matrix of intercorrelations of the variables in the larger set, and $\underline{R}_{22}$ the matrix of intercorrelations of the smaller set.

## A Computer Program for Canonical Correlation Analysis

A computer program for CCA has been written by Mitchell (1965) and is available from the IBM 1620 Users Group Library. This program calculates and prints out means, standard deviations, all the canon R's and their significance tests, and the complete matrices of simple correlations, weights, loadings between sets, and factor scores ($\underline{P}$ and $\underline{Q}$) for all subjects. The program does not print out loadings within sets, though these can easily be calculated by dividing the loadings between sets for a pair by the canon R for that pair. The program also prints out the $\underline{R}_{11}^{-1}\ \underline{R}_{12}$ matrix of multiple regression weights, and could be modified fairly easily also to print out $\underline{R}_{2.1}$, so that a complete MRA would be available in addition to the complete CCA. The program is written in FORTRAN for the IBM 1620 with card input; output is via cards and IBM typewriter. In its present form the program is suitable for 60K storage, but it can be adapted to 20K or 40K, and it will take up to a total of 50 variables divided into two sets which may be equal or unequal in size.

# CHAPTER IV

## APPLICATION OF CANONICAL CORRELATION ANALYSIS TO EMPIRICAL DATA

In this chapter CCA will be applied to data from a study of the effects of prolonged chlorpromazine treatment on a sample of 51 chronic schizophrenic female patients in a state mental hospital. The study was double-blind and 21 concurrent placebo-control subjects were also used. A detailed report of the selection of subjects and the experimental design is found in Clark _et al_, (1967).

Among the measures used in the study was a Psychologist's Scale, with five items labelled as follows:

1. Ideational fluency
2. Expressiveness and gestures
3. Affective status
4. Motor activity
5. Ease of testing

Each item is ratable on a 13-point scale from -6 to +6, with zero representing the rating assigned a hypothetical normal person. The entire scale, with concrete descriptions of the rating categories, is given in Appendix III.

Each patient was rated on each item of this scale at "baseline" (prior to start of treatment) and again at "peak dose" (after 12 weeks of treatment). Ratings were made on the basis of behavior observed during the administration of some simple psychological tests. Response to treatment was expressed in terms of change scores adjusted for changes also found in the placebo-control group (these are called "psi" scores in Ragland & Costiloe, 1963, where their derivation is given). Each of the treated patients, therefore, had five baseline scores and also five adjusted change scores on the same scale items. The baseline scores constitute a battery of predictors and the change scores a battery of criterion variables.

Some questions which are of interest in this situation include:

1. How well can response to treatment be predicted?
2. When the batteries are factored by CCA, what is the process by which the factors are interpreted?
3. What are the interpretations for this set of data?

## Prediction

The first question can be answered in terms of each criterion variable separately and of the criterion battery as a whole, and also in terms of factors derived from the two batteries.

### Simple Correlations and $\overline{r^2}$

The $\underline{R}_{12}$ matrix of simple correlations, shown in the upper right corner of Table 7, provides one answer to the question of how well response can be predicted. The highest correlation in the first column of $\underline{R}_{12}$ is underlined--it is the correlation of -.51 between baseline "Ideational fluency"and adjusted change on this same item. When squared, it indicates that about 26% of the criterion variance for this item is accounted for by the predictor. The highest correlations in subsequent columns of $\underline{R}_{12}$ are underlined. In every case

TABLE 7.--Psychologist's Scale: Upper triangular matrix of simple correlations among predictors and criteria (decimal points omitted)

| | Predictors | | | | | Criteria | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Idea | Expr | Aff | Mtr | Ease | Idea | Expr | Aff | Mtr | Ease |
| Idea | 100 | 54 | 51 | 55 | 73 | -51 | -37 | -45 | -34 | -46 |
| Expr | | 100 | 71 | 76 | 60 | -19 | -63 | -55 | -49 | -32 |
| Aff | | | 100 | 81 | 67 | -15 | -35 | -80 | -38 | -40 |
| Mtr | | | | 100 | 72 | -25 | -45 | -72 | -63 | -51 |
| Ease | | | | | 100 | -31 | -36 | -59 | -40 | -65 |
| Idea | | | | | | 100 | 54 | 40 | 41 | 62 |
| Expr | | | | | | | 100 | 57 | 49 | 59 |
| Aff | | | | | | | | 100 | 38 | 62 |
| Mtr | | | | | | | | | 100 | 45 |
| Ease | | | | | | | | | | 100 |

they fall on the diagonal, which means that change in each item is predicted best by initial status on that item.

Prediction with MRA will be discussed later, since the mult R's occur as a by-product of CCA.

A rough index of how well response in an overall sense can be predicted can be provided by taking the sum of the squares of the underlined elements of $\underline{R}_{12}$, divided by the total variance of the criterion set. The total variance is 5.00, since each score contributes a variance of 1.00. Thus:

$$(56)\quad \overline{r^2} = (1/5)\left[(-.51)^2 + (-.63)^2 + (-.80)^2 + (-.63)^2 + (-.65)^2\right]$$

$$\overline{r^2} = .42.$$

Since each squared element gives the proportion of variance in one criterion measure accounted for by one predictor measure, $\overline{r^2}$ gives a crude estimate of the proportion of variance in the entire criterion set accounted for by the entire predictor set. $\overline{r}^2$ is not a precise estimate, however, because the squared correlations which go into it do not in general represent independent components of variance, since the criterion variables are intercorrelated and thus share variance among themselves.

The average squared multiple correlation will be taken up in connection with CCA. It too involves the adding together of non-independent portions of variance.

Canonical Correlation: Weights and Loadings Within Sets

The results of the CCA are given in Tables 8-12. The tests of significance summarized in Table 8 show that all five canon R's are significant. More exactly, all five of the roots account for a significant amount of variance between sets, and as each of the

TABLE 8.--Psychologist's Scale: Tests of significance for canonical correlations

| Number of Roots Removed | Largest Remaining Root (canon $R^2$) | 1 - canon $R^2$ | Lambda | Chi-square | df | p |
|---|---|---|---|---|---|---|
| 0 | .80 | .20 | -.03 | 157.5 | 25 | .01 |
| 1 | .58 | .42 | .16 | 85.0 | 16 | .01 |
| 2 | .38 | .62 | .37 | 45.5 | 9 | .01 |
| 3 | .30 | .70 | .59 | 24.1 | 4 | .01 |
| 4 | .16 | .84 | .84 | 7.7 | 1 | .02 |

first four roots is removed, there still remains a significant amount of variance accounted for by the ones, or the one, remaining.

In Table 9, the canonical regression weights for the predictor set are in the upper half of the table, and the weights for the criteria are in the lower half. Table 10 gives the loadings within sets for the two sets in the upper and lower halves of the table, and also sums of squares for rows and columns. Each half of Table 10 follows the usual format for reporting the results of an FA, except that an extra row has been added at the bottom of each half of the table. These extra rows will be explained later.

TABLE 9.--Psychologist's Scale: Canonical Weights

| | Factor I | Factor II | Factor III | Factor IV | Factor V |
|---|---|---|---|---|---|
| Predictors | | | | | |
| 1. Idea | -.18 | - .08 | - .09 | -1.46 | - .30 |
| 2. Expr | - .60 | -1.28 | .26 | .44 | - .61 |
| 3. Aff | .76 | .21 | -1.50 | .12 | - .61 |
| 4. Mtr | .62 | - .13 | .73 | - .64 | 1.69 |
| 5. Ease | .21 | .69 | .93 | 1.18 | - .72 |
| Criteria | | | | | |
| 1. Idea | .25 | - .27 | .13 | 1.26 | .20 |
| 2. Expr | .55 | 1.07 | - .19 | -. 31 | .63 |
| 3. Aff | - .99 | .12 | .87 | .21 | .14 |
| 4. Mtr | - .34 | .35 | - .43 | .20 | - .97 |
| 5. Ease | - .24 | - .81 | - .99 | - .73 | .33 |

Comparison of Table 9 and 10 shows that a variable with a high loadings on a given factor will not necessarily have a high weight on that same factor. For example, on predictor factor I, the "Ease of testing" item has a relatively low weight of .21 and a relatively high loading of .68. A low weight can mean either that a variable is irrelevant to the factor, or that it has been partialed out because

