PSYCHOMETRIC MONOGRAPH NUMBER 10

FORMAL MODELS FOR A CENTRAL PREDICTION SYSTEM

The William Byrd Press, Inc.
Richmond, Virginia

FORMAL MODELS FOR A CENTRAL PREDICTION SYSTEM

By

LEDYARD R TUCKER

UNIVERSITY OF ILLINOIS

FORMAL MODELS FOR A CENTRAL PREDICTION SYSTEM

LEDYARD R TUCKER

UNIVERSITY OF ILLINOIS

FOREWORD

Investigation by the author of models for a central prediction system was stimulated by the joint interest of the College Entrance Examination Board and the Educational Testing Service in the possibility of developing a service, using secondary school grades and supplementary data, to aid colleges in the prediction of academic success of applicants. This interest, while existing for many years, was activated by the report of Drs. Benjamin S. Bloom and Frank R. Peters, *The Improvement of Academic Prediction through the Use of a National Grading Scale*, issued in 1959. A committee was established by the Educational Testing Service to investigate the technical feasibility of a central prediction system. In addition to the author, the committee included Drs. Henry S. Dyer, John French, Frederic M. Lord, and William B. Schrader, of the Educational Testing Service, and Dr. Samuel S. Wilks of Princeton University. From a study of the Bloom and Peters report and from the considerations of the committee, the author undertook an investigation of the mathematical structures of possible prediction systems. *Formal Models for a Central Prediction System*, which was first issued as a Research and Development Report by the CEEB while the author was on the staff of the Educational Testing Service, summarizes the results of these investigations.

The author is indebted to a number of individuals on the staff of the Educational Testing Service for their comments and criticisms of the theoretical developments, and for their aid in the conduct of the illustrative study. Contributions by the members of the above named committee were especially helpful. During the computational phases of the study, most excellent assistance was given by Mrs. Ruth Bredon, who carried through the intricate steps on the IBM 650. Grateful acknowledgment is due to the officials of the colleges who supplied the data used in the study.

Financial support for the study and, in part, for the publication of the monograph was provided by the College Entrance Examination Board.

LEDYARD R TUCKER

TABLE OF CONTENTS

LIST OF TABLES AND FIGURE

FORMAL MODELS FOR A CENTRAL PREDICTION SYSTEM

Introduction

For many years, secondary school grades have been used as a basis to form expectations as to the performances of students in colleges. Traditionally, this process has been conducted as a subjective and judgmental operation with, and only relatively recently, some assistance from statistical studies. One serious problem in this field is the existence of extensive differences among various secondary schools and among colleges with respect to their grading standards. This is not the place to review the ways in which the enforcement of some common minimum standard of instruction has been attempted, or the success of these attempts.

Our problem is to recognize the existence of these differences among schools and among colleges, and to develop a system for processing the data of secondary school grades so as to obtain as effective predictions of college performance as feasible. In addition, such a system should possess flexibility so that additional data, such as test scores, could be included so as to arrive at maximally valid predictions. Further, it is desired to provide predictions of performance for a number of colleges. Thus the system should encompass many secondary schools, a number of tests and other predictive variables, and a number of colleges. The present monograph describes several formal models for such a system.

Some work has been done on our problem and related problems over the past 30 years or so. In an early report, Toops (1933) proposed to equate grades at different institutions for students with equal intelligence. This type of technique has been utilized in a number of situations including the equating of grades for different class sections of a course at one institution and, on an informal basis, the judging of levels of schools sending students to a college. An alternative approach attempted by some colleges has been to set up a separate regression equation for each school to predict grades at that college. Burnham (1954) describes some aspects of such a program developed for Yale University. A common experience for colleges that attempt to set up such regression systems is to find that only a few schools provide a large enough sample of students to establish stable regression coefficients. One of the hopes of a more general and, consequently, centralized system using data from a number of colleges is that the number of cases can be built up, and stable coefficients can be determined for a larger number of schools.

The secondary schools share the problem in that they desire to evaluate their grades and also need information with which to guide students in their choice of a college or a field of study. The job that a school would have in following up its graduates is complicated by the scattering of students to numerous colleges, with too few cases at most of those colleges to establish stable relations between school grades and college grades.

A major step toward solving these problems was taken in the study by Bloom and Peters (1961) for the National Council of Independent Schools and the National Registration Office of Independent Schools. In this study a procedure was developed for adjusting both secondary school grades from a group of secondary schools and college grades from a variety of colleges. They termed this procedure the "internal method." The results from this study are exceedingly promising in that high validities obtained for one sample of students held up in subsequent samples.

The major emphasis in this monograph will be on possible complete systems and their properties. These are described (i) to provide an understanding of the interrelations of parts of the systems and (ii) to anticipate features of each system that will be important when it is applied to data and placed in operation. Furthermore, a comprehensive analysis of the complete systems should lead to the discovery of their more subtle aspects and consequences. An investigation of the formal structure of the models underlying the systems may provide a base for further developments.

Consider the following two equations,

$$(1) \qquad (\mathrm{SG})_i w_s + u_s = (\mathrm{ASG})_i , \qquad \text{for } i \text{ from } s,$$

$$(2) \qquad (\mathrm{CG})_i w_c + u_c = (\mathrm{ACG})_i , \qquad \text{for } i \text{ at } c,$$

where $(\mathrm{SG})_i$ is an individual's secondary school grade in terms of the grading scale of his school s, and $(\mathrm{ASG})_i$ is the school grade on an adjusted scale, w_s being a scaling factor and u_s an additive constant for school s grades. Similarly, $(\mathrm{CG})_i$ is an individual's college grade on the grade scale for his college c, and $(\mathrm{ACG})_i$ is the college grade on an adjusted scale, w_c being a scaling factor and u_c an additive constant for college c. The adjusted grade scales are to be established in such a way that they apply across schools or colleges and are not particularized by individual institutions. Differences between institutions are to be absorbed in the values of the scaling factors w and additive constants u.

A number of past suggestions have dealt, separately, with one or the other of these equations. Toops' proposal was to establish the values of the w's and u's in terms of the relations of the grades at each institution with scores made by these students on an intelligence test. The regression approach used at individual colleges involves the use of (1) for the separate schools, the w's and u's being regression coefficients for predicting grades at the college from grades at the schools. In this formulation, the second equation is not utilized. A parallel system could be employed by an individual secondary school to develop information on the validity of its grading system by a backward prediction of the grades at that school from the grades its students earned in the various colleges they attended. This type of system employs (2) to obtain adjusted college grades with which the secondary school grades are correlated.

The "internal method" of Bloom and Peters (1961) used both equations, alternating between them in a successive approximation scheme. Each trial started with conversion of the college grades by (2), using guessed weights, to a temporary adjusted college grade scale. Then regression weights were developed for predicting these adjusted college grades from the various secondary school grades, this yielding the w's and u's in (1) for the schools and adjusted secondary school grades. Finally, the regression equations were determined for predicting the adjusted school grades from the college grades which yielded new w's and u's for the colleges. In their application of this system to data from the files of the National Registration Office, Bloom and Peters carried through three trials.

With the method formulated in the foregoing manner, difficulties exist in the development of properties of the system as a whole and in anticipating the nature of the solution when carried to convergence. One possibility is that all w's, for schools and colleges, will iterate to zero and all u's, for schools and colleges, will iterate to a single constant. Thus, all adjusted grades would be the same constant, that of the u's. Evidence for this possibility is given by the decrease in variances of the adjusted grades from trial to trial in the study by Bloom and Peters. With a slight modification, that of rescaling the adjusted college grades at the beginning of each trial to some fixed variance, this method would yield the first total covariance and variance canonical solution to be discussed in a subsequent section.

The quantitative models to be considered here involve the simultaneous treatment of systems of equations (1) and (2), or, in particular, equations derived from them. Equations (1) and (2) may be termed the separate variable formulation of the problem. A construction will be employed to facilitate the development of these models which encompass the aggregate of separate variable equations for the various schools and colleges. Included in this construction are "scattered data matrices" and associated transformation matrices which will be defined and discussed in the section on mathematical constructions. An isomorphism exists between the system of separate variable equations and the operations possible with the scattered data matrices. Consideration of the quantitative models in terms of the scattered data matrices offers several advantages. Considerable knowledge concerning the properties of any particular model may be derived on deductive bases and then checked with application to actual data. When several models have been explored in this fashion, it will be possible to choose the one which most nearly meets the requirements of each practical situation. Alternatively, the results from the study of these models may indicate important aspects of the practical situations which were not otherwise anticipated. As a result, the statements of the practical problems may be reformulated.

Plan of an Illustrative Study

A number of the developments described in later sections will be illustrated with results from a small study involving data on a total of 387 students in six college groups who graduated from 19 secondary schools. All of these students entered college in the fall of 1956. Following is a short description of the six college groups (hereafter referred to as "colleges").

1. Men at a liberal arts college
2. Students at a university in an urban environment
3. Men at a technical college
4. Men at a technical college
5. Men entering a university in an urban environment who indicated a desire to major in the college of liberal arts
6. Men entering the same university as no. 5 but who indicated a desire to major in the college of engineering

None of these colleges is state or municipally supported, and all are located in the northeast United States. Note that students in colleges nos. 5 and 6 are, in fact, registered at one university and are differentiated only by their statements as to intended majors as upperclassmen.

Table 1 presents, for the students in the study, the number from each school who attended each college, N_{sc}, the total number from each school, $N_{s.}$, the total number at each college, $N_{.c}$, and the total number in the study $N_{..}$. Equations (3)–(5) give the relations among these counts.

$$N_{s.} = \sum_c N_{sc} , \tag{3}$$

$$N_{.c} = \sum_s N_{sc} , \tag{4}$$

$$N_{..} = \sum_s \sum_c N_{sc} = \sum_s N_{s.} = \sum_c N_{.c} . \tag{5}$$

The data in Table 1 are fairly representative of the larger supply of data from which the sample was drawn. Data for the study were obtained from validity study files at Educational Testing Service. The six colleges were selected to provide as complete data as possible on a diverse sample of liberal arts and technical colleges from the northeastern section of the United States. All secondary schools considered had at least four graduates at one or another of the six colleges. Final selection of the schools involved an attempt to have the schools utilized represent the full pool of schools which satisfied the preceding requirement.

TABLE 1
Number of Students in Study

	College						
School	1	2	3	4	5	6	Total
1	4		2				6
2	1		10	4	2	3	20
3		2	16	27	3	10	58
4	2	1	3	4	4	4	18
5	3	1		6	1		11
6	2	2		2	5		11
7	4		9	17	10	8	48
8	3		3	11	1	1	19
9	2		6	2	9	5	24
10	3		8	3	8	2	24
11		14	1		1	1	17
12		10		2			12
13		22	3	1			26
14		9		1		2	12
15	4			1	2	1	8
16	1		13		1	1	16
17	5		4	2	2	2	15
18	22		1	2		1	26
19	10		5		1		16
Total	66	61	84	85	50	41	387

A point of some interest exhibited in Table 1 is the varying extent to which the colleges draw their students from the same secondary schools. Colleges 3 and 4 draw relatively heavily from the same schools. College 2, in contrast, draws the majority of its students in this study from schools 11–14 which, in turn, send the majority of their students in the study to college 2. This varying degree of overlap of students between the schools and colleges in the study has a considerable effect on the results for some of the models for a central prediction system.

Data for each student included his rank in secondary school class and his freshman average grade in college. For the students from each school, their ranks in secondary school class were converted to percentile ranks and then to corresponding deviates for a normal frequency distribution with zero mean and unit standard deviation. The average freshman grades for the students in the study at each college were converted linearly to standard scores (zero mean and unit standard deviation). For approximately 90 per cent of the students at each college, Verbal and Mathematical scores on the CEEB Scholastic Aptitude Test were entered into the analyses. Test scores were not available for some students and test scores were discarded for a random sampling of other students so as to reach the aforementioned level of approximately 90 per cent. Our purpose was to illustrate the effects of missing test data and the manner in which the models allowed for these missing data.

Mathematical Constructions

Before consideration is given to particular models for a central prediction system, several mathematical constructions, which will form a general basis for these models, will be discussed. A more precise notation will be adopted than that employed in (1) and (2); special matrices will be defined, including scattered data matrices and associated transformation matrices.