TABLE 10.--Psychologist's Scale: Canonical loadings within sets, sums of squares, and calculations leading to $H^2_{21}$ and $H^2_{12}$

| | Factor I | Factor II | Factor III | Factor IV | Factor V | Sum of Squares |
|---|---|---|---|---|---|---|
| Predictors | | | | | | |
| 1. Idea. | .38 | -.07 | .36 | -.64 | -.55 | 1.00 |
| 2. Expr. | .44 | -.78 | .26 | -.04 | -.35 | 1.00 |
| 3. Aff. | .88 | -.32 | .14 | -.05 | -.31 | 1.00 |
| 4. Mtr. | .84 | -.40 | .34 | -.15 | .05 | 1.00 |
| 5. Ease | .68 | .02 | .55 | -.01 | -.49 | 1.00 |
| Sum of Squares | 2.29 | .88 | .63 | .45 | .76 | 5.00 |
| Proportion of Total Variance | .46 | .18 | .13 | .09 | .15 | 1.01 |
| $R^2_k$ (Proportion of Total Variance) | .39 | .10 | .05 | .03 | .02 | .59 = $H^2_{12}$ |
| Criteria | | | | | | |
| 1. Idea. | -.14 | -.00 | -.42 | .80 | .40 | 1.00 |
| 2. Expr. | -.19 | .68 | -.44 | .14 | .54 | 1.00 |
| 3. Aff. | -.86 | .25 | .03 | .16 | .41 | 1.00 |
| 4. Mtr. | -.46 | .44 | -.60 | .30 | -.38 | 1.00 |
| 5. Ease | -.54 | -.11 | -.69 | .08 | .48 | 1.00 |
| Sum of Squares | 1.30 | .73 | 1.20 | .79 | .98 | 5.00 |
| Proportion of Total Variance | .26 | .15 | .24 | .15 | .20 | 1.00 |
| $R^2_k$ (Proportion of Total Variance) | .21 | .08 | .09 | .05 | .03 | .46 = $H^2_{21}$ |

it correlates highly with some other variable which <u>does</u> have a high weight on that factor. These are two very different circumstances, but they can not be discriminated simply by looking at the weight. If a variable has both a low weight and a low loading on a given factor, it can be clearly stated that it is irrelevant to that factor. If it has a low weight and a high loading, it can be clearly stated that it has been partialed out in the calculation of the factor scores for that factor. Though partialed out, it is still of use in interpreting the factor.

Canonical Correlation: Loadings Between Sets and Mult $R^2$

The complete CCA provides information about the prediction of the original criterion variables taken singly, which may be of interest in some cases, though this is not the prime purpose of CCA. In Table 11 the loadings between sets for the criterion set are

TABLE 11.--Psychologist's Scale: Loadings between sets for the criterion set

| Criteria | Factor I | Factor II | Factor III | Factor IV | Factor V | Sum of Squares (Mult $R^2$'s) |
|---|---|---|---|---|---|---|
| 1. Idea | -.13 | -.00 | -.26 | .44 | .16 | .44 |
| 2. Expr | -.17 | .52 | -.27 | .08 | .21 | .54 |
| 3. Aff | -.77 | .19 | .02 | .09 | .16 | .66 |
| 4. Mtr | -.41 | .33 | -.37 | .17 | -.15 | .46 |
| 5. Ease | -.48 | -.08 | -.42 | .04 | .19 | .45 |

given. These are the correlations of the original (observed) criterion variables with each of the factored predictors. The entry in row one, column one is -.13, which is the correlation between adjusted change in "Ideational fluency" and the first predictor factor. The rest of row one contains the correlations of this change item with the other four predictor factors, and the sum of squares of this row is the squared multiple correlation between this first change item and all the original predictors.

In effect, what is illustrated in Table 11 is how CCA takes the original multiple regression equation and breaks it into five orthogonal components, each of which then predicts an independent portion of each of the original criterion variables.

From the last column of Table 11, $\overline{\text{mult } R^2}$ can be calculated as an index of overall goodness of prediction. In this case it is:

$$(57)\qquad \overline{\text{mult } R^2} = (1/5)(.44 + .54 + .66 + .46 + .45),$$
$$\overline{\text{mult } R^2} = .51.$$

Canonical Correlation: Prediction of Factors

When the predictor and criterion are both factored by CCA, the canon $R^2$'s show how well each change factor can be predicted by its associated predictor factor. The five significant canon $R^2$'s, from Table 8, are .80, .58, .38, .30, and .16. The first of these is higher than any of the squared simple or multiple correlations obtained from these data. The square roots of the above values give the five canon R values of .89, .76, .61, .55, and .39. (This author prefers to work directly with squared correlations, since these provide a direct estimate of shared variance.)

Canonical Correlation: $H^2$ as an Overall Index of Relationship

To show how CCA can furnish an estimate of how well the criterion set as a whole can be predicted, it is necessary to discuss Table 10 in more detail. Since the Table consists of two complete factor analyses, only the bottom half pertaining to the criterion set will be explained. Similar remarks would apply to the top half.

The first row of the lower half of Table 10 gives the within-sets loadings for "Ideational fluency." These are -.14, -.00, -.42, .80, and .40, and their sum of squares, as indicated in the last column, is 1.00, which is the total variance of this variable. The last three of the five loadings are of substantial size, which means that the bulk of the variance for this variable is distributed among the last three change factors.

Subsequent rows in the lower half of Table 10 show how the variance of each of the other four change score items is distributed among the five change factors. The sum of squares of each row is one, and the total sum of squares is 5.00, which is the total variance of the criterion set.

The first column in the lower half of Table 10 gives the loadings of all five change items on the first canonical change factor. These are -.14, -.19, -.86, -.46, and -.54, and their sum of squares is 1.30. This sum of squares, when divided by the total of 5.00, gives the proportion of the total criterion variance contributed by the first change factor, which in this case is .26.

Subsequent columns contain the loadings for change factors II through V, with the column sums of squares and proportion of total variance contributed by each factor given at the bottom of each column. Whereas the row sums of squares are always 1.00 (within rounding error), the column sums of squares may not be, since some factors may contribute more to the overall variance than do others. This uneven contribution of the factors is obscured when the factor scores are standardized or when their variances are arbitrarily set equal to the corresponding canon $R^2$'s. The next to the last row at the bottom of Table 10 shows that the five change factors respectively contribute proportions of .26, .15, .24, .15, and .20, and these add up to 1.00 or 100% of the total variance.

The entries in the last row at the bottom of Table 10 shows how much of the total criterion variance is accounted for by each predictor factor. The first entry, .21, is arrived at as follows: the first predictor factor accounts for .80 of the variance of the first criterion factor (canon $R_1^2 = .80$), and in turn the first criterion factor contributes .26 to the total criterion variance; therefore the first predictor factor accounts for (.80)(.26) = .21 of the total criterion variance. In the same way, predictor factor II through its relationship with criterion factor II accounts for (.58)(.15) = .08 of the total criterion variance, and this .08 is independent of the previous .21 because factors I and II are orthogonal. The remaining three predictor factors also account for independent and significant portions of the total criterion variance because they are orthogonal to the first two factors, and to one another, and because their associated canonical correlations are significant.

The sum of all the entries in the last row of Table 10 is .46. This sum, symbolized by $H^2_{21}$, is analogous to a squared correlation between criterion and predictor sets _each_ _taken as_ _a_ _whole_. It is to be compared with the $\overline{r}^2$ and $\overline{\text{mult } R}^2$ values of .42 and .51 computed previously, but is a more precise index of overall goodness of prediction since it contains no redundant variance. The computing formula is:

$$(58)\quad H^2_{21} = \sum_{k=1}^{s} \text{canon } R^2_k \left[ \sum_{j=1}^{m_2} g^2_{jk} \right]$$

where s is the number of _significant_ canon R's and $m_2$ is the number of variables in the criterion set.

The subscript 21 on H indicates that it is the variance of the second battery which is being accounted for by the first. In general, this will not be the same as the proportion of variance in the first battery accounted for by the second, because of differences in factor structure in the two batteries, or because of a different number of variables in the two sets, or both. That is, $H^2_{21} \neq H^2_{12}$ in general. For these data $H^2_{12} = .59$, which is the sum of entries in the last row in the top half of Table 10. The difference between $H^2_{12}$ and $H^2_{21}$ in this instance is largely due to the fact that predictor factor I has a much larger sum of squares (2.29) than that of criterion factor I (1.30).

The question of how well response can be predicted has now been answered exhaustively in terms of simple, multiple, and canonical correlations and $H^2_{21}$, and in terms of the original variables and of factor variables derived from them. More will be said about these measures of predictiveness in the section on interpretation of the factors.

## The Process of Interpretation

Most published reports of factor analyses proceed with interpretation on the basis of loadings such as are given in Table 10. In many cases loadings smaller than some arbitrary absolute magnitude (it is often $\pm$ .30) are considered negligible. Rather than set such a cutting point which is applied to the whole table of loadings, it might be more meaningful to consider each loading in relation to the size of the factor of which it is a part.

In the upper half of Table 10, for example, "Ideational fluency" has loadings of .38 on factor I and .36 on factor III. However, factor III has a much smaller sum of squares than factor I, so that relatively speaking the loading of .36 is much more important than the loading of .38. If the loadings were expressed as percentages of their respective factor sums of squares, their true importance in the analysis would be more readily apparent. Thus the loading of .38 contributes (100)(.38)/(2.29) = 6% to the factor I sum of squares, while the loading of .36 contributes (100)(.36)/(.63) = 21% to the factor III sum of squares. Therefore "Ideational fluency" may be said to be of negligible value in defining factor I, while it seems quite important in defining factor III.

In Table 12 the loadings from Table 10 have been converted

TABLE 12.--Psychologist's Scale: Percentage of sum of squares contributed by each item to each factor [a,b]

| | Factor I | Factor II | Factor III | Factor IV | Factor V |
|---|---|---|---|---|---|
| Predictors | | | | | |
| 1. Idea | | | 21 | -93 | -41 |
| 2. Expr | | -70 | | | -16 |
| 3. Aff | 34 | | | | |
| 4. Mtr | 31 | -18 | 18 | | |
| 5. Ease | 20 | | 48 | | -32 |
| Total | 85 | 88 | 87 | 93 | 88 |
| Criteria | | | | | |
| 1. Idea | | | -15 | 82 | 16 |
| 2. Expr | | 64 | -15 | | 29 |
| 3. Aff | -57 | | | | 17 |
| 4. Mtr | -16 | 26 | -30 | | |
| 5. Ease | -22 | | -39 | | 23 |
| Total | 95 | 90 | 100 | 82 | 85 |

[a] Entries smaller than 15 have been omitted.

[b] A negative percentage means that the loading from which the percentage was calculated was negative.

to percentages of the respective factor sums of squares in the maner just described. The first column in the top half of Table 12 shows that 85% of the sum of squares for predictor factor I comes from three items--"Affective status," "Motor activity," and "Ease of testing." These same three items, as adjusted change scores, make up 95% of the sum of squares for criterion factor I in the lower half of the table. Factor I, both with respect to initial status and change, will be interpreted in terms of what these three items have in common, exclusive of what they contribute to the other factors.

Each of the remaining four factor pairs, like the first, involve the same items in both the predictor and criterion sets. Factor IV, for example, is made up largely of "Ideational fluency" in both halves of the table. It is not accurate to say therefore that factor IV means the same at the original item, since this item also contributes to factors III and V. For purposes of interpretation, an attempt must be made conceptually to split the item into three uncorrelated components to be parceled out to factors III, IV, and V.

## Interpretation of the Factors

Considering then each column of Table 12 in the light of the table as a whole, the following interpretations are offered for the factors.

Factor I. Elation-depression. This is affective status along the elation-depression dimension, expressed more motorically than verbally. To the extent that it may be expressed verbally it has nothing to do with rate of speech. Its extreme manifestations can interfere with the process of testing. Initial status on this factor is predictive of change in this same factor, and the relation is negative, since the loadings have opposite signs in the upper and lower halves of Table 12. Prediction is better than with the "Affective status" item which contributes most heavily to this factor: the canon $R^2$ is .80, while the mult $R^2$ for the item (from Table 11) is .66 and the highest simple $r^2$

(from Table 7) is $(-.80)^2 = .64$.

Factor II. Psychotic posturing. This is what the "Expressiveness" item was intended to measure, but the item also picks up motoric restlessness, and that part of it is relegated to factor V. This factor pulls psychotic posturing out of "Expressiveness" and to a lesser extent out of the "Motor activity" item. It predicts change in itself better than the "Expressiveness" item does: the canon $R^2$ is .58, the mult $R^2$ for the item is .54, and the highest simple $r^2$ for the item is .40.

Factor III. Undercontrol vs. overcontrol of behavior. This factor has to do with ability to regulate both verbal and motor behavior in a situation where there is a demand placed on one to do so (the testing situation). It is unrelated to elation-depression and to psychotic posturing as such, but rather reflects the patient's capacity to modulate behavior (through tightening or easing of controls) in order to meet the task demands despite her standing on the first two factors. Change in this factor is not quite as predictable as change in the "Ease of testing" item which contributes most heavily to it: the canon $R^2$ is .38, the mult $R^2$ for the item is .45, and the highest simple $r^2$ is .42.

Factor IV. This is Verbal speed and spontaneity, apart from verbal expression of affect and apart from comprehensibility of speech which would make testing more or less difficult. The canon $R^2$ is .30, while the mult $R^2$ for the "Ideational fluency" item which loads highly on this factor is .44 and the highest simple $r^2$ for the item is .26.

Factor V. General restlessness. This is the most tentatively defined of the factors, and it is defined in terms of what seems to be left over from the other factors. It could be dismissed simply as "error" except for two related facts: it contributes more to the total sum of squares (in both the predictor and criterion sets) than some of the previous factors, and its canon R is significant. The canon $R^2$ is .16. No single item loads highest on this factor in both sets.

## Discussion of results

The results indicate that the Psychologist's Scale is useful in the study of prediction of response to chlorpromazine treatment in chronic schizophrenia. In particular, response along the bipolar dimensions of elation-depression and psychotic posturing may be predicted fairly well, since persons who are initially toward either extreme on these dimensions will tend to shift toward the middle (normal) as a result of treatment. Of course there is a natural tendency for extreme ratings to shift toward the middle on a bipolar scale, but this natural tendency was partialed out through the use of adjusted change scores. Thus the shift actually found was due to treatment.

To apply the above findings with respect to elation-depression to a new sample, it would suffice to administer the Psychologist's Scale to each new subject prior to treatment, and to calculate a $p_1$ score using the weights from column one in the top half of Table 9. Letting $x_1$ , $x_2$, ... , $x_5$ stand for the pretreatment scores on the five items, then for each subject:

$$(59) \quad p_{i1} = (-.18)x_{i1} + (-.60)x_{i2} + (.76)x_{i3} + (.62)x_{i4} + (.21)x_{i5}.$$

If the resultant $p_{i1}$ score for a given subject were near zero, one would not recommend treatment, since

$$(60) \quad \text{predicted } q_{i1} = \hat{Q}_{i1} = (\text{canon } R_1)p_{i1} = .89p_{i1},$$

so that if $p_{i1}$ is near zero, the predicted amount of change is also near zero. If however $p_{i1}$ is quite deviant from zero, one would expect good results from treatment. (Cross-validation with a new sample of subjects will be taken up in the next chapter.)

Response in terms of the bipolar dimensions of undercontrol-over-control, verbal speed and spontaneity, and general restlessness is not predictable in a practical sense, in each case a significant but relatively small amount of the criterion variance is accounted for. On the basis of this analysis alone, chlorpromazine treatment would seem to be more strongly indicated for schizophrenics who manifest symptoms along the first two dimensions than for those whose symptoms are primarily in the last three categories. Before making such a definite conclusion, however, it would be well to look at mean change along the last three dimensions, since it is possible to have a significant mean change and still have little correlation between initial value and change. Such a thing could happen if change was constant, regardless of starting level.

Apart from the pragmatic issue of predicting response, knowledge of the factor structure of the scale may be of interest. Such knowledge could lead to an "orthogonalizing" of the scale; that is, a re-wording of the items so that they more clearly reflect the factored behavior categories rather than the original categories. (Of course such a re-wording would mean that in effect a new scale had been created, and this scale would have to be validated in a separate study.) Or the scale could be used as it stands in future studies, and the ratings converted to factor scores using the weights from Table 9, as illustrated in equation (59).

It cannot be concluded from this analysis that the five factors are truly more basic, "real," or substantive than the original five items. Such a conclusion would only perhaps be defensible after a series of analyses in which different samples of subjects were used and in which other measures besides the Psychologist's Scale were included. The factors from this analysis represent an alternative way of dividing up the total variance which may have more pragmatic and heuristic value than the original. In any case, further study of the scale would seem desirable.

# CHAPTER V

## DISCUSSION

It has been illustrated through the use of artificial data how CCA is an extension of MRA to the case of multiple criteria, and also how CCA can reproduce the factor structure known to underly a set of variables. The author has applied CCA to a number of other artificial problems (not reported here), each using normally distributed factor variables, varying numbers of "observed" variables in each set, an N of 50 observations, and an explicit error term (see discussion of error below). In each case the loadings within sets were good approximations of the original factor loadings.

It has also been shown in this dissertation that CCA, when applied to empirical data can yield factors capable of meaningful interpretation. A method for interpreting factors in terms of the percentage of sum of squares contributed by each variable to each factor, rather than in terms of loadings, has been suggested. Finally, $H^2$ has been proposed as an overall index of association between batteries which takes factor structure into account as well as the number of significant canonical correlations. In this chapter, a number of practical and theoretical issues raised by CCA will be discussed, and recommendations made for further research.

## Methods of Cross-Validation

When CCA is used for prediction, the question of cross-validation arises. That is, how well will the results of this analysis hold up with a new sample of subjects? At least two approaches to cross-validation are possible with CCA. The most direct approach would be to repeat the entire analysis using the same two sets of variables and a new sample of subjects. The loadings within sets could be compared to see if the same factor-pairs emerged from both analyses, and $\underline{\Lambda}$, the diagonal matrix of canon $R^2$'s, could be compared for the two analyses. All comparisons would be more qualitative than quantitative; no statistical tests for differences between loadings or between canon R's from separate analyses are known to this author.

A second approach to cross-validation with CCA would be to consider the criterion weights (one set for each factor) from the first analysis to be sufficiently good estimates of the "true" weights that no further validation is needed for them. In the new sample, one could apply these weights to the criterion variables to get a set of composite criteria (q's), for however many of the composites might be of interest in a particular study. A separate MRA could then be carried out for predicting each q from all the observed predictors (this could actually be accomplished by simultaneous MRAs using the $\underline{R}_{2.1}$ matrix, as indicated in Chapter II). The mult R's from the new sample could then be compared with the corresponding canon R's from the original sample. This second approach would not directly produce any loadings in the MRA for comparison with those from the CCA. However, to the extent that the cross-validation holds up, the MRA weights for each composite predictor in the new sample should be roughly proportional to the corresponding predictor weights from the CCA.