Indexing Subscripts

A number of indexing subscripts will be used as follows:

$$
\begin{aligned}
s &= 1, 2, \cdots, n_s &&: \text{secondary school;}\\
c &= 1, 2, \cdots, n_c &&: \text{college;}\\
t &= 1, 2, \cdots, n_t &&: \text{test;}\\
i &= 1, 2, \cdots, N_{..} &&: \text{individual student;}\\
j \text{ or } k &= 1, 2, \cdots, n_j &&: \text{predictor variable;}\\
g \text{ or } h &= 1, 2, \cdots, n_g &&: \text{criterion variable;}\\
p \text{ or } q &= 1, 2, \cdots, n_p &&: \text{array of composite scores.}
\end{aligned}
$$

The meanings of the first four of these indices are obvious; they indicate the basic elements to be considered in a central prediction system. A special point concerning the index i is that a unique value of i will be assigned to each student irrespective of the secondary school or college he attended. At times it will be convenient to indicate the school or college, or both, that he attended by use of s and c as subscripts to the index i. Thus, i_s is to be interpreted as extending over all individuals who attended school s; i_c is to be interpreted as extending over all individuals who are attending college c; and i_{sc} is to be interpreted as extending over all individuals who attended school s and are attending college c.

The last three indices will be used as generalized indicators over fields of variables that will be constructed. A possibility exists that the school grades might be combined into several categories rather than into a single average grade or rank in class. Different uses might be made of average English grades from uses made of average mathematics and science grades. Each category of grades from each school will be assigned a unique value of j. In the present formulation it is not necessary that the grades of each school to be grouped into the same categories as the grades of every other school. Values of j will be assigned to whatever categories of grades exist at each school. For convenience in some of the equations to follow, the subscript s will be attached to j to indicate consideration of only those values of j associated with school s. Thus, j_s is to be interpreted as extending over all categories of grades j at school s.

The index j will be used for scores on tests, a unique value of j being assigned to each category of scores on each test entered into the system as a predictor variable. The values of j for a particular test will be indicated by j_t in several subsequent equations.

The index g will be used for indicating grades at the colleges just as the index j is used for indicating grades at the schools. As in the case of the schools, it may be of value to consider several categories of grades at each college. A unique value of g will be assigned to the grades at each college, and g_c is to be interpreted as extending over the g for college c. These variables are the criteria to be predicted by the system.

The index p represents a further expansion in the considerations of a central prediction system. When particular values are given to the weights and additive constants in equations (1) and (2), an array of adjusted school grades and an array of adjusted college grades may be determined. If the weights and additive constants are changed, the adjusted grades will be changed. Each collection of particular values of the weights and additive constants and associated arrays of adjusted grades will be designated by the index p. As will be discussed in detail later, it is possible to consider several of these collections simultaneously. On intuitive grounds, it may seem reasonable that the grades for each school may be differently adjusted for best relations to grades at liberal arts colleges than for best relations to grades at technical colleges. If this be the case, there would be two optimal solutions for the school weights and additive constants. Each of the models to be considered can encompass several collections of weights, etc., simultaneously.

A change in terminology is indicated, also, in the definition of the index p. Instead of "adjusted grade," the term "composite score" will be used to designate the weighted sum of the variables. "Predictor composite score" will designate a weighted sum of the predictor variables, and "criterion composite score" will designate a weighted sum of the criterion variables.

Predictor Variable Constructions

Initial consideration will be given to each school, separately. Subsequently, a system to encompass all schools will be described. For simplicity, the construction to include test scores in the system will be delayed until later in this section.

Predictor variables will be indicated by the lower case letter x; thus, $x_{i_s j_s}$ will designate the secondary school grade of individual i_s in grade category j_s, both being for school s. This notation replaces the $(SG)_i$ in (1). The scaling weights w_s of (1) will be replaced by $a_{j_s p}$, which is the weight applied to category of grades j_s for array p of composite scores. The first term in (1), $(SG)_i w_s$, will be written in the present notation as $\sum_{j:s} x_{i_s j_s} a_{j_s p}$ in which $\sum_{j:s}$ is to be read "the sum of all elements having j for a given s." The index under the summation symbol in front of the colon indicates the variable index for the summation while the index following the colon indicates a limitation on the variable index. This notation permits the consideration of several categories of grades at each school and consideration of several composite scores.

The additive constant u_s of (1) will be designated in the present notation by $a_{J_s p}$. Note the use of the capital letter J which is to have for each school a single, unique value of the index j. Initially, let j_s include every category of grades at school s but not include J_s. A dummy variable $x_{i_s J_s}$ is defined to have a value of unity for each individual from school s. Equation (1) can be rewritten in the present notation as (6).

$$\sum_{j:s} x_{i_s j_s} a_{j_s p} + x_{i_s J_s} a_{J_s p} = f_{i_s p} \qquad (j_s \neq J_s), \tag{6}$$

where $f_{i_s p}$ is the predictor composite score for individual i_s in array p. The dummy variable $x_{i_s J_s}$ was introduced so that the additive constant for the school could be included within the summation by dropping the restriction on j_s and letting j_s include every category of grades at school s and J_s. Then

$$\sum_{j:s} x_{i_s j_s} a_{j_s p} = f_{i_s p} . \tag{7}$$

By defining matrix X_s to contain the $x_{i_s j_s}$, matrix A_s to contain the $a_{j_s p}$, and matrix F_s to contain the $f_{i_s p}$, (7) may be stated in matrix form.

$$X_s A_s = F_s . \tag{8}$$

Note that there are matrices X_s, A_s, and F_s for each school and that (8) applies to each school.

In order to consider all schools simultaneously, the following matrices are defined. These matrices will be called the *scattered data matrices.*

$$X = \begin{bmatrix} X_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & X_s & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & X_{n_s} \end{bmatrix} . \tag{9}$$

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_s \\ \vdots \\ A_{n_s} \end{bmatrix} . \tag{10}$$

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_s \\ \vdots \\ F_{n_s} \end{bmatrix} . \tag{11}$$

The scattered data matrix equation corresponding to (8) is

$$XA = F. \tag{12}$$

Note, for any given school s, that all matrix products are zero except for the product X_sA_s which corresponds to (8). Thus, for any collection of weights whatsoever in the A_s matrices for the various schools, (12) will yield the corresponding predictor composite scores. The importance of this construction is that it brings together all of the separate equations for the various schools into a single equation and provides a means for studying the properties of the system as a whole.

Up to this point test scores have not been included in the constructions. The matrices X and A may be extended to include sections for test scores and their weights. Let x_{ij_t} be the score of individual i on category j_t of test scores. If individual i took the test, x_{ij_t} will be his observed score; if individual i did not take the test, x_{ij_t} will be set at zero. This construction parallels the insertion of zero sections in the scattered data matrix X in (9). Let x_{iJ_t} be a dummy variable for test t with a value of unity for each individual who took the test and a value of zero for each individual who did not take the test. Let the weights for these variables be a_{j_tp} and a_{J_tp} . A contribution of the test scores to predictor composite score p may be written as

$$\sum_{j:t} x_{ij_t}a_{j_tp} + x_{iJ_t}a_{J_tp}$$

when j_t does not include J_t . For individuals who did not take the test this contribution is zero since all their x's are set at zero. Note that taking the test might not always lead to a positive contribution to an individual's composite score. The weights may be negative, especially $a_{J_t p}$. If $a_{J_t p}$ is negative, an individual has to make high enough scores, x_{ij_t} , so that the weighted sum of these scores overcomes the effect of the negative $a_{J_t p}$ in order to obtain a positive contribution to his composite score.

If j_t is permitted to include J_t , the contribution to the composite score can be stated more simply by

$$\sum_{j:t} x_{ij_t} a_{j_t p} .$$

This contribution can be incorporated into the scattered data formulation by adding columns to matrix X for the x_{ij_t} and adding rows to the matrix A for the $a_{j_t p}$. Equation (12) will apply to the extended matrices, and the matrix F will contain the composite scores which are the additive results of the contributions from school grades and from the test scores. Solutions for optimal weights in the matrix A will provide, consequently, optimal relative contributions to the composite scores from the school grades and from the test scores.

A point to be noted is that after the scattered data matrices X and F have been established, the rows of these matrices may be rearranged so that the individuals are in any desirable order. Such a rearrangement is necessary, in fact, so that the individuals may have the same order in the matrices for predictor variables and the matrices for criterion variables. This rearrangement of rows of A and F has no effect on the accuracy of (12).

Criterion Variable Constructions

The criterion variable constructions parallel closely those for the predictor variables and will not be presented in detail. As for the secondary schools, several categories of grades at each college may be established and entered into the system. These grades at college c for individuals attending that college will be designated by $y_{i_c g_c}$. The scaling weights for these grades will be designated by $b_{g_c p}$.

A dummy variable will be established for each college to provide for the additive constant for the college. Scores $y_{i_c G_c}$ will be unity for all individuals attending college c. The additive constant will be handled by a weight $b_{G_c p}$ applied to the dummy variable. Equation (13), which parallels (6), yields the criterion composite scores $z_{i_c p}$ for individuals attending college c.

$$\sum_{g:c} y_{i_c g_c} b_{g_c p} + y_{i_c G_c} b_{G_c p} = z_{i_c p} , \tag{13}$$

where g_c does not include G_c . When g_c is permitted to include G_c , (13) reduces to

$$\sum_{g:c} y_{i_c g_c} b_{g_c p} = z_{i_c p} . \tag{14}$$

Scattered data matrices Y, B, and Z for the criterion variables are defined parallel to scattered data matrices X, A, and F, respectively, for the predictor variables. These definitions and (14) yield (15), which parallels (12).

$$YB = Z. \tag{15}$$

In the present formulation no provision will be made for tests on the criterion side of the system. If a situation arises in which test scores or other variables, that extend beyond single colleges, become relevant, these variables could be included by extending the matrices Y and B in a manner parallel to the extension of the matrices X and A for test scores on the predictor side of the system.

Illustration of Scattered Data Matrices

An illustration of the scattered data matrices X and Y is given in Table 2 in which the grades and test scores are given for ten individuals from the illustrative study. Note that the school grade variables, the additive constant dummy variables, and the test score variables are numbered consecutively from $j = 1$ to $j = 41$ in the matrix X. The college grade variables and additive constant dummy variables are numbered consecutively from $g = 1$ to $g = 12$ in the matrix Y. The students are numbered consecutively within college, and the rows of the predictor variable matrix X have been arranged accordingly. Consequently, the students from each school do not appear as a block as they do for each college in the criterion variable matrix Y. The arrangement of the rows of X, however, does not affect the properties of this matrix.

Each student has one grade in one school in matrix X and a score of unity in the additive constant dummy variable for the same school. The entries in the row for each student in columns for the schools he did not attend are set at zero in the construction of the matrix. Nine of the students have scores on the College Entrance Examination Board Scholastic Aptitude Test; only individual 70 does not have a test score. These scores for the nine individuals are entered in the test variable columns of matrix X, and unity is entered in the additive constant dummy variable for the test. All three entries for individual 70 are left blank and are set at zero.

In matrix Y, the college grade for each student is entered in the grade column for the college he attended and unity is entered in the corresponding additive constant dummy variable. All other entries are left blank and defined as zero.

A row of weights is given beneath each of the matrices X and Y. The determination of these weights will be discussed in a later section. These weights are used here to illustrate use of the scattered data matrices in obtaining the composite scores given at the right of each matrix. The predictor composite score for each student is obtained by finding the sum of products

TABLE 2

Illustration of Scattered Data Matrices*

Predictor Score Matrix X

Student	School 1 Grade $j = 1$	School 1 Const. $j = 2$	School 2 Grade $j = 3$	School 2 Const. $j = 4$	···	SAT Scores Verb. $j = 39$	SAT Scores Math. $j = 40$	SAT Scores Const. $j = 41$	Predictor Composite Score
1	.00	1.00				−.22	−.17	1.00	.40
2	.25	1.00				.90	1.98	1.00	.59
3	.25	1.00				.95	.22	1.00	.39
4	.00	1.00				1.00	.81	1.00	.65
5			1.25	1.00		1.43	1.49	1.00	.37
67	.25	1.00				−.75	1.59	1.00	.36
68	.25	1.00				1.69	2.17	1.00	.70
69			1.75	1.00		.95	1.69	1.00	.48
70			1.50	1.00					.43
71			1.25	1.00		1.69	1.69	1.00	.42
Weights†	−.74	.77	.28	.01		.11	.12	−.33	

Criterion Score Matrix Y

Student	College 1 Grade $g = 1$	College 1 Const. $g = 2$	···	College 3 Grade $g = 5$	College 3 Const. $g = 6$	···	College 6 Grade $g = 11$	College 6 Const. $g = 12$	Criterion Composite Score
1	−.97	1.00							.45
2	−.78	1.00							.45
3	.14	1.00							.45
4	−.13	1.00							.45
5	1.09	1.00							.45
67				−.10	1.00				.47
68				.46	1.00				.50
69				.28	1.00				.49
70				.00	1.00				.48
71				−1.56	1.00				.40
Weights†	.00	.45		.05	.48		.46	.36	

*All blank cells are to be interpreted as filled with zeros.