The second approach to cross-validation is obviously cruder than the first, but it may be preferred as a less expensive alternative if funds or computer time (or both) are not available for repeating the CCA. A complete MRA is cheaper than a complete CCA. It may also be preferred if only the first few out of a large total number of factor-pairs are of interest, so that repetition of the entire CCA would involve a great deal of wasted computer time. It should be borne in mind that the second approach involves switching from a variable-predictor model (CCA) to a fixed-predictor model (MRA).

If the investigator is satisfied with the cross-validation, however it is carried out, he has in hand for future application the matrices of predictor and criterion weights. For simplicity, consider only $\underline{v}$ and $\underline{w}$, the vectors of weights associated with p and q. Given a new individual, who was neither in the original sample nor in the cross-validation sample, and given that a set of x-measurements is available for this individual, it is desired to predict his standing on $q_1$.

Now $p_1$ and $q_1$ are in standard form, and canon $R_1$ is the product-moment correlation between them as well as the regression weight for predicting one from the other. Since $p_1$ and $q_1$ are both free to vary, the bivariate correlation model is applicable, and

(61) $\hat{Q}_{i1} = (\text{canon } R_1)p_{i1}$, where $\hat{Q}_{i1}$ predicts $q_{i1}$,

which has the same form as equation (2) in Chapter I. To predict $q_1$ for individual $\underline{i}$, apply the vector $\underline{v}_1$ to his scores on the x variables to get his $p_1$ score, then apply equation (61) to get his score on $\hat{Q}_1$.

Note that $\hat{Q}_1$, which is not in standard form, predicts $q_1$, which is in standard form. If it is desired that the predictor of $q_1$ be in standard form, simply divide (61) through by canon $R_1$, and

(62) $\hat{q}_{i1} = \hat{Q}_{i1}/(\text{canon } R_1) = p_{i1}$.

That is, if one has the standardized predictor in hand, the best guess (assuming a linear relationship) as to the value of the standardized criterion is that it will be identical with the standardized predictor.

Using either (61) or (62), if a person is one standard deviation above the mean on $p_1$, it will be predicted that he is one standard deviation above the mean on the criterion. That is, if $p_{i1} = 1$, then $\hat{Q}_{i1}$ = canon $R_1$, and $\hat{q}_{i1} = 1$.

The vector $\underline{w}_1$ of weights for the y measurements do not come explicitly into the foregoing discussion at all. The y measurements are not actually taken; it is predicted that if they were taken, and if they were combined into one score using $\underline{w}_1$, then the resulting composite score is best estimated by $Q_{i1}$ or $q_{i1}$

(whichever the investigator prefers).

## Parsimony, Orthogonality, and Error

A common use made of FA is in "data reduction," in which a large number of variables is replaced by a smaller number of factors which account for almost all the variance in the original variables. If the sum of squares of loadings for a given factor is quite small in relation to the total sum of squares of loadings, that factor may be dismissed as due to "error" or simply as negligible. If several factors are eliminated in this way, then the investigator may claim that he has replaced the original variables with a more parsimonious set of factor variables which account for the same essential phenomena. Some factor analysts (Thurstone, 1947) feel that such data reduction is one of the chief functions of FA. The use of communalities rather than ones in the diagonal of the correlation matrix matrix is defended because (among other things) it enhances the prospects of eliminating factors.

The common factor model, as distinguished from the principal components FA model so far referred to in this dissertation, has a fundamental equation of the following form:

$$(63) \qquad x_{i1} = f_{11}p_{i1} + f_{12}p_{i1} + \ldots + f_{1r}p_{ir} + f_1 s_{i1} + e_{i1},$$

where there are r common factors underlying the x measurements, represented by the p's, a factor specific to $x_1$, represented by $s_1$, and an explicit error term. The f coefficient for $s_1$ has a single subscript because there is only one such coefficient for each specific factor. Only the common factors are analyzed, since there are more or less subjective methods for removing the variance contributed by the specific factors and by the error term. If the resulting number of common factors is fewer than the number of original variables, a claim for parsimony is made.

Actually, such a claim for parsimony ignores the fact that under the common factor model there are as many specific factors as there are original variables (Harman, 1960). Adding these to the common factors gives more factors than variables. Nor is all error removed from the common factor model by the explicit error term; the factor loadings resulting from the analysis are still estimates of population values. Finally, the use of the common factor model, rather than the FA model used here, tremendously complicates the calculations required for getting weights from loadings, and loadings from weights.

Using the principal components FA model or CCA, the investigator may find that in some cases some of the factors or factor-pairs extracted contribute very little to the total sum of squares and may be ignored. However, even if all the factors are relatively large, there may still be considerable value in having transformed the original variables into a set of orthogonal variables. The alternative structuring of the data may be conceptually simpler if not more parsimonious than the original data, and, in addition, orthogonal variables have convenient algebraic properties which may be desirable.

The question of whether "nature" is orthogonal or whether the structuring provided by an FA or CCA is more "real" or closer to nature need not be raised if the analysis provides results which are meaningful in terms of the context in which it is carried out.

With respect to CCA, Meredith (1964), in the article referred to in Chapter II, has shown one way in which error due to test unreliability can be removed from the analysis. For those who subscribe to the notion that an observed scores can be represented as the sum of a "true" and an "error" component, and who find the present methods of assessing test reliability acceptable, Meredith's treatment may prove satisfactory.

One effect which error may have on a CCA may be seen in Table 10 with reference to the analysis of empirical data. Predictor factor I and criterion factor I have the same structure--that is, "Affective status" has the highest loading in both, and "Motor activity" and "Ease of testing" are the next most important in defining the factor-pair. However, the loadings are all somewhat smaller in the criterion set, with the result that the criterion sum of squares of loadings for factor I is 1.30, as compared with a predictor sum of squares of loadings for factor I of 2.29. Thus error may lead to differences in saturation of members of a factor-pair without destroying essential structure. The same sort of effect might be seen in an FA repeated using the same variables and the same or different subjects.

### $H^2_{21}$ and Battery Reliability

If the two sets of variables in a CCA consist simply of the same battery of measurements given twice to the same subjects, or if the second set consists of "alternate forms" of the tests in the first set, then $H^2_{21}$ can be considered an index of the reliability of the entire battery. It would be expected that the factor structure would be much the same in the two batteries, though of course the two sets of loadings should be inspected to see if such is indeed the case. Even when the factor structure is the same, a factor may be more strongly saturated in one battery than in the other, as indicated in the preceding section. Such differences in saturation will of course affect the value of $H^2_{21}$.

## Limitations of Canonical Correlation Analysis

Now that electronic computers are becoming more generally available, it is possible to apply CCA to much larger problems than formerely. There are still some limitations with respect to sample size relative to number of variables in both sets, and with respect to arithmetic accuracy, which must be taken into account.

From looking at the chi-square test for significance of canonical correlations in equation (54b), it is clear that the number of subjects necessary for a valid test must at least be greater than $.5(m_1 + m_2 + 1)$ or the resulting chi-square will be negative. This seems to be a rather generous limitation. In practice, however, it has been the author's experience that N must be about twice as large as the total number of tests in both sets to detect significance. Canonical correlations of .99 have been found insignificant when the total number of tests approached the total number of subjects.

Computer accuracy is another factor which can cut down the size of problems even though the computer has the storage capacity for large problems. The CCA program for the IBM 1620 described at the end of Chapter III has the storage for up to fifty variables divided into two sets. However, with floating-point arithmetic and eight-digit accuracy, the results with problems larger than a total of twenty variables are questionable. The inversion of large matrices leads to considerable rounding error, and the iterative procedures for finding roots and vectors for a matrix lead to more such error. As a result, computer outputs of CCAs should not be accepted uncritically. The author has found a satisfactory check to consist of ascertaining that the row sums of squares of loadings within sets are all tolerably close to 1.00; this check could be built into future computer programs or (with small problems) made by desk calculator.

## Suggestions for Further Research

Horst (1961) has provided a generalization of CCA to more than two sets of variables. The treatment is largely mathematical, and the development is in terms of canonical weights with no consideration of problems of interpretation of results. It would be worthwhile to see if the approach to CCA in this dissertation could be extended to the general case developed by Horst, and to discover if new problems arise as a result of his generalization.

Another field requiring investigation is that of working out significance tests for the loadings within sets in CCA ( or alternatively for the weights). The significance tests known to this author, such as those of Lawley (given in Harman, 1960) or Rao (1958) assume a common factor model rather than the principal components or CCA model, though perhaps they could be adapted for use with the latter models.

It might also be desirable to develop a significance test for $H^2_{21}$ though it seems plausible to argue that if $H^2_{21}$ is calculated using only significant canon R's, one should have as much confidence in it as in the canon R's. In general, this dissertation has had little to say about tests of significance, estimation of parameters, or the distribution of such statistics as $H^2_{21}$. Such matters could profitably be pursued with more rigor now that the meaning of CCA and its relationships to more familiar methods of analysis have been clarified.

Finally, the present results could be related to "canonical analysis" in the general sense as discussed by Bartlett (1947) and McKeon (1962), in which the basic CCA model, through the use of pseudo-variates and conditional inverses, may be applied to a whole range of statistical techniques including analysis of variance and covariance and discriminant function analysis.