†The weights are for the first total covariance and variance canonical variate (to be described in a later section).

between the values in his row of X and the values in the row of weights at the bottom of matrix X. The criterion composite score for each individual is the sum of products between the values in his row in matrix Y and the values in the row of weights at the bottom of matrix Y.

The construction for including test scores in the predictive composite scores is illustrated also in Table 2. One concern to a number of people when they are first introduced to this construction is that a student receives an automatic bonus when he takes the test. Remember that the test scores were scaled in the study so that a 500 on the CEEB scale was translated to zero and that a difference of 100 on the CEEB scale was made unity. The SAT scores of individual 1 on the CEEB scale were 478 and 483 which were transformed to the $-.22$ and $-.17$ recorded in matrix X and used in the study. The contribution of the test scores to the composite score for this individual is

$$(-.22)(.11) + (-.17)(.12) + (1.00)(-.33) = -.3746.$$

This individual would have had a predictive composite score of .77 if he had not taken the test. The test scores for individual 2 were high enough to counterbalance the additive constant of $-.33$. Other individuals made high enough scores to obtain a positive contribution to their predictive composite scores, as for example individuals 68 and 71. Thus, taking the test need not necessarily yield a positive contribution (a bonus) to the student's predictive composite. The construction provides adjustments of the predictive composite scores for the test scores made by the students.

Linear Scaling Transformation of Predictor Variables and Criterion Variables

Procedures for linear scaling transformations of variables in a scattered data matrix is a topic of special importance. The grades at each school, for example, could be expressed in terms of two different scales such as the raw grades given to the students and in terms of deviations from the mean grade given at the school. From these two statements of the school grades, two scattered data matrices can be constructed. Major questions involve the relations between these two scattered data matrices and the appropriate transformations of weight matrices A in order to maintain the composite scores in F as constants. Other questions involve procedures for obtaining composite scores which are deviation scores from the mean for all students in the study, or which are deviation scores for the students at each college from the mean for that college.

Special constructions and procedures are defined to accomplish linear scaling transformations of the predictor variables and criterion variables in scattered data matrices. The form of these constructions and procedures will be discussed explicitly for a predictor variable scattered data matrix with grades from a number of schools, but no test scores. Parallel constructions and procedures apply to the criterion variable scattered data matrices and will not be discussed. The constructions and procedures for a predictor variable scattered data matrix having both school grades and test scores are a simple extension of the constructions and procedures for the scattered data matrix having only school grades and will not be discussed.

Let the transformed school grades be designated by x^*_{i,k_s} and be defined by (16).

$$x^*_{i,k_s} = \sum_{j:s} x_{i,j_s} t_{j_s k_s} + x_{i,J_s} t_{J_s k_s} , \tag{16}$$

where $t_{j_s k_s}$ are the transformation coefficients and $t_{J_s k_s}$ is the additive constant, all for school s, and j_s does not include J_s. Note that the translation of grades is accomplished by adding the product of the additive constant and the dummy variable x_{i,J_s}, and not just adding the additive constant. This construction provides a functional difference between observed zero grades, as for individual 1 at school 1 in Table 2, and the zeros entered in a grade column for individuals who did not attend the particular school. All observed grades may be translated by the foregoing procedure but the artificial zeros for students who did not attend the school will remain zero since the value of the dummy variable is zero for them.

The subscript k_s is used to designate the transformed grades at school s. The number of k_s is to be the same as the number of j_s for each school. The construction used here provides not only for linear rescaling of the grades in each category but also for combination grades. For example, if there were two categories of grades for a school, the first j_s might be for verbal subjects and the second j_s might be for quantitative subjects; the first k_s might be for a sum of the two j_s categories and the second k_s might be for a difference between the two j_s categories. A restriction is made in the present topic that the combination grades involve the grades at each school separately and not at two or more schools. A second restriction is that the transformations are linearly independent for each school.

The additive constant dummy variable is carried into the transformed group of grades by defining the coefficients $t_{j_s K_s}$ and $t_{J_s K_s}$ as follows:

$$t_{j_s K_s} = 0, \tag{17}$$

$$t_{J_s K_s} = 1, \tag{18}$$

where K_s is the additive constant dummy variable. The score x^*_{i,K_s} is defined by (16) when K_s is substituted for k_s. The preceding definitions yield, then,

$$x^*_{i,K_s} = x_{i,J_s} . \tag{19}$$

Thus, the additive constant dummy variable remains unchanged in the linear rescaling transformations.

The coefficients $t_{j_s k_s}$, $t_{j_s K_s}$, $t_{J_s k_s}$, and $t_{J_s K_s}$ form a square matrix which may be designated T_s. The linear independence restriction stated previously is to be taken to imply that T_s is nonsingular so that an inverse exists. Let the inverse of T_s be designated by M_s with entries $m_{k_s i_s}$, $m_{k_s J_s}$, $m_{K_s i_s}$, and $m_{K_s J_s}$. A consequence of the definitions in (17) and (18) is that

$$m_{k_s J_s} = 0, \tag{20}$$

$$m_{K_s J_s} = 1. \tag{21}$$

Solution of (16) and (19) for $x_{i_s j_s}$ yields

$$x_{i_s i_s} = \sum_{k:s} x^*_{i_s k_s} m_{k_s i_s} + x^*_{i_s K_s} m_{K_s i_s} . \tag{22}$$

Equations (16) and (19) may be stated in matrix form as

$$X^*_s = X_s T_s , \tag{23}$$

and (19) and (22) may be stated in matrix form as

$$X_s = X^*_s M_s , \tag{24}$$

where matrix X^*_s is the matrix of transformed grades $x^*_{i_s k_s}$ and $x^*_{i_s K_s}$.

The effects of the linear rescaling transformations of the grades on the weights for determining the composite scores can be traced most easily in matrix form. Substituting in (8) the value of X_s from (24),

$$X^*_s M_s A_s = F_s . \tag{25}$$

Define

$$A^*_s = M_s A_s ; \tag{26}$$

then (25) yields

$$X^*_s A^*_s = F_s . \tag{27}$$

Equation (26) defines the transformed weights which, when applied to the transformed grades, yields the same predictive composite scores. Equation (27) is in a form which is identical to (8).

In order to carry out the linear scaling transformation of grades at every school with the scattered data matrices the following matrices are defined.

$$T = \begin{bmatrix} T_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & T_s & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & T_{n_s} \end{bmatrix} , \tag{28}$$

(29)
$$M = \begin{bmatrix} M_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & M_s & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & M_{n_s} \end{bmatrix} .$$

Scattered data matrices X^* and A^* are defined for the transformed grades and weights similar to the matrices X and A in (9) and (10). The foregoing definitions and (23) through (27) yield

(30) $$X^* = XT,$$

(31) $$X = X^*M,$$

(32) $$X^*MA = F,$$

(33) $$A^* = MA,$$

(34) $$X^*A^* = F.$$

Note that the defined zero sections of matrix X are still zero in matrix X^* of transformed grades.

Linear Transformation of Composite Scores

The material in this section will be limited to one class of transformations, that of transforming an array of composite scores to deviation form from the mean of a group of individuals. Two subclasses will be considered: in the first subclass the deviation scores for all individuals are taken from the mean for the entire group of individuals, in the second subclass the deviation scores for the students attending each college are taken from the mean score for the group attending that college. Since parallel constructions apply to the predictor variable matrix X and the criterion variable matrix Y, only those for the matrix X will be considered in detail. However, certain matrices will be defined for the criterion variable matrix.

Two derived procedures will be considered for obtaining deviation composite scores from the mean composite score for all individuals. In both procedures, A_p will be considered as a single column of weights from a matrix A. There are no restrictions on the contents of this column (aside from the general assumption of all numbers being real), and, by successive applications of the following principles to various possible columns, the following results can apply to any and all possible columns of weights. All columns must be of length n_i, the number of predictor variables. F_p is the column of composite scores corresponding to A_p so that, by (12),

$$XA_p = F_p . \tag{35}$$

A convenient matrix construction involves the definition of a row vector S having as many entries as there are individuals in the study. Every entry in S is unity. If the column vector, F_p , is premultiplied by S, the result is a single number, the sum over individuals of the entries f_{ip} in F_p .

$$SF_p = \sum_i f_{ip} . \tag{36}$$

The mean f_{ip} over all individuals will be designated by $\bar{f}_{.p}$ and is given in (37).

$$\bar{f}_{.p} = \frac{1}{N_{..}} \sum_i f_{ip} , \tag{37}$$

and, by (36),

$$\bar{f}_{.p} = \frac{1}{N_{..}} SF_p . \tag{38}$$

The deviation composite score for individual i from the mean for all individuals, $\bar{f}_{.p}$, will be designated by f^t_{ip} and will be defined by (39).

$$f^t_{ip} = f_{ip} - \bar{f}_{.p} , \tag{39}$$

or, in matrix form,

$$F^t_p = F_p - S'\bar{f}_{.p} , \tag{40}$$

in which the function of S' in the last term is to produce a column vector with an entry for each individual. Since S', being the transpose of S, has an entry of unity for each individual, and $\bar{f}_{.p}$ is a scalar quantity, the product $S'\bar{f}_{.p}$ produces a column vector with an entry of $\bar{f}_{.p}$ for each individual. F^t_p is a column vector of deviation composite scores from $\bar{f}_{.p}$.

The first derived procedure for obtaining the column vector F^t_p depends on noting that each individual has one and only one entry of unity in the matrix X on the additive constant dummy variables J_s . His entries on the other dummy variables have been set at zero. This observation depends on an implicit assumption that each individual entered as a regular candidate into a central prediction system will have graduated from one and only one secondary school. Let the column vector A_J be defined as a column of weights with an entry of unity for each school's additive constant dummy variable, J_s , and entries of zero for all other variables. The preceding observation yields the equation

$$XA_J = S', \tag{41}$$

that is, the sum of the entries in the dummy variables for the schools is unity for every individual.

In case it is desired to enter some students who have attended more than one school into a central prediction system, and it is desired to give different weights to their grades according to the weights for the schools they did attend, a construction is possible which will satisfy all of the preceding equations. This construction would lead to a number of complexities and will not be given here.

Substitution in (40) for F_p from (35) and for S' from (41) yields

$$F_p^t = XA_p - XA_J\bar{f}_{.p} , \tag{42}$$

or

$$F_p^t = X(A_p - A_J\bar{f}_{.p}). \tag{43}$$

Let

$$A_p^t = (A_p - A_J\bar{f}_{.p}); \tag{44}$$

then,

$$F_p^t = XA_p^t . \tag{45}$$

Equation (44) gives the procedure for revising a column of weights, A_p , so that the revised weights, A_p^t , will yield deviation composite scores from the mean for all individuals.

In the second derived procedure for obtaining the column vector F_p^t , a substitution is made in (40) from (38) for the value of $\bar{f}_{.p}$.

$$F_p^t = F_p - S' \frac{1}{N_{..}} SF_p , \tag{46}$$

$$= \left(I - S' \frac{1}{N_{..}} S\right)F_p . \tag{47}$$

Substitution from (35) for F_p yields

$$F_p^t = \left(I - S' \frac{1}{N_{..}} S\right)XA_p . \tag{48}$$

Define a matrix ξ^t as

$$\xi^t = \left(I - S' \frac{1}{N_{..}} S\right)X. \tag{49}$$

If the matrix X were completely full of observed scores, the entries in ξ^t would be termed deviations from the means of the variables; however, since matrix X is a scattered data matrix with a number of artificial zero entries, this interpretation of the entries in ξ^t is not proper. The matrix operations leading to the matrix ξ^t, as well as those using it, are proper. Equations (48) and (49) yield

$$F_p^t = \xi^t A_p . \tag{50}$$

The matrix ξ^t is a matrix such that postmultiplication by any column vector A_p will yield the corresponding deviation composite score vector F_p^t.

An important property of the matrix ξ^t is developed by forming the product $\xi^t A_J$ and substituting for the value of ξ^t from its definition in (49).

$$\xi^t A_J = \left(I - S' \frac{1}{N_{..}} S\right) X A_J , \tag{51}$$

$$= \left(I - S' \frac{1}{N_{..}} S\right) S' \tag{52}$$

when substitution is made from (41) for XA_J. By the construction of vector S with $N_{..}$ entries of unity

$$SS' = N_{..} . \tag{53}$$

When the S' in (52) is multiplied into the parentheses and a substitution is made from (53) for the SS' in the last term,

$$\xi^t A_J = S' - S' \frac{1}{N_{..}} N_{..} = 0. \tag{54}$$

Consequently, the rank of matrix ξ^t is at least one less than the number of columns n_j. This result must be remembered when operating with the matrix ξ^t in the following models.