## CHAPTER VI

## SUMMARY

Although canonical correlation analysis (CCA) has been available as a statistical tool for more than thirty years, relatively little use of it is found in the psychological literature, and there seems to be no instance in which it has been used for the joint purpose of prediction and exploration of relationships among variables. CCA is like multiple regression analysis (MRA) in its predictive function, and it is like factor analysis (FA) in that it can be used for the conceptual restructuring of data. Where attempts have been made in the literature to give FA-type interpretations to the results of a CCA, however, these have almost always made use of canonical weights, rather than loadings which may be derived from the analysis. The inappropriateness of weights for interpretation, and the possibility of obtaining loadings from a CCA, seem not to be generally appreciated. Mathematical treatments of CCA have not focused explicitly on problems of interpretation, and further have lacked uniformity of notation and terminology. Methodological studies have indicated relationships between CCA and MRA, or between CCA and FA, but there has been no unified presentation of all three, and no studies have exhaustively explored the implications of these relationships.

In this dissertation it is shown how CCA is an extension of MRA to the case of several criterion variables, with a relaxation of the restriction that the predictors be fixed, and also how a canonical correlation coefficient is only a particular kind of multiple correlation. It is thus made clear that the weights obtained from a CCA are partial regression weights. Whereas MRA produces such weights only for the observed predictors, CCA produces weights for both sets of observed variables, and either set may be treated as "predictor" or "criterion." It is made clear that such weights, either in MRA or CCA, are useful in calculating composite scores for each individual in the sample, but that they are not very helpful in speculating about the meaning of these composite scores. The implications of using standardized or non-standardized composite scores is discussed, and matrix equations for standardizing them presented.

It is shown how factor loadings, which <u>are</u> appropriate for interpretation of the composite scores obtained from a CCA, may be computed. Two kinds of loadings--loadings within sets which link CCA to principal components FA, and loadings between sets which link it to MRA--are described and the relation between them given. Both MRA and CCA are applied to the same artificial data problem, which is small enough so that it is possible to see what happens to each variable and to each subject in the sample. The similarities and differences between MRA and CCA are illustrated with this problem, and it is also used to show how CCA can reproduce the factor structure known to underly two sets of variables. All computational steps for a complete CCA are shown, and reference is made to an available computer program which calculates loadings as well as weights. CCA is then applied to empirical data to show that it is capable, through the use of loadings, of producing meaningful factors in a situation where the structure is not known beforehand.

An index of overall association between two batteries of measures, which is a function of the number and size of the significant canonical correlations and of factor structure and saturation, is developed. A method for interpreting factors, whether from a CCA or an FA, in terms of the percentage of sum of squares contributed by each variable to each factor, rather than in terms of loadings as such, is demonstrated. Alternative cross-validation methods for CCA are described, as are certain pragmatic limitations on the number of variables that may be analyzed even with electronic computers. The question of whether CCA and FA are useful primarily as a means of reducing a larger number of variables to a smaller number of factors, or primarily as a means of transforming sets of correlated variables into new sets of orthogonal variables, is discussed. In any case, CCA would seem to be the method of choice for matching factors from different sets of variables.

Algebraic proofs and a summary matrix notation of all important definitions and relationships are given separately for MRA and CCA in the appendices.

## REFERENCES


Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Bartlett, M. S. Internal and external factor analysis. British Journal of Psychology, Statistical Section, 1947, 1, 73-81.

Browne, E. T. Introduction to the theory of determinants and matrices. Chapel Hill: University of North Carolina, 1958.

Burt, Cyril. Factor analysis and canonical correlations. British Journal of Psychology, Statistical Section, 1947, 1, 96-106.

Cattrell, R. B. Factor analysis. New York: Harper, 1952.

Clark, M. L., Ray, T. S., Paredes, A., Ragland, R. E., Costiloe, J. P., Smith, C. W., and Wolf, S. Chlorpromazine in chronic schizophrenic women: The effect on cholesterol levels and cholesterol-behavior relationships. Psychosomatic Medicine, 1967, in press.

Cooley, W. W., and Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Das, R. S. An application of factor and canonical analysis to multivariate data. British Journal of Mathematical and Statistical Psychology, 1965, 18, 57-69.

Fruchter, Benjamin. Introduction to factor analysis. New York: Van Nostrand, 1954.

Graybill, F. A. An introduction to linear statistical models. New York: McGraw-Hill, 1961.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Horst, Paul. Relations among m sets of measures. Psychometrika, 1961, 26, 129-149.

Hotelling, Harold. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 1933, 24, 417-441, 498-520.

Hotelling, Harold. The most predictable criterion. <u>Journal of Educational Psychology</u>, 1935, <u>26</u>, 139-142.

Hotelling, Harold. Relations between two sets of variates. <u>Biometrika</u>, 1936, <u>28</u>, 321-377.

Hotelling, Harold. The relations of the newer multivariate statistical methods to factor analysis. <u>British Journal of Statistical Psychology</u>, 1957, <u>10</u>, 69-79.

Johnson, P. O., and Jackson, R. W. B. <u>Modern statistical methods: Descriptive and inductive</u>. Chicago: Rand McNally, 1959.

Jones, Robert. A canonical analysis of criterion and predictor factors of talented behavior. Unpublished Doctor's Dissertation, Department of Psychology, University of Texas, 1964.

King, F. J., Bowman, Barbara, and Moreland, H. J. Some intellectual correlates of biochemical variability. <u>Behavioral Science</u>, 1961, <u>6</u>, 297-302.

Koons, P. B., jr. Canonical analysis. In Borko, Harold (Ed.), <u>Computer applications in the behavioral sciences</u>. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

McKeon, J. J. Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory. <u>Psychometric Monographs</u>, No. 13 (undated). Mimeographed paper, Department of Psychology, University of Michigan, 1962.

Meredith, William. Canonical correlations with fallible data. <u>Psychometrika</u>, 1964, <u>29</u>, 55-65.

Mitchell, Cara. Canonical correlation program, one-pass version. <u>1620 Users Group Library</u>, 1965.

Mukherjee, B. N. Application of canonical correlation analysis to learning data. <u>Psychological Bulletin</u>, 1966, <u>66</u>, 9-21.

Ragland, R. E., and Costiloe, J. P. On the analysis of repeated measurements on the same subjects. <u>Journal of Chronic Diseases</u>, 1963, <u>16</u>, 1215-1216.

Rao, C. R. Estimation and tests of significance in factor analysis. <u>Psychometrika</u>, 1955, <u>20</u>, 93-111.

Seal, Hilary. <u>Multivariate statistical analysis for biologists</u>. New York: Wiley, 1964.

Thomson, Godfrey. The maximum correlation of two weighted batteries. British Journal of Psychology, Statistical Section, 1947, 1, 27-34.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Ward, J. H., jr. Comments on "The paramorphic representation of clinical judgment." Psychological Bulletin, 1962, 59, 74-76.

Wittenborn, J. R. Pretreatment individual differences as potential predictors of response to pharmacology. Journal of Nervous and Mental Disease, 1964, 139, 186-194.

## APPENDIX I

### MULTIPLE REGRESSION ANALYSIS: DEFINITIONS, SUMMARY OF ALGEBRAIC RELATIONSHIPS, AND PROOFS

The definitions and equations which follow are given for the general case of several criteria predicted separately by the same set of predictors. To apply them to the case of only one criterion, replace $\underline{R}_{12}$ and $\underline{R}_{21}$ with $\underline{r}_{12}$ and $\underline{r}_{21}$ in what follows.

#### Definitions

$N$ = number of subjects

$n = N - 1$ = degrees of freedom

$m_1$ = number of variables in the predictor set

$m_2$ = number of variables in the criterion set

$x_j$ = an observed score in the predictor set, in standard form

$y_j$ = an observed score in the criterion set, in standard form

$\underline{X}$ = N by $m_1$ matrix of standard scores of all subjects on all x variables

$\underline{Y}$ = N by $m_2$ matrix of standard scores of all subjects on all y variables

$\underline{R}_{11}$ = $m_1$ by $m_1$ matrix of correlations within the x set

$\underline{R}_{12}$ = $m_1$ by $m_2$ matrix of correlations of the variables in the x set with the variables in the y set

$\underline{R}_{21}$ = $m_2$ by $m_1$ matrix which is the transpose of $\underline{R}_{12}$

$\underline{R}_{22}$ = $m_2$ by $m_2$ matrix of correlations within the y set

## Relationships and Proofs

The matrix $\underline{B}$ of standard partial regression weights is given by:

(1) $$\underline{B} = \underline{R}_{11}^{-1} \underline{R}_{12}.$$

The squared multiple correlations (mult $R^2$'s) are found on the diagonal of $\underline{R}_{2.1}$, where

(2) $$\underline{R}_{2.1} = \underline{R}_{21} \underline{R}_{11}^{-1} \underline{R}_{12}.$$

Proof 1

When the observed predictors (x's) are uncorrelated, the regression weights are also correlation coefficients. That is, if $\underline{R}_{11} = \underline{I}$, then $\underline{R}_{11}^{-1} = \underline{I}$. Then from equation (1), $\underline{B} = \underline{R}_{11}^{-1} \underline{R}_{12} = \underline{I}\ \underline{R}_{12} = \underline{R}_{12}$. Thus in this case the weights are identical with the simple correlations between the x's and the y's.

Q. E. D.

Proof 2

The matrix of unstandardized composite predictor scores is $p* = \underline{X}\ \underline{B}$. It can be shown that the variance-covariance matrix for p* equals $\underline{R}_{2.1}$, which means that the variance of each unstandardized composite predictor equals the squared multiple correlation between it and its criterion. The variance-covariance matrix is:

(3a) $$(1/n)\ \underline{P}^{*\prime}\ \underline{P}^{*} = (1/n)\ (\underline{X}\ \underline{B})'(\underline{X}\ \underline{B}),$$

(3b) $$= (1/n)\ \underline{B}'\ \underline{X}'\ \underline{X}\ \underline{B},$$

(3c) $$= \underline{B}'\ \underline{R}_{11}\ \underline{B},$$

(3d) $$= (\underline{R}_{11}^{-1}\ \underline{R}_{12})'\ \underline{R}_{11}\ (\underline{R}_{11}^{-1}\ \underline{R}_{12}),$$

(3e) $$= \underline{R}_{21}\ \underline{R}_{11}^{-1}\ \underline{R}_{11}\ \underline{R}_{11}^{-1}\ \underline{R}_{12},\ \text{where}\ (\underline{R}_{11}^{-1})' = \underline{R}_{11}^{-1},$$

(3f) $$= \underline{R}_{21}\ \underline{R}_{11}^{-1}\ \underline{R}_{12},$$

(3g) $$= \underline{R}_{2.1}.$$

Q. E. D.

It follows that the composite predictors may be standardized by dividing each column of $\underline{B}$ through by the corresponding mult R, to give a new matrix of weights, $\underline{V}$:

(4) $$\underline{V} = \underline{B}\ \underline{Z},\ \text{where}\ \underline{Z} = \begin{bmatrix} (\text{mult } R_1)^{-1} & 0 & \cdots & 0 \\ 0 & (\text{mult } R_2)^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & (\text{mult } R_{m_2})^{-1} \end{bmatrix}.$$

(5) $$\text{Then}\ \underline{P} = \underline{X}\ \underline{V},$$

where the diagonal elements of $(1/n)\ \underline{P}'\ \underline{P}$ are ones.

To intercorrelate the composite predictors:

(6a) $$\underline{R}_{pp} = (1/n)\ \underline{P}'\ \underline{P},$$

(6b) $$= (1/n)\ \underline{V}'\ \underline{X}'\ \underline{X}\ \underline{V},$$

(6c) $$= \underline{V}' \, \underline{R}_{11} \, \underline{V}, \text{ where } \underline{V} = \underline{B} \, \underline{Z},$$

(6d) $$= \underline{Z}' \, \underline{B}' \, \underline{R}_{11} \, \underline{B} \, \underline{Z}, \text{ where } \underline{Z}' = \underline{Z}.$$

But from the second proof above, $\underline{B}' \, \underline{R}_{11} \, \underline{B} = \underline{R}_{2.1}$, hence,

(6e) $$\underline{R}_{pp} = \underline{Z} \, \underline{R}_{2.1} \, \underline{Z}.$$

To correlate the observed criteria with the composite predictors:

(7a) $$\underline{R}_{yp} = (1/n) \, \underline{Y}'\underline{P},$$

(7b) $$= (1/n) \, \underline{Y}' \, \underline{X} \, \underline{V},$$

(7c) $$= \underline{R}_{21} \, \underline{V}.$$

## APPENDIX II

## CANONICAL CORRELATION ANALYSIS: DEFINITIONS, SUMMARY OF ALGEBRAIC RELATIONSHIPS, AND PROOFS

### Definitions

$N$ = number of subjects

$n = N - 1$ = degrees of freedom

$m_1$ = number of variables in the x set

$m_2$ = number of variables in the y set, where $m_2 \leq m_1$

$x_j$ = an observed score in the x set, in standard form

$y_j$ = an observed score in the y set, in standard form

$\underline{X}$ = N by $m_1$ matrix of standard scores of all subjects on all x variables

$\underline{Y}$ = N by $m_2$ matrix of standard scores of all subjects on all y variables

$\underline{D} = \left[\underline{X}\ \underline{Y}\right]$ = matrix of <u>all</u> the scores (data matrix)

$\underline{R}_{11}$ = $m_1$ by $m_1$ matrix of intercorrelations within the x set

$\underline{R}_{12}$ = $m_1$ by $m_2$ matrix of intercorrelations of the variables in the x set with the variables in the y set

$\underline{R}_{21}$ = $m_2$ by $m_1$ matrix which is the transpose of $\underline{R}_{12}$

$\underline{R}_{22}$ = $m_2$ by $m_2$ matrix of intercorrelations within the y set

$$(1) \qquad (1/n)\ \underline{D}'\underline{D} = \underline{R} = \begin{bmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{21} & \underline{R}_{22} \end{bmatrix}.$$

## The Problem

Find a pair of linear combinations of the observed scores of the form $p_1 = \sum v_{j1} x_j$ and $q_1 = \sum w_{j1} y_j$, with $v_{j1}$ and $w_{j1}$ not all zero, such that $p_1$ and $q_1$ are in standard form and the correlation between them is a maximum. Find this correlation, which is called a canonical correlation.

Then find a second pair of standardized linear combinations, $p_2$ and $q_2$, which are maximally correlated subject to the restriction that they are orthogonal to p and q. Repeat this process until there are $m_2$ correlated pairs of p's and q's.

## First Solution

There are two approaches to a solution. The first finds the squared canonical correlations ($\lambda_j$ values) and the weights for the y's from:

$$(2a) \qquad (\underline{R}_{22}^{-1}\ \underline{R}_{21}\ \underline{R}_{11}^{-1}\ \underline{R}_{12} - \lambda_j\ \underline{I})\underline{b}_j = 0,$$

$$(2b) \qquad \text{where } \lambda_j = (\underline{b}_j'\ \underline{R}_{21}\ \underline{R}_{11}^{-1}\ \underline{R}_{12}\ \underline{b}_j)(\underline{b}_j'\ \underline{R}_{22}\ \underline{b}_j)^{-1},\ j = 1, 2, \ldots, m_2.$$

The vectors $\underline{b}_1, \underline{b}_2, \cdots, \underline{b}_{m_2}$ constitute the $m_2$ by $m_2$ matrix $\underline{B}$ of weights to be applied to the y set. The weights for the x set are found by:

$$(3) \qquad \underline{A} = (\underline{R}_{11}^{-1}\ \underline{R}_{12}\ \underline{B})(\underline{B}'\ \underline{R}_{22}\ \underline{B})^{-\frac{1}{2}},$$

where $\underline{A}$ is an $m_1$ by $m_2$ matrix and $(\underline{B}'\ \underline{R}_{22}\ \underline{B})$ is a diagonal matrix. The $\underline{A}$ and $\underline{B}$ weights, if applied to $\underline{X}$ and $\underline{Y}$, will not produce standardized composites. Rather, the composites will have zero means and variances equal to $\underline{A}'\ \underline{R}_{11}\ \underline{A}$ and $\underline{B}'\ \underline{R}_{22}\ \underline{B}$ respectively (these are diagonal matrices with the variances on the diagonal). If

$$(4) \qquad \underline{V} = \underline{A}\ (\underline{A}'\ \underline{R}_{11}\ \underline{A})^{-\frac{1}{2}} \text{ and } \underline{W} = \underline{B}\ (\underline{B}'\ \underline{R}_{22}\ \underline{B})^{-\frac{1}{2}},$$

$$(5) \qquad \text{then } \underline{P} = \underline{X}\ \underline{V} \qquad \text{and } \underline{Q} = \underline{Y}\ \underline{W}$$

are N by $m_2$ matrices of standardized composites, and

$$(6a) \qquad (1/n)\ \underline{P}'\underline{P} = (1/n)\ \underline{Q}'\underline{Q} = \underline{I}, \text{ and}$$

$$(6b) \qquad (1/n)\ \underline{P}'\underline{Q} = (1/n)\ \underline{Q}'\underline{P} = \underline{\Lambda}^{\frac{1}{2}}, \text{ where}$$

$$(6c) \qquad \underline{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \lambda_2^{\frac{1}{2}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_{m_2}^{\frac{1}{2}} \end{bmatrix}.$$

In the above solution, $\underline{V}$ is rectangular ($m_1$ by $m_2$) while $\underline{W}$ is square ($m_2$ by $m_2$), and $\underline{P}$ and $\underline{Q}$ have the same number of columns ($m_2$). Each column vector of composite scores in $\underline{P}$ has a matching column vector in $\underline{Q}$ which correlates with it and with no other column vector in $\underline{P}$ or $\underline{Q}$.

## Second Solution

The alternative approach is to find the $\lambda_j$'s and $\underline{A}$ from:

(7a) $$(\underline{R}_{11}^{-1}\,\underline{R}_{12}\,\underline{R}_{22}^{-1}\,\underline{R}_{21} - \lambda_j\,\underline{I})\underline{a}_j = \underline{0}$$

(7b) where $\lambda_j = (\underline{a}_j'\,\underline{R}_{12}\,\underline{R}_{22}^{-1}\,\underline{R}_{21}\,\underline{a}_j)(\underline{a}_j'\,\underline{R}_{11}\,\underline{a}_j)^{-1}$, $j = 1, 2, \cdots, m_1$.