Consider the deviation of a student's composite score from the mean composite score for all individuals attending the college which he attends. This will be termed his *within-college deviation composite score* and will be designated by f_{icp}^c. Note the use of the superscript c to denote quantities related to the within-college deviation composite scores. The mean of the original composite scores for a given college c, designated by $\bar{f}_{cp}$, is given by

$$\bar{f}_{cp} = \frac{1}{N_{.c}} \sum_{i:c} f_{icp} . \tag{55}$$

Then, f_{icp}^c is defined by

$$f_{icp}^c = f_{icp} - \bar{f}_{cp} . \tag{56}$$

A convenient construction is to define a row vector S_c^c for each college (the subscript and not the superscript c will denote the college) with an entry of unity for each individual who is attending that college and of zero for each other individual. These row vectors are collected into a matrix (S^c). Note that by this construction

$$(S^c)(S^c)' = [N_{.c}], \tag{57}$$

where $[N_{.c}]$ is a diagonal matrix containing the number of students at each college. With the preceding construction, a column vector $\bar{F}_{cp}$ containing the mean composite scores of the groups of students attending the colleges

$\bar{f}_{cp}$ may be determined by

(58) $$\bar{F}_{cp} = [N_{.c}]^{-1}(S^c)F_p \,.$$

The matrix F_p^c of within-college deviation composite scores f_{icp}^c is given by

(59) $$F_p^c = F_p - (S^c)'\bar{F}_{cp} \,.$$

Substituting for $\bar{F}_{cp}$ in (59) from (58),

(60) $$F_p^c = F_p - (S^c)'[N_{.c}]^{-1}(S^c)F_p \,,$$

(61) $$= \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}F_p \,.$$

Substituting for F_p from (35),

(62) $$F_p^c = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}XA_p \,.$$

Define the matrix ξ^c by

(63) $$\xi^c = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}X;$$

then

(64) $$F_p^c = \xi^c A_p \,.$$

Note the similarity between ξ^t defined in (49) and ξ^c defined in (63). Also note the similarity between equations (50) and (64). Postmultiplication of ξ^t by A_p yields deviation composite scores from the mean for all individuals, and, in comparison, postmultiplication of ξ^c by A_p yields within-college deviation composite scores. Insofar as A_p was defined to be any possible column vector of weights, (50) gives every possible column F_p^t of deviation composite scores from the mean of all individuals, and (64) gives every possible column F_p^c of within-college deviation composite scores. Furthermore, every F_p^t resulting from (50) is a possible column of deviation composite scores from the mean of all individuals, and every F_p^c resulting from (64) is a possible column of within-college deviation composite scores. As a consequence, the matrix ξ^t forms a basis for the field of vectors F_p^t, and the matrix ξ^c forms a basis for the field of vectors F_p^c. These matrices will be used in this sense in the development of models for a central prediction system.

An important property of ξ^c is developed by forming the product $\xi^c A_J$ and substituting for ξ^c from (63),

(65) $$\xi^c A_J = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}XA_J \,,$$

and, by substituting for XA_J from (41),

(66) $$\xi^c A_J = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}S'.$$

Define a row vector [1] with an entry of unity for each college. Assuming that each student in a central prediction study will attend one and only one of the colleges, the construction of the matrices S and (S^c) yields

(67) $$[1](S^c) = S.$$

Substituting in (66) for S' from (67),

(68) $$\xi^c A_J = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}(S^c)'[1]'.$$

Multiplication of the last $(S^c)'$ into the quantity in braces and substituting for the resulting (S^c) $(S^c)'$ in the last term from (57)

(69) $$\xi^c A_J = \{(S^c)' - (S^c)'[N_{.c}]^{-1}[N_{.c}]\}[1]',$$

(70) $$= 0.$$

Thus, as is true of ξ^t, the rank of ξ^c is at least one less than the number of columns n_j .

Before going on to consider specific models for a central prediction system, several matrices will be defined and equations presented for the criterion side by analogy with the preceding matrices for the predictor side.

Let Z_p^t be a column vector of deviation criterion composite scores from the mean for all individuals

(71) $$Z_p^t = Z_p - S'\bar{z}_{.p} ,$$

where Z_p is a column vector of criterion composite scores corresponding to a column vector of weights B_p , and $\bar{z}_{.p}$ is the mean over all individuals of the entries in Z_p . This definition is analogous to that of F_p^t in (40). Let B_p^t be defined by

(72) $$B_p^t = B_p - B_J\bar{z}_{.p} ,$$

where B_J is a column vector like A_J with unit weights for the additive constant dummy variables and zero weights for the other variables. This definition is analogous to that of A_p^t in (44). By analogy with (45)

(73) $$Z_p^t = YB_p^t .$$

By analogy with the definition of ξ^t in (49), let the matrix ζ^t be defined

(74) $$\zeta^t = \left(I - S' \frac{1}{N_{..}} S\right)Y.$$

Then, by analogy with (50),

(75) $$Z_p^t = \zeta^t B_p .$$

The matrix ζ^t has all the properties of the matrix ξ^t, but involves the criterion variables.

By analogy with $\bar{F}_{cp}$, let $\bar{Z}_{cp}$ be a column vector of means of the criterion composite scores for the students at the several colleges, and, by analogy with (58),

(76) $$\bar{Z}_{cp} = [N_{.c}]^{-1}S_cZ_p .$$

By analogy with (59), a column vector of within-college deviation criterion composite scores is defined as

$$(77) \qquad Z_p^c = Z_p - (S^c)'\bar{Z}_{cp} .$$

By analogy with the definition of ξ^c in (63), let the matrix ζ^c be defined

$$(78) \qquad \zeta^c = \{I - (S^c)'[N_{.c}]^{-1}(S^c)\}Y.$$

Then, by analogy with (64),

$$(79) \qquad Z_p^c = \zeta^c B_p .$$

A point to note about the matrix ζ^c is that all entries in the additive constant dummy variable columns become zero. This result is most easily understood from the similarity of the operation defining matrix ζ^c and that of obtaining deviation scores within each college. The entries in each additive constant dummy variable in matrix Y for each college are a constant, unity for the dummy variable for the college and zero for the dummy variables for other colleges. The mean entry for each such variable for each college has the same value, unity or zero, as the entries in the column; consequently, the deviations in each column from the mean for the college are all zero. These additive constant dummy variables may be eliminated from the matrix ζ^c, thus reducing its order. The corresponding rows of B_p are to be eliminated.

Canonical Correlation Model I: Total Covariances and Variances

The basic principle for the first quantitative model for a central prediction system is that the weights be determined so as to maximize the correlation between the predictor composite scores and the criterion composite scores. Granting the desirability of high correlations, since the goodness of predictions is measured by the size of the correlations involved, the foregoing principle seems to have considerable face validity. A solution on this basis could provide a means for scaling both the predictor variables and the criterion variables. This corresponds to the objectives of much of the previous work.

Before this system is adopted, however, its properties should be investigated. The formal model may be established in terms of the constructs developed in the preceding sections. Since the correlation coefficient is most simply stated in terms of deviation scores, consider the column vectors F_p^t and Z_p^t of deviation composite scores from the means of all individuals for the predictor variables and the criterion variables respectively. These column vectors were defined in (40) and (71). The parameters in this system are the column vectors of weights, A_p and B_p , which are to be determined so that the correlation between the entries in F_p^t and Z_p^t is maximized.

Equations (50) and (75) provide the most convenient statement of the dependence of the column vectors F_p^t and Z_p^t on the weight vectors A_p and B_p , respectively. These statements involve the matrices ξ^t and ζ^t defined in (49) and (74). A very helpful observation is to note the identity in form between the model stated in terms of the matrices ξ^t and ζ^t and the canonical correlation model. [For a discussion of canonical correlations see Kendall (1951), Vol. 2, pp. 348–358.] This identity between the two models provides access to the known properties of the canonical correlation model. These properties may be applied to the model being discussed here.

The most important property in the present context of the canonical correlation model is the existence of a number of solutions corresponding to the characteristic roots of a derived matrix. This property raises questions concerning the more or less traditional formulation of our problem as that of a rescaling of the grades at the various schools and colleges to a single, common scale. In case the traditional formulation is correct, there should be only one large canonical correlation. Several large canonical correlations would cast considerable doubt on the traditional formulation. It is reasonable, however, to attempt to think of possible sources of several large canonical correlations. One plausible possibility is that different types of schools and colleges might generate several types of scales of adjusted grades. That is, different adjustments of grades relatively between general college preparatory secondary schools and technical secondary schools may be appropriate when the adjusted grades are being related to grades at liberal arts colleges than when the adjusted grades are being related to grades at technical colleges. Grades at college preparatory schools may be more highly related to grades at liberal arts colleges than to grades at technical colleges. The reverse may be true for technical schools. To the extent that these differences exist, two large canonical correlations could be generated.

A closer inspection of the situation reveals another possibility. In an extreme case the schools and colleges could separate into two subgroups so as to form two systems such that students from school subgroup I go only to colleges in college subgroup I and students from school subgroup II go only to colleges in college subgroup II. This would result in a canonical correlation of unity. The only nonzero weights in A_p would be for the additive constant dummy variable for which weights would be a constant, say f_{I}^t, for the schools in subgroup I and another constant, say f_{II}^t, for the schools in subgroup II. As a consequence, the predictive composite scores for all students from subgroup I of schools would be f_{I}^t, and the predictive composite scores for all students from subgroup II of schools would be f_{II}^t. An identical situation would exist for the colleges. The values of f_{I}^t and f_{II}^t would be such that

$$N_{\text{I}}f_{\text{I}}^t = N_{\text{II}}f_{\text{II}}^t \,, \tag{80}$$

$$N_{\text{I}}f_{\text{I}}^{t^2} + N_{\text{II}}f_{\text{II}}^{t^2} = N_{..} \,, \tag{81}$$

where N_{I} and N_{II} are the numbers of students in systems I and II, respectively. In the foregoing case every student would obtain an identical predictor composite score and criterion composite score. This canonical variate would be independent of any other variate related to differences in grades within schools and colleges.

A third possible effect is also related to the additive constants, especially those for the colleges. This effect might or might not produce a second large canonical correlation. The guidance programs at secondary schools may affect the college-going plans of their students such that the students having high grades apply to higher prestige colleges, students with middle grades apply to middle prestige schools, and students with low grades apply to low prestige colleges. The admission actions of the colleges tend to help enforce this tendency, each college accepting the most able applicants up to its capacity. Thus, generally, only the most able students are accepted at the high prestige colleges. It is unprofitable, usually, for a student with low grades to apply to a high prestige college. The purpose here is not to comment on any moral issue, but to point out a possible source of correlation between secondary school grades and the colleges to which the students go. The additive constants, or weights for the additive constant dummy variables, for the colleges can produce criterion composite scores which reflect the prestige levels of the colleges if these prestige levels are correlated with the secondary school grades. This effect could affect the values of one or more canonical correlations.

In the usual solution for the canonical correlations and weights, inverses are used of the matrices of intercorrelations of the predictor variables and of the criterion variables. Because the ranks of the matrices ξ^t and ζ^t are at least one less than the numbers of variables, as per the implication of (54), these matrices of intercorrelations are singular and inverses do not exist. A solution for the present case is given in Appendix A.

In the introductory section of this report a suggestion was made that the internal method of Bloom and Peters (1961) could be considered as a variant of the total covariances and variances canonical model. In their procedure they assumed a vector of weights for the college grades and computed a vector of adjusted college grades, which can be identified here as a column of Z_p. They established trial school weights as the regression weights for predicting the trial adjusted grades from the school grades. This step is the same as indicated in equation (A.9) in Appendix A except that Bloom and Peters did not provide for the scaling of the adjusted school grades. In the next half of one of their cycles they reversed the function of the school and college grades which is analogous to (A.11) except, again, for the scaling of the new adjusted grades. In that the correlation between the two arrays of adjusted grades is increased at every half cycle, it may be concluded that the Bloom and Peters internal method would converge to the largest canonical correlation solution for total covariances and variances provided that the adjusted grades were scaled so as not to vanish.

TABLE 3
Canonical Correlations
(Canonical Correlation Model I: Total Covariances and Variances)

Variate Number	Canonical Correlation	Variate Number	Canonical Correlation
1	.90	7	.44
2	.80	8	.42
3	.68	9	.30
4	.60	10	.29
5	.56	11	.23
6	.46		

Results for the illustrative study are presented in Tables 3–6 for the Canonical Correlation Model I for total covariances and variances. The eleven canonical correlations are given in Table 3, the number of coefficients being equal to the rank of the basis matrix for college grades, one less than twice the number of colleges. Several of these canonical correlations seem quite high in terms of predictive validity. While there is considerable capitalization on chance due to the small number of cases used in the study, several of these correlations probably would remain high for a larger sample. This evidence is indicative of the existence of more than one adjusted grade scale.

Tables 4 and 5 give the weights for the first three canonical variates. In Table 4 the weights are given as in matrices A and B with one column for each canonical variate and the original variables in a single list with consecutive index numbers. This is the format specified in (10) for use with the scattered data matrices X and Y. This format does not aid inspection and interpretation of the results, however, and the weights are listed in a different format in Table 5 so as to be easier to inspect. The format in Table 5 is more appropriate for the separate variable formulation with the two weights for one school for each canonical variate being given on one line. All of the weights on each canonical variate for the grades at the schools are given in one column, and all of the additive constants for the schools are given in another column. The same is true for the colleges.

Note the large negative additive constant on the first canonical variate for college 2 and the large negative additive constants for schools 11–14. As noted for Table 1, most of the students from these schools went to college 2 and most of the students at college 2 came from these schools. None of the other colleges has a negative additive constant and only three other schools, and these are schools from which one or two students went to college 2, have negative additive constants. This result corresponds to the possible effects

TABLE 4
Weights for Variables, Scattered Data Matrix Form
(Canonical Correlation Model I: Total Covariances and Variances)

Predictor Variable Weights

School	Variable	Index No.	Canonical Variate 1	2	3
1	Grade	1	−.74	6.48	−2.43
	Const.	2	.77	−2.93	.45
2	Grade	3	.28	.35	.29
	Const.	4	.01	−.29	−1.02
3	Grade	5	.44	.17	−.15
	Const.	6	−.16	.06	−.18
4	Grade	7	.52	−.01	.13
	Const.	8	−.32	.05	−.76
5	Grade	9	−.04	.38	.51
	Const.	10	.24	−1.10	−1.16
6	Grade	11	1.21	.26	.96
	Const.	12	−2.34	−.76	−2.91
7	Grade	13	.16	.26	.37
	Const.	14	.28	−.15	−1.03
8	Grade	15	.00	.16	.04
	Const.	16	.44	.22	−.41
9	Grade	17	.23	.60	.21
	Const.	18	.34	−.73	−.69
10	Grade	19	.06	.09	.15
	Const.	20	.62	.03	−.66
11	Grade	21	.32	−.12	.73
	Const.	22	−2.22	.22	−1.08
12	Grade	23	.58	.30	1.99
	Const.	24	−1.78	.30	1.78
13	Grade	25	.53	.07	1.06
	Const.	26	−1.93	.30	.48
14	Grade	27	1.16	.21	1.53
	Const.	28	−2.95	−.18	−3.15
15	Grade	29	.48	−.40	.64
	Const.	30	.25	−.91	−.79
16	Grade	31	−.08	.05	−.09
	Const.	32	.58	−.02	−.55
17	Grade	33	−.01	.79	−.06
	Const.	34	.53	−.64	−.37
18	Grade	35	−.15	1.85	−.86
	Const.	36	.60	−2.84	.32
19	Grade	37	.02	.98	−.27
	Const.	38	.61	−1.22	−.16
SAT	Verb.	39	.11	.00	.22
	Math.	40	.12	.30	−.09
	Const.	41	−.33	−.18	.25

Criterion Variable Weights

College	Variable	Index No.	Canonical Variate 1	2	3
1	Grade	1	.00	1.31	−.41
	Const.	2	.45	−1.72	.24
2	Grade	3	.54	.38	2.32
	Const.	4	−2.20	.17	.56
3	Grade	5	.05	.05	−.04
	Const.	6	.48	.48	−.19
4	Grade	7	.09	.36	.30
	Const.	8	.32	.44	−.08
5	Grade	9	.48	.47	.34
	Const.	10	.45	.22	−.24
6	Grade	11	.46	.38	.42
	Const.	12	.36	.34	−.40

TABLE 5
Weights of Variables, Separate Variable Form
(Canonical Correlation Model I: Total Covariances and Variances)

Predictor Variable Weights

	Canonical Variate 1		Canonical Variate 2		Canonical Variate 3	
School	Grade	Constant	Grade	Constant	Grade	Constant
1	−.74	.77	6.48	−2.93	−2.43	.45
2	.28	.01	.35	−.29	.29	−1.02
3	.44	−.16	.17	.06	−.15	−.18
4	.52	−.32	−.01	.05	.13	−.76
5	−.04	.24	.38	−1.10	.51	−1.16
6	1.21	−2.34	.26	−.76	.96	−2.91
7	.16	.28	.26	−.15	.37	−1.03
8	.00	.44	.16	.22	.04	−.41
9	.23	.34	.60	−.73	.21	−.69
10	.06	.62	.09	.03	.15	−.66
11	.32	−2.22	−.12	.22	.73	−1.08
12	.58	−1.78	.30	.30	1.99	1.78
13	.53	−1.93	.07	.30	1.06	.48
14	1.16	−2.95	.21	−.18	1.53	−3.15
15	.48	.25	−.40	−.91	.64	−.79
16	−.08	.58	.05	−.02	−.09	−.55
17	−.01	.53	.79	−.64	−.06	−.37
18	−.15	.60	1.85	−2.84	−.86	.32
19	.02	.61	.98	−1.22	−.27	−.16
SAT Scores						
Verb.	.11		.00		.22	
Math.	.12		.30		−.09	
Const.	−.33		−.18		.25	

Criterion Variable Weights

	Canonical Variate 1		Canonical Variate 2		Canonical Variate 3	
College	Grade	Constant	Grade	Constant	Grade	Constant
1	.00	.45	1.31	−1.72	−.41	.24
2	.54	−2.20	.38	.17	2.32	.56
3	.05	.48	.05	.48	−.04	−.19
4	.09	.32	.36	.44	.30	−.08
5	.48	.45	.47	.22	.34	−.24
6	.46	.36	.38	.34	.42	−.40

of the existence of separate systems of schools and colleges. In a similar manner, the second canonical variate appears to be related to school-college relations as to common paths for students to follow, in that college 1 is singled out for a negative additive constant with the schools from which the students came who went to college 1.

As a further indication of the values of the composite scores determined by the several total covariance and variance canonical variates, the correlations between the predictor composite scores and the grades at each college were computed. These correlations are presented in Table 6. The correlations

TABLE 6
Correlations of Predictor Composite Scores on Canonical Variates with Grades at Each College
(Canonical Correlation Model I: Total Covariances and Variances)

Canonical Variate	College 1	2	3	4	5	6
1	.04	.58	.13	.23	.75	.54
2	.70	.67	.11	.49	.68	.52
3	−.53	.78	.00	.24	.51	.57

in the first row for the first canonical variate are markedly less than the canonical correlation of .90. Only one coefficient in this row is greater than .60. The correlations in the second row appear in general to be somewhat higher than those in the first row. Three correlations in the second row are greater than .60. Another peculiar feature of the system of canonical variates is indicated by the negative correlation in the third row for the first college. This coefficient corresponds to the negative weight given in Tables 4 and 5 to grades at this school for the third canonical variate. When the correlations in the third row are considered in absolute value they appear to be of the same order of magnitude as the correlations for the first canonical variate.

When Table 6 is inspected by columns, the low correlations at college 3 are noteworthy. None of the canonical variates would provide a worthwhile predictor of grades at this college. Of the other colleges, only at college 5 is the correlation with the first variate higher than the correlation with other variates. The highest correlation at college 1 and college 4 is with the second variate while the highest correlation at colleges 2 and 6 is with the third variate.

A possible interpretation of the results given in Tables 4–6 is that the uneven linkages as to numbers of students for each school-college combination is producing considerable disturbing effects on the system.

Canonical Correlation Model II:
Within-College Covariances and Variances

An attempt is made in this model to eliminate the seemingly untoward effects of the uneven linkages between schools and colleges as to numbers of students associated with the various school-college combinations. The mechanism of this effect seemed to be the additive constants. Another possible effect associated with the additive constants for the colleges is the effect of the prestige of the colleges inducing a correlation between the secondary school grades and the colleges to which the students went. Some people might argue that the basis of a central prediction system should not include this possible effect. In order to avoid these effects associated with the additive constants and to make the basis of the system depend on the differences in grades within each college, the within-college deviation composite scores are used in the model to be discussed in this section.

Matrices ξ^c and ζ^c of (63) and (78) form the bases of the within-college deviation composite scores in F_p^c and Z_p^c as given in (64) and (79). The solution for this model is obtained by use of the superscript c for the matrices in Appendix A.

Results for the illustrative study are given in Tables 7–10. The canonical correlations in Table 7 are distinctly lower than those given in Table 3 for

TABLE 7
Canonical Correlations
(Canonical Correlation Model II:
Within-College Covariances and Variances)

Variate Number	Canonical Correlation
1	.73
2	.65
3	.59
4	.44
5	.38
6	.30

the total covariances and variances model discussed in the preceding section. The highest coefficient has dropped from .90 in the preceding model to .73 for the present model, the second highest has dropped from .80 to .65, etc. This effect was anticipated when the between-colleges contributions to the covariances and variances were excluded from the present model, whereas they had been included in the total covariances and variances model. The present correlations represent, however, relations between differences on the predictor variables and differences in grades within the colleges. In this context, the correlations are quite high for the present model.

A point to note is that there are only six canonical variates for the illustrative study. This result is associated with the reduction of rank of the criterion basis matrix ζ^c due to the vanishing of the additive constant dummy variables for the colleges, as noted following (79) at the end of the section on Mathematical Constructions.

Table 8 presents the weights for the first three canonical variates for the within-college covariances and variances model. This table is set up in the separate variable form to facilitate inspection of the weights and additive constants. Note for the first canonical variate that the weights for the grades at every school except 16 are positive and that the additive constants for every school are negative. This is a much more regular situation than was true for any of the canonical variates for the total covariances and variances model given in Table 5. The positive weights for the school grades should be expected from the general positive correlation between school grades and college grades. The negative additive constants are a reflection of the scaling of the school grades such that the mean grade for all students in a class at a school was zero, but, generally, only the students with higher school grades entered the colleges in the study. The means of the school grades for the students in the study generally are positive. In order to obtain deviation composite scores for the students in the study, it is necessary to subtract some constant from the school grades.

Note in the criterion variable section of Table 8 that for the first canonical variate for every college the weights for the grades are positive. This reflects the same observation as pointed out above that generally there is a positive correlation between school grades and college grades. No entries are made in the additive constant column for the colleges. This is a result of the scaling of the college grades as standard scores for the group of students in the study and in each college, separately by college, coupled with the use of within-college deviation composite scores. Since the grades at each college were in deviation form, no additive constant was necessary to obtain the deviation composite scores.

The appearance of negative weights in Table 8 on the second and third canonical variate for the grades at some schools and colleges is to be expected from the property of the canonical model that the composite scores for each variate are uncorrelated with the composite scores on the preceding variates. This necessitates some changes in signs of the weights and additive constants. There is no necessity, however, that the second and third canonical correlations have high values. One possible explanation for the combination of results obtained in the study might be that the canonical variates are rotations of some other variates representing the two types of schools and colleges, liberal arts and technical. More data on this possibility will be presented

TABLE 8

Weights of Variables, Separate Variable Form
(Canonical Correlation Model II: Within-College Covariances and Variances)

Predictor Variable Weights

	Canonical Variate 1		Canonical Variate 2		Canonical Variate 3	
School	Grade	Constant	Grade	Constant	Grade	Constant
1	2.50	−1.02	4.49	−2.51	−8.77	2.67
2	.78	−1.23	.57	−.93	2.11	−3.62
3	.56	−.68	.48	−.76	1.21	−1.67
4	.54	−.83	−.07	.14	.44	−.60
5	.82	−1.85	.95	−2.03	.01	−.99
6	1.75	−4.31	−.53	1.06	−.24	−.15
7	.73	−.92	.37	−.42	1.21	−1.70
8	.48	−.11	.97	−.32	−.42	−.04
9	.90	−.91	.74	−.99	1.92	−1.78
10	.44	−.31	.52	−.64	.58	.03
11	.52	−2.17	−.87	1.14	−.54	−.34
12	1.71	−.21	−1.29	−.94	−.83	−1.42
13	.97	−1.02	−.91	.09	−.50	−.73
14	1.75	−4.04	−1.35	2.55	−.67	.73
15	1.15	−1.10	1.27	−1.11	1.25	−1.70
16	−.09	−.10	.14	−.41	.24	−.57
17	.56	−.21	.97	−.62	−.24	.48
18	.58	−1.06	2.57	−2.83	−2.48	1.79
19	.48	−.25	1.13	−.74	−.18	.49
SAT Scores						
Verb.	.25		−.08		.12	
Math.	.20		.32		.08	
Const.	−.12		.00		−.12	

Criterion Variable Weights

	Canonical Variate 1		Canonical Variate 2		Canonical Variate 3	
College	Grade	Constant	Grade	Constant	Grade	Constant
1	.68		1.85		−1.36	
2	1.93		−1.37		−.83	
3	.16		.13		.36	
4	.59		.38		.28	
5	1.08		.72		1.67	
6	.95		.47		1.27	

later. A second possibility is that there are some remnant effects of the uneven pattern of number of students for the school-college combinations. This seems to be borne out in that college 2 is the only college to have a negative weight on the second canonical variate and the schools from which college 2 receives its students have the more negative weights for grades.