Note that in (7b) the index j ranges from 1 to $m_1$, whereas in (2b) it ranged only from 1 to $m_2$. If $m_1 > m_2$, (7a) and (7b) will yield $m_2$ nonzero values of $\lambda_j$ plus $m_1 - m_2$ values which are identically zero. For example,

(8) If $m_1 = 3$ and $m_2 = 2$, $$\underline{\lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix of $\underline{A}$ weights from (7a) will be square ($m_1$ by $m_1$). The matrix of $\underline{B}$ weights is found by:

(9) $$\underline{B} = \underline{R}_{22}^{-1}\,\underline{R}_{21}\,\underline{A}\,(\underline{A}'\,\underline{R}_{11}\,\underline{A})^{-\frac{1}{2}}.$$

$\underline{B}$ will be a $m_2$ by $m_1$ matrix, but the last $m_1 - m_2$ columns will contain zeros, and can be dropped, leaving B a square ($m_2$ by $m_2$) matrix.

$\underline{V}$ and $\underline{W}$ are found from $\underline{A}$ and $\underline{B}$ as before, only now both $\underline{V}$ and $\underline{W}$ are square, being $m_1$ by $m_1$ by $m_2$ by $m_2$, respectively. $\underline{P}$ now has $m_1 - m_2$ more columns than $\underline{Q}$. For example, if $m_1 = 3$ and $m_2 = 2$, (10a) $$\underline{V} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \underline{v}_3 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix},$$

(10b) $$\underline{W} = \begin{bmatrix} \underline{w}_1 & \underline{w}_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}, \text{ and}$$

(10c) $$\underline{P} = \begin{bmatrix} \underline{p}_1 & \underline{p}_2 & \underline{p}_3 \end{bmatrix}, \quad \underline{Q} = \begin{bmatrix} \underline{q}_1 & \underline{q}_2 \end{bmatrix}.$$

The set of N composite scores symbolized by $\underline{p}_3$ does not correlate with any other vector in $\underline{P}$ or with any vector in $\underline{Q}$.

## Proofs

In the proofs that follow, $\underline{V}$ refers to the $m_1$ by $m_1$ square matrix of weights which, when applied to the matrix of x scores, gives $\underline{P} = \underline{X}\,\underline{V}$, and $\underline{P}$ will refer to the full N by $m_1$ matrix of factor scores. The reasons for so defining $\underline{V}$ and $\underline{P}$ will become apparent.

The basic equations obtained from the CCA are:

(11a) $$\underline{P} = \underline{X}\,\underline{V} \qquad\qquad \underline{Q} = \underline{Y}\,\underline{W}$$

(11b) $$(1/n)\ \underline{P}'\underline{P} = \underline{I}_1 \qquad\qquad (1/n)\ \underline{Q}'\underline{Q} = \underline{I}_2,$$

where $\underline{I}_1$ is an $m_1$ by $m_1$ matrix and $\underline{I}_2$ is an $m_2$ by $m_2$ matrix;

(11c) $$(1/n)\ \underline{P}'\underline{Q} = \underline{\Lambda}_1^{\frac{1}{2}}, \qquad\qquad (1/n)\ \underline{Q}'\underline{P} = \underline{\Lambda}_2^{\frac{1}{2}},$$

where $\underline{\Lambda}_1^{\frac{1}{2}}$ is an $m_1$ by $m_1$ matrix and $\underline{\Lambda}_2^{\frac{1}{2}}$ is an $m_2$ by $m_2$ matrix.

It is important to note that $\underline{\Lambda}_1^{\frac{1}{2}}$ and $\underline{\Lambda}_2^{\frac{1}{2}}$ are diagonal matrices having the same numerical values in the corresponding diagonal elements, except that when $m_1 > m_2$, $\underline{\Lambda}_1^{\frac{1}{2}}$ will have $m_1 - m_2$ more rows and columns than $\underline{\Lambda}_2^{\frac{1}{2}}$, and all the elements of these extra rows and columns, including the diagonal elements, will be zero.

Substituting (11a) into (11c) gives:

$$
\begin{array}{llll}
\text{(11d)} & \underline{\Lambda}_1^{\frac{1}{2}} = (1/n)\ \underline{P}'\underline{Q}, & & \underline{\Lambda}_2^{\frac{1}{2}} = (1/n)\ \underline{Q}'\underline{P}, \\
\text{(11e)} & \quad = (1/n)(\underline{X}\ \underline{V})'(\underline{Y}\ \underline{W}), & & \quad = (1/n)(\underline{Y}\ \underline{W})'(\underline{X}\ \underline{V}), \\
\text{(11f)} & \quad = (1/n)\ \underline{V}'\underline{X}'\underline{Y}\ \underline{W}, & & \quad = (1/n)\ \underline{W}'\underline{Y}'\underline{X}\ \underline{V}, \\
\text{(11g)} & \quad = \underline{V}'\underline{R}_{12}\ \underline{W}. & & \quad = \underline{W}'\underline{R}_{21}\ \underline{V}.
\end{array}
$$

The loadings within sets, which are correlations of x's with p's and y's with q's are:

$$
\begin{array}{llll}
\text{(12a)} & \underline{F} = (1/n)\ \underline{X}'\ \underline{P} & & \underline{G} = (1/n)\ \underline{Y}'\ \underline{Q} \\
\text{(12b)} & \quad = (1/n)\ \underline{X}'\ \underline{X}\ \underline{V} & & \quad = (1/n)\ \underline{Y}'\ \underline{Y}\ \underline{W} \\
\text{(12c)} & \quad = \underline{R}_{11}\ \underline{V} & & \quad = \underline{R}_{22}\ \underline{W}
\end{array}
$$

## Proof 1

From (12c) it follows immediately that when the observed x and y variables are orthogonal within sets (that is $\underline{R}_{11} = \underline{I}$, $\underline{R}_{22} = \underline{I}$), the loadings within sets are identical with the canonical weights:

(12d) If $\underline{R}_{11} = \underline{I}$, If $\underline{R}_{22} = \underline{I}$,

(12e) $\underline{F} = \underline{I}\ \underline{V} = \underline{V}$. $\underline{G} = \underline{I}\ \underline{W} = \underline{W}$.

Q. E. D.

The loadings between sets, which are the correlations of the x's with the q's and the y's with the p's, are found by:

(13a) $\underline{M} = (1/n)\ \underline{X}'\ \underline{Q}$, $\underline{N} = (1/n)\ \underline{Y}'\ \underline{P}$,

(13b) $= (1/n)\ \underline{X}'\ \underline{Y}\ \underline{W}$, $= (1/n)\ \underline{Y}'\ \underline{X}\ \underline{V}$,

(13c) $= \underline{R}_{12}\ \underline{W}$. $= \underline{R}_{21}\ \underline{V}$.

For the proofs that follow, certain relations involving $\underline{V}$ and $\underline{W}$ must be demonstrated. Both $\underline{V}$ and $\underline{W}$ are square matrices which are non-singular since they consist of columns of eigenvectors. Therefore they have inverses. In the rare instance when all the roots of (7a) are identically one, so that it would appear that all the eigenvectors would be equal in $\underline{V}$ and in $\underline{W}$, $\underline{V}$ and $\underline{W}$ can still be made non-singular (Browne, 1958; pp. 89 ff.).

From (11b):

(14a) $(1/n)\ \underline{P}'\underline{P} = \underline{I}_1$, $(1/n)\ \underline{Q}'\ \underline{Q} = \underline{I}_2$,

(14b) $(1/n)\ \underline{V}'\ \underline{X}'\ \underline{X}\ \underline{V} = \underline{I}_1$, $(1/n)\ \underline{W}'\ \underline{Y}'\ \underline{Y}\ \underline{W} = \underline{I}_2$,

(14c) $\underline{V}'\ \underline{R}_{11}\ \underline{V} = \underline{I}_1$, $\underline{W}'\ \underline{R}_{22}\ \underline{W} = \underline{I}_2$,

The matrix products $(\underline{R}_{11}\ \underline{V})$ and $(\underline{R}_{22}\ \underline{W})$ are square and nonsingular, hence;

(14d) $\underline{V}' = (\underline{R}_{11}\ \underline{V})^{-1}$, $\underline{W}' = (\underline{R}_{22}\ \underline{W})^{-1}$,

$$\text{(14e)} \qquad \underline{V}' = \underline{V}^{-1}\,\underline{R}_{11}^{-1}, \qquad \underline{W}' = \underline{W}^{-1}\,\underline{R}_{11}^{-1},$$

$$\text{(14f)} \qquad \underline{V}\,\underline{V}' = \underline{R}_{11}^{-1}. \qquad \underline{W}\,\underline{W}' = \underline{R}_{22}^{-1}.$$