The differences in level between colleges is of interest to many people; objective evidence on the magnitude of these differences would be important to them. By working with within-college deviation composite scores, the present model appears to exclude the possibility of evidence as to differences between colleges. As a consequence, an initial preference might be given to the total covariances and variances canonical model which, explicitly, includes the differences between colleges. A possibility exists, however, for obtaining information from the within-college covariances and variances canonical model as to the differences between colleges.

For each canonical variate, a vector of predictor weights A_p is determined by this latter model. The weights for three such vectors are given in Table 8. Using this weight vector and the original predictor scattered data matrix X, the vector of predictor composite scores F_p may be determined from (12) (for a single vector of weights this equation is $XA_p = F_p$). These predictor composite scores are not deviation scores from the means of the colleges, and the means for the colleges are not zero. Table 9 presents the mean for

TABLE 9

Means of Predictor Composite Scores for Canonical Variates for Group of Students at Each College (Canonical Correlation Model II: Within-College Covariances and Variances)

Canonical Variate	College 1	2	3	4	5	6
1	−.34	−1.17	.45	.32	.34	.29
2	−1.10	.17	.30	.19	.16	.32
3	.74	−.88	.11	−.27	.18	.25

each college on each of the three first canonical variates from the within-college covariances and variances canonical model for the illustrative study. Thus, the use of within-college deviation composite scores to determine the vectors of weights does not preclude using these weights to determine composite scores that are not deviation scores from the college means.

The mean scores for the colleges on the predictor composite scores in the vectors F_p provide, in a sense, evidence concerning differences in level between the colleges. The mean predictor composite score for the students going to a college indicates the general level of these students in whatever function is reflected by the predictor composite score. For example, insofar as the first canonical variate for the present model is a measure of general academic aptitude (so interpreted because of the general positive weights for the grades), the mean composite scores given in Table 9 for each college for the first canonical variate is an indication of the general level of the students going to that college as to general academic ability. In this sense, college 2 appears to attract the least able group of students of any of the colleges in the study, and college 3 appears to attract the most able group of students of any of the colleges in the study. Only small differences occur, however, between the positive means. This type of data appears to have promise for interpretations relevant to the differences in levels of the colleges.

In order to investigate the relations of the predictor composite scores on the canonical variates for the within-college model with the college grades of the students at each college, the correlations given in Table 10 were computed. This table is analogous to Table 6 for the total covariances and variances model. Note that the correlations in the first row of Table 10 are higher than those in the first row of Table 6. The higher correlations for each college with the within-college model occur even though the first canonical correlation for this model is lower than the first canonical correlation for the total covariances and variances model. The correlations in the other rows also appear to be higher in absolute value in Table 10 than in Table 6. Thus, as anticipated, the within-college covariances and variances model provides predictor composite scores that correlated more highly with the grades at the colleges.

TABLE 10
Correlations of Predictor Composite Scores on Canonical Variates With Grades at Each College
(Canonical Correlation Model II: Within-College Covariances and Variances)

Canonical Variate	College 1	2	3	4	5	6
1	.69	.79	.25	.55	.82	.72
2	.72	−.73	.16	.33	.70	.45
3	−.59	−.66	.24	.22	.82	.71

Comparisons between the weights for the college grades as given in Table 8 and the correlations given in Table 10 indicate good agreement between these two types of coefficients. In particular, wherever there is a negative weight for a college for one of the canonical variates, the corresponding correlation is negative. These negative weights and correlations produce problems in the interpretation of the second and later canonical variates.

One possible interpretation of the second and third canonical variates for the within-college model must be discarded in light of the correlations in Table 10. These variates are not associated with differential correlations for the liberal arts colleges, 1, 2, and 5, as compared with the technical colleges, 3, 4, and 6. The relatively low correlations on all three variates for colleges 3 and 4 (especially for 3) are quite disappointing. A problem remains whether either of the canonical models can determine composite scores appropriate to the variety of schools and colleges.

Predictive Model

A reformulation of the mathematical treatment of the criterion side of the system is involved in the predictive model as compared with the treatment of the criteria in the canonical models. Instead of attempting to establish criterion composite scores which correlate highly with the predictor composite scores, as in the canonical models, an attempt will be made to predict the college grades from the predictor composite scores. This treatment appears to be a more accurate translation of an important aim of a central prediction system, that of providing predictions of students' performances in college from available information such as the school grades and test scores.

The predictor variable constructions previously described will be employed in the predictive model, including the scattered data matrix X defined in (9), the weight matrix A of (10), and the matrix of composite scores F of (11), these matrices being related as in (12). In contrast to the use of column vectors A_p and F_p in the canonical models, A and F will be considered to have several columns, the number of columns being n_p. Extensive use will be made of the basic matrix ξ^c, defined in (63), for within-college deviation composite scores. The several column vectors of within-college deviation composite scores, F_p^c, corresponding to the columns of the matrix A will be collected into a matrix F^c. For convenience, the superscript c will not be used in the following equations, being left as implied wherever appropriate. The only possibility of confusion exists between the two matrices of composite scores and this may be eliminated by using F^r to designate the matrix of *raw* composite scores of (11). With these changes, (64) becomes

$$F = \xi A. \tag{82}$$

This equation may be written in summation notation

$$\sum_j \xi_{i_c j} a_{jp} = f_{i_c p} . \tag{83}$$

The tag second subscript c is used with the index i for subsequent convenience.

On the criterion side, regression equations will be used for predicting the grades at each college from the predictor composite scores. Let the regression coefficient of predictive composite p for predicting grade category g_c at college c be w_{pg_c} . The regression equation is

$$\sum_p f_{i_c p} w_{p g_c} = \hat{y}_{i_c g_c} , \tag{84}$$

where $\hat{y}_{i_c g_c}$ is the predicted value of the observed grade $y_{i_c g_c}$. Note that the additive constant dummy variables, G_c , are not included in the criterion variables. Substitution for $f_{i_c p}$ from (83) yields

$$\sum_p \sum_j \xi_{i_c j} a_{jp} w_{p g_c} = \hat{y}_{i_c g_c} . \tag{85}$$

The error of estimate is the difference between $y_{i_c g_c}$ and $\hat{y}_{i_c g_c}$ and is denoted by $e_{i_c g_c}$.

$$e_{i_c g_c} = y_{i_c g_c} - \hat{y}_{i_c g_c} . \tag{86}$$

The coefficient E^2 is defined as the sum of squares of all errors of estimate.

$$E^2 = \sum_c \sum_{i:c} \sum_{g:c} e^2_{i_c g_c} . \tag{87}$$

Both the weights a_{jp} for determining the predictive composites and the regression weights w_{pg_c} are considered as unknown parameters in the predictive model and are to be established so that E^2 is a minimum. The number of predictive composites, n_p , is also unknown. A solution for the a_{jp} and w_{pg_c} for any fixed value of n_p is given in Appendix B. Numerical solution for any particular data involves a series of successive approximations. An unsolved mathematical problem is the uniqueness of the solution for any fixed value of n_p . However, the least minimum E^2 for each value of n_p decreases or does not increase as the value of n_p is increased. This is true since the system for any given value of n_p includes, as special cases, all systems having a smaller value of n_p . An upper bound for n_p is n_c , the number of colleges, in which case there would be a separate predictive composite for each college determined entirely by the students who go to that college. A reasonable hope is that the number of predictive composites could be much smaller than n_c . A desirable solution might be to use as small a value of n_p as possible for which there would be little further reduction in E^2 for an increased value of n_p .

The weighting of the squared errors in E^2 presents an unsolved problem, also. Mechanically, the weight of the errors for any category of grades at a college can be controlled by the scaling of the college grades. All errors for a category of grades at a college can be multiplied by a constant by multiplying the grades in this category by the constant. This will result, through the solution of the normal equations for the regression weights, in the multiplication of the regression weights by this constant, and, thus, the multiplication of the predicted grades by the constant. Since the errors of estimate by (86) are the differences between the observed grades and the predicted grades, and since both of the types of quantities were multiplied by the constant, the errors would be multiplied by the constant. Such an operation could affect the a_{jp}'s however. It may be that the data are so structured that the solution will be relatively insensitive to changes in the weights for the various categories of grades at the various colleges. This is a point that warrants further investigation.

A feature of the predictive model may be developed by translating (84) into matrix form.

$$FW = \hat{Y}, \tag{88}$$

where W is a matrix of the w_{pg_c} and Y is a matrix of the predicted grades $\hat{y}_{ig_c}$. Parallel to the development of (85), substitution is made for F from (82).

$$\xi AW = \hat{Y}. \tag{89}$$

Let Γ be a square, nonsingular matrix of order n_p and let the matrices A_r and W_r be defined so that

$$A = A_r\Gamma^{-1}, \tag{90}$$

$$W = \Gamma W_r\,. \tag{91}$$

Substitution in (89) for A and W yields

$$\xi A_r W_r = \hat{Y}, \tag{92}$$

which returns to the form of (89). Consequently, the matrices A and W may be transformed by any square matrix Γ which satisfies the restriction as to being of order n_p and being nonsingular. This indeterminacy of the weight matrices is analogous to the rotation of axes problem in factor analysis. The analogy of the transformation problem to the rotation of axes problem in factor analysis suggests a solution, that of applying the principle of simple structure. In case a transformation were possible so that the regression weights exhibited a simple structure, each predictive composite might be interpreted in terms of a limited type of category of grades or college for which this composite had nonzero regression weights. Such a result would be very interesting and could be of considerable practical importance. It would aid in the interpretation of predictive composite scores for individuals and in guiding students on choices of college and of major field of study.

The formal aspects of the foregoing analogy may be completed by identifying W' as analogous to Thurstone's (1947) factor matrix F and Γ' as analogous to Thurstone's matrix T of primary trait vectors. The rotation of the W' matrix may be investigated by any of the rotation of axes procedures and will result in Thurstone's matrix Λ. Then, by Thurstone's equation (7) in Chapter 15 (1947, p. 351)

$$\Gamma' = D\Lambda^{-1}, \tag{93}$$

where D is a diagonal matrix such that row vectors Γ' are of unit length. If the unrotated predictive composites have standard scores and are uncorrelated, the correlations between the rotated predictive composites are the entries in the matrix product $\Gamma'\Gamma$.

Results for the illustrative study are given in Tables 11–13 and Fig. 1. Only the case for two predictive composites was considered in the numerical solution and the iterations were stopped after five trials. Only small changes appeared in the weights from trial 4 to trial 5, but complete convergence was not achieved. The first trial weights in the matrix A were the first and fourth within-college covariances and variances canonical variates, these two being selected because of their representation of the liberal arts colleges in the first variate and the technical colleges in the fourth variate as judged by the college canonical weights. The weights in A were refined by the computing system derived in Appendix B and described in Appendix C. No investigation was made as to the uniqueness of the solution given. Such an investigation might involve trying several different pairs of starting vectors of weights in the matrix A and seeing whether each pair led to the same computed weights upon convergence. The pair of starting vectors used were chosen (in accordance with the author's judgment) so as probably to lead to the least minimum for two composites.

Table 11 gives the weights of the predictor variables for determining the predictive composites. Weights for the unrotated composites are given in the left half of the table, and weights for the rotated composites are given in the right half of the table. The rotated composites will be discussed in subsequent paragraphs. Table 12 gives the regression weights for predicting the grades at the six colleges: Also given in Table 12 are the multiple correlations and the additive constants for use with the raw scores on the predictive composites computed from the scattered data matrix X. Several of these multiple correlations appear to be rather high. Because of the small sample and large number of degrees of freedom utilized by the model, there may have been considerable capitalization on chance and there would be considerable shrinkage in the correlations if the weights were applied to a new sample. One gratifying result of the present solution is the multiple correlation of .49 for college 3. The correlations with the grades at this college for the composites determined by the canonical models were markedly below this value.

The additive constants for the college regression equations can be interpreted in terms of differences between the grade scales of the colleges. To accomplish this interpretation most reasonably the signs of the additive constants should be reversed. As the signs stand in Table 12, for example, a positive amount is added to the weighted sum of the predictive composite scores for college 1 and a negative amount is added to the weighted sum of the composite scores for college 5 (.66 is the difference between .40 and $-.26$). The regression weights for these two colleges are very similar. These additive constants indicate that the predicted grade at college 1 would be .66 higher than at college 5. Usual interpretation would say that .66 would have to be

TABLE 11

Weights of Predictor Variables, Separate Variable Form (Predictive Model)

School	Unrotated Predictive Composites I Grade	I Constant	II Grade	II Constant	Rotated Predictive Composites A Grade	A Constant	B Grade	B Constant
1	4.95	−1.70	23.68	−2.74	−1.87	−.86	18.17	−3.00
2	.73	−1.14	−1.72	4.28	1.18	−2.29	−.45	1.66
3	.69	−.86	.44	−.06	.54	−.81	.81	−.73
4	.43	−.74	.17	−.68	.36	−.52	.44	−.99
5	1.07	−2.29	1.74	−3.31	.54	−1.27	1.90	−3.82
6	1.80	−4.23	1.90	−3.20	1.20	−3.17	2.58	−5.30
7	.79	−.98	1.68	−1.83	.29	−.43	1.64	−1.88
8	.78	−.29	1.95	−.29	.20	−.19	1.79	−.40
9	.95	−.97	.11	.54	.88	−1.08	.83	−.45
10	.48	−.39	.03	.35	.45	−.47	.40	−.10
11	.46	−1.89	.72	.43	.24	−1.94	.80	−1.26
12	1.66	−.09	−.01	.54	1.60	−.24	1.32	.26
13	.87	−.85	−.01	.27	.84	−.89	.69	−.52
14	1.64	−3.65	−.16	.90	1.62	−3.76	1.22	−2.37
15	.60	−.96	−2.83	1.07	1.37	−1.23	−1.21	−.13
16	.01	−.19	1.72	−1.49	−.47	.23	1.04	−1.05
17	.81	−.26	.95	.84	.51	−.49	1.22	.29
18	.65	−1.15	−1.94	.76	1.17	−1.32	−.64	−.46
19	.58	−.38	.35	.88	.46	−.61	.67	.23
SAT Scores								
Verb.	.24		−.22		.29		.06	
Math.	.23		−.10		.25		.13	
Const.	−.18		−.36		−.06		−.35	

TABLE 12

Regression Systems for Predicting College Grades for Predictive Composites (Predictive Model)

College	Regression Weights for Predictive Composites: Unrotated I	Unrotated II	Rotated *A*	Rotated *B*	Additive Constant	Multiple Correlation
1	.87	−.18	.83	.09	.40	.74
2	.47	.03	.32	.20	.49	.79
3	.58	.35	.09	.62	−.35	.49
4	.46	.43	−.09	.68	.00	.69
5	.81	−.32	.92	−.10	−.26	.86
6	.78	−.02	.60	.25	−.24	.74

added to the grades at college 5 for the grades at the two colleges to be comparable. Thus, in a sense, the additive constants give the amounts that should be subtracted from the observed grades at the colleges to obtain comparable grades from college to college.

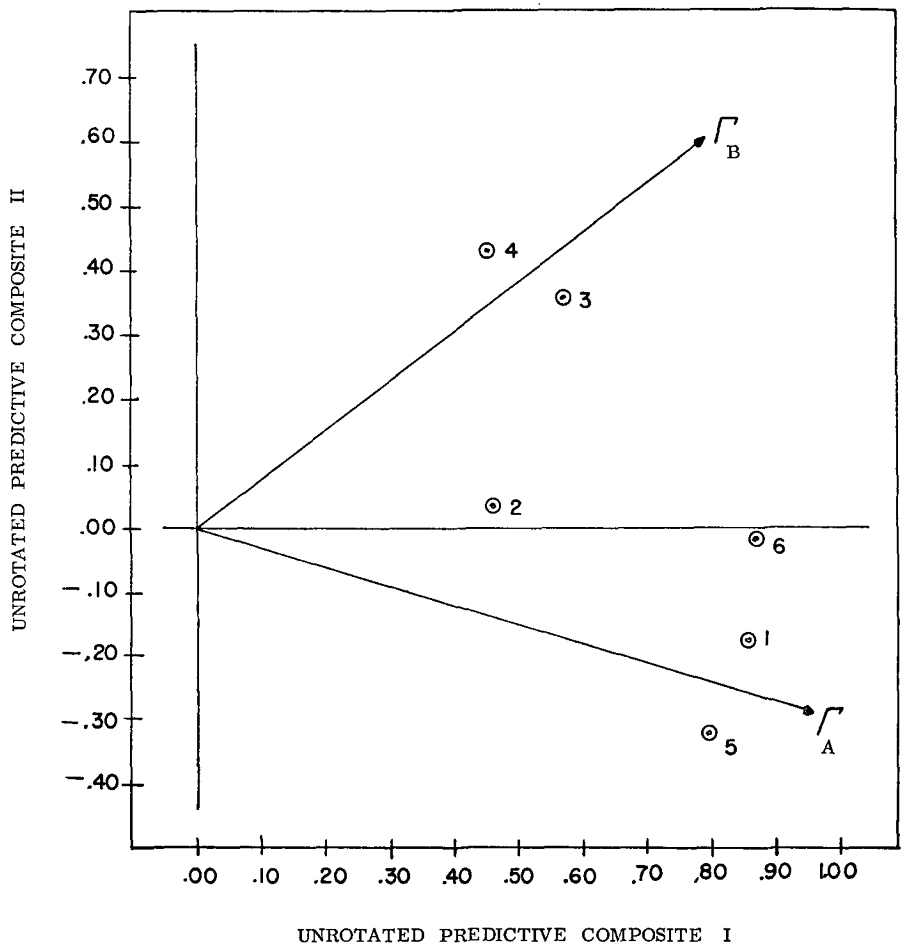

FIGURE 1. Plot of College Regression Weights on Unrotated Predictive Composites

A graph as for rotation of axes was made for the unrotated regression weights at the left of Table 12. This graph is presented in Fig. 1. The abscissa represents the regression weights on the first unrotated composite, and the ordinate represents the regression weights on the second unrotated composite. The regression weights of each college are represented by a numbered point, the number being that of the college. The configuration of points has an interesting structure in that the points for the two technical colleges, 3 and 4, separate from the points for the definitely liberal arts colleges, 1 and 5. It is of interest that the point, 6, for the students at the same college as 5 but who plan to enter engineering tends toward the technical schools. The location of the point for college 2 suggests that a number of the students at this university are concentrating in science and technical areas.

The two vectors drawn in Fig. 1 and labeled Γ_A and Γ_B were chosen so that one rotated composite would be effective for the liberal arts colleges and the other rotated composite would be effective for the technical colleges. These vectors are of unit length and the coordinates of their termini are recorded in the appropriate columns of the matrix Γ given at the left in Table 13. The inverse of this matrix is given at the right of Table 13. The

TABLE 13

Matrices for Rotating Predictive Composites
(Predictive Model)

Matrix Γ			Matrix Γ^{-1}		
Unrotated Predictive Composites	Rotated Predictive Composites A	B	Rotated Predictive Composites	Unrotated Predictive Composites I	II
I	.96	.80	*A*	.75	−1.00
II	−.28	.60	*B*	.35	1.20

Correlation Between Rotated Predictive Composites = .60

correlation between the two rotated composites is .60 as found by obtaining the sum of products between the two columns of Γ. This is also the cosine of the angle between the two Γ vectors in Fig. 1. The matrix A of predictor variable weights for the unrotated predictive composites at the left in Table 11 was postmultiplied by the matrix Γ as per equation (90) to obtain the matrix A_r of rotated weights at the right in Table 11. The matrix W of unrotated regression weights in Table 12 was premultiplied by Γ^{-1} as per (91) to obtain the matrix W_r of rotated regression weights. This transformation had no effect on the additive constants nor on the multiple correlations given in Table 12.

It is of interest that most of the predictor variable weights given in Table 11 for the rotated predictive composites are positive for the school grades and negative for the school additive constants. These weights, however, probably are somewhat unstable due to the small sample of students used in the study. Very little confidence should be placed, in particular, on the weights for school 1 since the study included only 6 students from this school, as given in Table 1. The sample size of 8 students for school 15 also is very small. Before this system is used in practice, the weights should be determined in a study using many more cases than was used in the present study.

The weights for the test scores for the rotated predictive composites are small, but positive, while the additive constants for the test are negative.

The regression weights in Table 12 for the rotated predictive composites are as anticipated from the graph of Fig. 1. Colleges 1 and 5 have high weights for composite A and near zero weights for composite B, while colleges 3 and 4 have high weights for composite B and near zero weights for composite A. From these results composite A might be called a liberal arts college predictor, and composite B might be called a technical college predictor. These interpretations could be very helpful in interpreting scores of individual students on the two predictive composites. While each college could set up its own regression equation for predicting grades at that college from the predictive composite scores, many colleges could use the scores, directly, on one composite or the other according to the type of college or for each student according to the field of interest of the student. Scores on the predictive composites might furnish the secondary school counselor with valuable information with which to guide students as to choice of major field and of colleges to which to apply.

Discussion

In the preceding sections a mathematical construction has been presented within which various quantitative models for a central prediction system could be investigated. Three such models have been discussed: canonical correlation model I for total convariances and variances, canonical correlation model II for within-college covariances and variances, and a predictive model. Properties of these models were illustrated with results from a small study.

The varying numbers of students for the various school-college combinations produced considerable effects in the canonical correlation models, especially in model I for the total variances and covariances. No such effect was observed for the predictive model.

While previous formulations of possible solutions for developing a universal grading scale have been limited to univariate models, the models discussed in this report are not so limited but may involve several adjusted grades or composite scores. The possible extension of the "internal method" by Bloom and Peters (1961) to several scales of adjusted grades was pointed out. These multiple solutions could have considerable importance, both in the general consideration of the problems involved in school and college grades and in the counseling of students by secondary schools and colleges as well as in the admission procedures of the colleges.

Of the three models discussed, the predictive model appears to provide the soundest basis for practical actions, both by the student counselors and the college admission officers. Both of these groups are interested in predictions of the performances of students in case given actions are taken. The predictive model is formulated to provide such predictions. Both canonical models are formulated to obtain pairs of variates, one from the predictor variables and one from the criterion variables, which correlated maximally. Prediction of student performances on specific criteria would be a secondary step and might not be accomplished effectively. For the predictive model, the predictor composites are derived so as to predict most effectively student performances in the specific criteria.

In future applications of the models for a central prediction system, serious consideration should be given to the possibility of using more categories of grades than a single average grade for each student in secondary school and, possibly, a single average grade for each student in college. Such differentiation of the input data coupled with the multivariate capabilities of the models could lead to substantial differentiations between several lines of study and furnish improved bases for practical actions such as are involved in counseling and admission.

Both extensive empirical studies and conceptual considerations of the models are indicated for future developments. Knowledge is needed on the stability of parameters of the models when estimated from various samples of individuals, schools, and colleges. This knowledge might be developed in part by mathematical statistical techniques and in part by empirical investigations. The latter would be required especially for changes that are associated in time with changes in the educational systems. It may be easier to obtain knowledge on stability under random sampling with empirical studies, both with actual data and with data derived from random numbers. There is a need, also, for the development of more effective computing systems. Knowledge is needed, too, on the useful dimensionality of the predictive situation, especially for the predictive model. This knowledge can come only from extensive empirical studies.