Proof 2

Using (14f) it can be shown that the fundamental equation of factor analysis holds for the loadings within sets. That is, $\underline{F}\,\underline{F}' = \underline{R}_{11}$, and $\underline{G}\,\underline{G}' = \underline{R}_{22}$:

$$\text{(15a)} \qquad \underline{F}\,\underline{F}' = (\underline{R}_{11}\,\underline{V})(\underline{R}_{11}\,\underline{V})', \qquad \underline{G}\,\underline{G}' = (\underline{R}_{22}\,\underline{W})(\underline{R}_{22}\,\underline{W})',$$

$$\text{(15b)} \qquad = \underline{R}_{11}\,\underline{V}\,\underline{V}'\,\underline{R}_{11}, \qquad = \underline{R}_{22}\,\underline{W}\,\underline{W}'\,\underline{R}_{22},$$

$$\text{(15c)} \qquad = \underline{R}_{11}\,\underline{R}_{11}^{-1}\,\underline{R}_{11}, \qquad = \underline{R}_{22}\,\underline{R}_{22}^{-1}\,\underline{R}_{22},$$

$$\text{(15d)} \qquad = \underline{R}_{11}. \qquad = \underline{R}_{22}.$$

Q. E. D.

Proof 3

Using (11g), (12c), and (14f) it can be shown that the loadings between sets equal the loadings within sets times the associated canonical correlation; that is, $\underline{M} = \underline{F}\,\boldsymbol{\lambda}_1^{\frac{1}{2}}$, and $\underline{N} = \underline{G}\,\boldsymbol{\lambda}_2^{\frac{1}{2}}$. For,

$$\text{(16a)} \qquad \underline{F}\,\boldsymbol{\lambda}_1^{\frac{1}{2}} = (\underline{R}_{11}\,\underline{V})(\underline{V}'\,\underline{R}_{12}\,\underline{W}), \qquad \underline{G}\,\boldsymbol{\lambda}_2^{\frac{1}{2}} = (\underline{R}_{22}\,\underline{W})(\underline{W}'\,\underline{R}_{21}\,\underline{V})$$

$$\text{(16b)} \qquad = \underline{R}_{11}\,(\underline{V}\,\underline{V}')\,\underline{R}_{12}\,\underline{W}, \qquad = \underline{R}_{22}\,(\underline{W}\,\underline{W}')\,\underline{R}_{21}\,\underline{V},$$

$$\text{(16c)} \qquad = \underline{R}_{11}\,\underline{R}_{11}^{-1}\,\underline{R}_{12}\,\underline{W}, \qquad = \underline{R}_{22}\,\underline{R}_{22}^{-1}\,\underline{R}_{21}\,\underline{V},$$

$$\text{(16d)} \qquad = \underline{M}. \qquad = \underline{N}.$$

Q. E. D.

Proof 4

Using (14f) it can be shown that the sum of the squares of the loadings between sets for a given observed variable (x or y) equals the squared multiple correlation between that observed variable and all the observed variables in the other set.

$$\text{(17a)} \qquad \underline{M}\,\underline{M}' = (\underline{R}_{12}\,\underline{W})(\underline{R}_{12}\,\underline{W})', \qquad \underline{N}\,\underline{N}' = (\underline{R}_{21}\,\underline{V})(\underline{R}_{21}\,\underline{V})',$$

$$\text{(17b)} \qquad = \underline{R}_{12}\,\underline{W}\,\underline{W}'\,\underline{R}_{21}, \qquad = \underline{R}_{21}\,\underline{V}\,\underline{V}'\,\underline{R}_{12},$$

$$\text{(17c)} \qquad = \underline{R}_{12}\,\underline{R}_{22}^{-1}\,\underline{R}_{21}, \qquad = \underline{R}_{21}\,\underline{R}_{11}^{-1}\,\underline{R}_{12},$$

$$\text{(17d)} \qquad = \underline{R}_{1.2}. \qquad = \underline{R}_{2.1}.$$

Q. E. D.

The expression for $\underline{N}\,\underline{N}'$ is the more familiar, since the y variables have here been treated as criteria. The diagonal element $d_{jj}$ of $\underline{R}_{2.1}$ is the squared multiple correlation between $y_j$ and all the x's. If the x's are treated as criterion variables, then the diagonal element $d_{jj}$ of $\underline{R}_{1.2}$ is the squared multiple correlation between $x_j$ and all the y's.

The final proof has no immediate application to this dissertation, but is included because some application for it may occur to subsequent investigators.

Proof 5

When the y variables are orthogonal to begin with, so that $\underline{R}_{22} = \underline{I}$, then the sum of the squared multiple correlations for predicting each of the y's from the x's (that is, the <u>trace</u> of

$\underline{R}_{2.1}$) equals the sum of the squared canonical correlations. This follows from the fact that for a characteristic equation of the form

(18a) $$(\underline{K} - \lambda_j \underline{I})\underline{b}_j = \underline{0}$$

the sum of the roots of $\underline{K}$ equals the trace of $\underline{K}$. Then for equation (2a),

(18b) $$\sum \lambda_j = \text{tr } (\underline{R}_{22}^{-1} \underline{R}_{21} \underline{R}_{11}^{-1} \underline{R}_{12}).$$

But if $\underline{R}_{22} = \underline{I}$, then $\underline{R}_{22}^{-1} = \underline{I}$, so that (18b) becomes

(18c) $$\sum \lambda_j = \text{tr } (\underline{R}_{21} \underline{R}_{11}^{-1} \underline{R}_{12}) = \text{tr } \underline{R}_{2.1}.$$

Q. E. D.

The foregoing can easily be adapted to the case in which the x variables are orthogonal to begin with. Then $\underline{R}_{11} = \underline{R}_{11}^{-1} = \underline{I}$, and using characteristic equation (7a) instead of (2a):

(18d) $$\sum \lambda_j = \text{tr } (\underline{R}_{12} \underline{R}_{22}^{-1} \underline{R}_{21}) = \text{tr } \underline{R}_{1.2}.$$

## APPENDIX III

## PSYCHOLOGICAL EXAMINER'S RATING SCALE FOR SCHIZOPHRENIC PATIENTS [a,b]

T. S. Ray, Ph.D., and J. S. Chappell, M. S.

1. Ideational fluency.

   +6 Patient chatters incessantly in disorganized way or even shouts unintelligibly.

   +4 Patient insisently presents psychotic ideas to examiner, is so preoccupied with own productions that examiner has difficulty breaking into stream of ideas.

   +2 Patient is overtalkative but interacts with examiner to satisfactory degree.

   0 Patient exchanges ideas spontaneously with examiner without loss of ease.

   -2 Answers questions without spontaneously elaborating. Does not carry her end of the conversation. Waits for examiner to lead conversation.

   -4 Verbally inhibited, replies with minimum answer to direct questions only, requires urging to reply, may speak in monotone or muffled or faint speech.

   -6 Muteness, or absence of meaningful speech (may mumble to self).

2. Expressiveness and gestures.

   +6 Active posturing and/or bizarre behavior.

   +4 Manneristic behavior and/or inappropriate facial expressions.

[a] Reproduced with permission of senior author.
[b] Abridged from original scale.

+2 Histrionic animation of facial expressions, gestures, etc., which are in line with the general ideas being expressed.

0 Spontaneous but modulated use of gestures and facial expressions.

-2 Speech not accompanied by associated gestures and facial expression; reduction in spontaneity, or effortful use of gestures.

-4 Economy of movement, impassive facial expression, etc.

-6 Expressionless, masked facies, waxy flexibility.

3. Affective status.

+6 Patient seems excited, high, or manic (possibly with flight of ideas).

+4 Patient is silly and/or inappropriate with laughter or giggling.

+2 Patient maintains a gay, carefree, or jovial front or may be quite labile (emotional) in affect.

0 Patient is fairly responsive in a well-modulated way, smiling at appropriate times, becoming serious when appropriate, etc.

-2 Patient is quite serious, sober, or subdued, unsmiling, uneasy, self-conscious.

-4 Patient seems unhappy, sad, depressed or fearful, even suspicious, or complaining and peevish; or markedly flattened in affect.

-6 Patient may be profoundly depressed, withdrawn, or uncommunicative.

4. Motor activity.

+6 Patient is grossly overactive, may leave chair to move about room restlessly.

+4 Patient engages in continuous random movements, is motorically active while able to stay seated.

+2 Patient is restless or fidgety, may busy self with something most of the time, fingering things, etc.

0 Motoric activity spontaneous and well-coordinated, absence of undue restlessness or inhibition.

-2 Movements have deliberate quality, slowness, lacking in spontaneity and ease of coordination or appears to be overcontrolled.

-4 Inhibited, jerky movements or stereotyped rocking motion. Extremely effortful, tired appearing, lacking in energy, etc.

-6 Spontaneous movements absent, extreme economy of movements, possibly waxy flexibility or rigidity of musculature, stuporousness.

5. Ease of testing.

+6 Testing attempted but not possible due to overactivity or testing not attempted due to patient being in seclusion.

+4 Testing impaired due to overactivity, some portion of test not valid (note on test protocol).

+2 Testing not seriously impaired even though the patient was overactive.

0 Testing accomplished without difficulty.

-2 Testing not seriously impaired (invalidated) even though patient was underactive or slow.

-4 Testing impaired due to unresponsiveness of patient, some portion of test inadequately responded to (invalidated; note on test protocol).

-6 Testing attempted but not possible due to lack of participation of patient, or testing not attempted due to refusal of patient to come to testing room. (Overactive patients should be scored +6 rather than -6.)