APPENDIX A

Solution for Canonical Correlation Models

The following procedure is appropriate for determining the canonical correlations and associated weights when the data matrices have ranks less than their orders. Equations (A.1) and (A.2) present the definitions of the column vectors of deviation composite scores F_p and Z_p as dependent on the basis matrices ξ and ζ, and on the column vectors of weights A_p and B_p. A slight shift in notation from that used in the body of the monograph has been introduced by the dropping of the superscripts to the matrices ξ and ζ as well as to the vectors of composite scores F_p and Z_p. The solution given here pertains to either the deviation composite scores from the means for all individuals or the within-college deviation composite scores and the appropriate superscripts may be added to particularize the material in this appendix to one or the other types of scores. If the superscript t were used with ξ, F_p, ζ, and Z_p, (A.1) and (A.2) would become (50) and (75) of the body of the monograph; if the superscript c were used with these matrices, (A.1) and (A.2) would become (64) and (79).

$$\xi A_p = F_p , \tag{A.1}$$

$$\zeta B_p = Z_p . \tag{A.2}$$

In the canonical correlation model, the weight vectors are established so as to maximize the correlation r_p between the composite score vectors F_p and Z_p. A series of solutions are found for which the F_p and the Z_p for each solution are uncorrelated with the composite score vectors for the preceding solutions. These solutions can be arranged in descending order of the correlation r_p. The usual method of solution (see Kendall, 1951, Vol. 2, pp. 348–358) involves the inverses of the intercorrelation matrices from the basis matrices. As noted from (54) and (70), the basis matrices ξ and ζ have ranks less than their orders; consequently, the matrices of intercorrelations computed from these bases would be singular and would not possess inverses. The following procedure is appropriate for this case.

If all entries in A_p were multiplied by a constant, then all entries in F_p given in (A.1) would be multiplied by the same constant. The correlation r_p between F_p and Z_p would not be affected, however, by this rescaling of the weights A_p and composite scores F_p. Similarly, the weights B_p and composite scores Z_p may be rescaled without affecting the correlation r_p. It is possible, and desirable, to define the scale for the weights such that the sum of squares of the composite scores over individuals equals the number of individuals. Thus,

$$F_p'F_p = Z_p'Z_p = N_{..} . \tag{A.3}$$

Since the bases ξ and ζ were established so that the composite scores F_p and Z_p would be deviation scores, the correlation between them is

$$r_p = \frac{1}{N_{..}} F_p'Z_p . \tag{A.4}$$

The regression equation for predicting Z_p from F_p is

$$F_p r_p = \hat{Z}_p , \tag{A.5}$$

where $\hat{Z}_p$ is the predicted Z_p. Similarly, the regression equation for predicting F_p from Z_p is

$$Z_p r_p = \hat{F}_p , \tag{A.6}$$

where $\hat{F}_p$ is the predicted F_p.

Substituting for F_p in (A.5) from (A.1)

$$\xi A_p r_p = \hat{Z}_p \ , \tag{A.7}$$

and substituting for Z_p in (A.6) from (A.2)

$$\zeta B_p r_p = \hat{F}_p \ . \tag{A.8}$$

Consider, temporarily, that the weights B_p and the composite scores Z_p are known. The $\hat{Z}_p$ would be the multiple regression predictions of Z_p from the basis matrix ξ with the regression weights being the entries in $A_p r_p$. The correlation r_p would be the multiple correlation. This relation to the multiple regression model occurs in order for r_p to be a maximum and provides a solution for the weights A_p. The multiple regression normal equations may be written as

$$(\xi'\xi)(A_p r_p) = (\xi' Z_p). \tag{A.9}$$

Substitution for Z_p from (A.2) and rearranging parentheses yields

$$(\xi'\xi)A_p r_p = (\xi'\zeta)B_p \ . \tag{A.10}$$

Similar reasoning considering the weights A_p and the composite scores F_p as known yields

$$(\zeta'\zeta)(B_p r_p) = (\zeta' F_p), \tag{A.11}$$

$$(\zeta'\zeta)B_p r_p = (\zeta'\xi)A_p \ . \tag{A.12}$$

Equations (A.10) and (A.12) provide two sets of equations in the weights A_p and B_p and are to be solved simultaneously for the weights.

The usual solution involves the inverses of the matrices $(\xi'\xi)$ and $(\zeta'\zeta)$. In the present case these inverses do not exist, and another mechanism is required to obtain the desired solution. Let the matrix $(\xi'\xi)$ be factored to obtain a matrix G such that

$$(\xi'\xi) = GG', \tag{A.13}$$

and

$$|G'G| \neq 0. \tag{A.14}$$

There are several methods for accomplishing this step, one of which is to obtain the characteristic roots and vectors of $\xi'\xi$. The zero roots are discarded so that

$$(\xi'\xi) = WDW', \tag{A.15}$$

where D is a diagonal matrix containing the nonzero characteristic roots and W is a portion of an orthogonal matrix containing, as columns, the corresponding characteristic vectors. Thus

$$|D| \neq 0, \tag{A.16}$$

(A.17) $$W'W = I.$$

Then

(A.18) $$G = WD^{1/2},$$

(A.19) $$GG' = WD^{1/2}D^{1/2}W' = WDW' = (\xi'\xi),$$

(A.20) $$G'G = D^{1/2}W'WD^{1/2} = D.$$

The matrix G as defined in (A.18) satisfies (A.13) as shown in (A.19) and satisfies (A.14) as shown in (A.20) and (A.16).

In a similar manner, the matrix $(\zeta'\zeta)$ is factored to obtain the matrix H.

(A.21) $$(\zeta'\zeta) = HH',$$

(A.22) $$|H'H| \neq 0.$$

The problem when the rank of a matrix is not equal to its order, such as is true for ξ and ζ, is that the vectors of the matrix do not span the entire space possible as implied by the dimensions of the matrix. The vectors A_p and B_p may be restricted to the spaces spanned by the column vectors of ξ and ζ by defining these weight matrices as follows.

(A.23) $$A_p = G(G'G)^{-1}U_p ,$$

(A.24) $$B_p = H(H'H)^{-1}V_p ,$$

where U_p is a column vector of order equal to the rank of ξ, and V_p is a column vector of order equal to the rank of ζ. The only restrictions on these vectors are that their lengths be such as to satisfy (A.3).

Substitution in (A.10) for $(\xi'\xi)$ from (A.13), for A_p from (A.23), and for B_p from (A.24) yields

(A.25) $$GG'G(G'G)^{-1}U_pr_p = (\xi'\zeta)H(H'H)^{-1}V_p ,$$

(A.26) $$GU_pr_p = (\xi'\zeta)H(H'H)^{-1}V_p .$$

Premultiplication of (A.26) by $(G'G)^{-1}G'$ yields

(A.27) $$U_pr_p = (G'G)^{-1}G'(\xi'\zeta)H(H'H)^{-1}V_p .$$

Similar substitutions into (A.12) from (A.21), (A.23), and (A.24) and manipulations as the preceding yield

(A.28) $$V_pr_p = (H'H)^{-1}H'(\zeta'\xi)G(G'G)^{-1}U_p .$$

Define a matrix M, not to be confused with the M of (29), as

(A.29) $$M = (G'G)^{-1}G'(\xi'\zeta)H(H'H)^{-1},$$

and substitute into (A.27) and (A.28),

(A.30) $$U_pr_p = MV_p ,$$

(A.31) $$V_p r_p = M' U_p \, .$$

Multiplication of (A.31) by r_p and substitution for $U_p r_p$ from (A.30) yields

(A.32) $$(M'M) V_p = V_p r_p^2 \, .$$

This equation is in standard form for V_p to be a characteristic vector of the matrix $(M'M)$ and r_p^2 to be the corresponding characteristic root. As for the usual canonical correlation solution there are multiple solutions for the present case. The number of nonzero canonical correlations cannot exceed the dimensionality of either of the basis matrices ξ or ζ.

APPENDIX B
Solution for Predictive Model

The basic equation for the solution for the predictive model is obtained by substituting for e_{icg_c} in (87) from (86).

(B.1) $$E^2 = \sum_c \sum_{i:c} \sum_{g:c} (y_{icg_c} - \sum_p \sum_j \xi_{icj} a_{jp} w_{pg_c})^2 .$$

Conditions for a minimum solution are obtained by setting the partial derivatives of E^2 with respect to the parameters a_{jp} and w_{pg_c} equal to zero.

Consider, first, the partial derivatives of E^2 with respect to w_{pg_c} .

(B.2) $$\frac{\partial E^2}{\partial w_{pg_c}} = -2 \sum_{i:c} (y_{icg_c} - \sum_q \sum_k \xi_{ick} a_{kp} w_{qg_c}) \sum_j \xi_{icj} a_{jp}$$
$$= 0,$$

where k is used as an alternate subscript to j to designate predictor variable and q is used as an alternate subscript to p to designate predictive composite. Algebraic manipulation of (B.2) and substitution from (83) yields

(B.3) $$\sum_{i:c} y_{icg_c} f_{icp} = \sum_q w_{qg_c} \sum_{i:c} f_{icq} f_{icp} \, .$$

Equation (B.3) is linear in the w_{qg_c} and may be solved when the a_{jp}'s are known, for each college, separately. Let the sections of the matrices F and Y for the students at college c be designated by F_c and Y_c . Equation (B.3) becomes, in matrix form,

(B.4) $$F_c' Y_c = F_c' F_c W_c \, ,$$

where W_c contains the w_{qg_c} for college c. Since the matrix product $F_c'F_c$ is obtained from empirical data, it probably will be nonsingular and possess an inverse. Then,

(B.5) $$W_c = (F_c'F_c)^{-1} F_c' Y_c \, .$$

A different equation obtained from (B.2) by algebraic manipulation will be of interest in a subsequent section and is given here for convenience.

(B.6) $$\sum_j a_{jp} \sum_{i:c} y_{i_c g_c} \xi_{i_c j} = \sum_j a_{jp} \sum_q w_{qg_c} \sum_k a_{kq} \sum_{i:c} \xi_{i_c k} \xi_{i_c j} .$$

Define the quantities $(\theta_{g_c j})_c$ and $(\phi_{kj})_c$ by

(B.7) $$(\theta_{g_c j})_c = \sum_{i:c} y_{i_c g_c} \xi_{i_c j} ,$$

(B.8) $$(\phi_{kj})_c = \sum_{i:c} \xi_{i_c k} \xi_{i_c j} .$$

Equation (B.6) becomes, then,

(B.9) $$\sum_j a_{jp} (\theta_{g_c j})_c = \sum_j a_{jp} \sum_q w_{qg_c} \sum_k a_{kq} (\phi_{kj})_c .$$

Consider the partial derivatives of E^2 with respect to a_{jp} .

(B.10) $$\frac{\partial E^2}{\partial a_{jp}} = -2 \sum_c \sum_{i:c} \sum_{g:c} (y_{i_c g_c} - \sum_q \sum_k \xi_{i_c k} a_{kq} w_{qg_c}) \xi_{i_c j} w_{pg_c} = 0.$$

Algebraic manipulation of (B.10) and substitutions from (B.7) and (B.8) yield

(B.11) $$\sum_k \sum_q a_{kq} \sum_c \sum_{g:c} w_{qg_c} w_{pg_c} (\phi_{kj})_c = \sum_c \sum_{g:c} w_{pg_c} (\theta_{g_c j})_c .$$

Equation (B.11) represents a group of simultaneous, linear equations in the weights a_{kq} and might be solved for these weights. Thus, (B.5) and (B.11) form a base for a possible successive approximation method for solving for the two types of weights, a_{jp} and w_{pg_c} . A trial group of weights w_{pg_c} might be entered into (B.11) and the weights a_{kq} determined. Use of these weights to determine trial scores on the predictive composites would determine the matrix F which could be entered into (B.5) to obtain a new group of weights w_{pg_c} . Such cycles of computations could be continued until convergence was obtained.

The number of unknowns and of simultaneous linear equations represented by (B.11) is $n_j \times n_p$ which might become prohibitively large for an extensive study. This product, 41×2, is 82 for the small illustrative study. A second procedure is possible that does not involve such large matrices. Algebraic manipulation of (B.11) yields

(B.12) $$\sum_c \sum_{g:c} w_{pg_c} (\theta_{g_c j})_c = \sum_c \sum_{g:c} w_{pg_c} \sum_q w_{qg_c} \sum_k a_{kq} (\phi_{kj})_c .$$

Note the similarity between (B.9) and (B.12).

While the quantities $(\phi_{kj})_c$ of (B.8) are defined for every college, the quantities $(\theta_{g_c j})_c$ of (B.7) are defined for each g_c for the college c, the college for which g_c is a category of grades. Let the quantities $(\hat{\theta}_{g_c j})_d$ be defined as in (B.13) when $d \neq c$.

(B.13) $$(\hat{\theta}_{gcj})_d = \sum_q w_{qgc} \sum_k a_{kq}(\phi_{kj})_d \qquad \text{(for } d \neq c\text{)}.$$

An equation similar to (B.9) may be written by multiplying (B.13) by a_{jp} and summing over j.

(B.14) $$\sum_j a_{jp}(\hat{\theta}_{gcj})_d = \sum_j a_{jp} \sum_q w_{qgc} \sum_k a_{kq}(\phi_{kj})_d \qquad \text{(for } d \neq c\text{)}.$$

Similarly, an equation similar to (B.12) may be written

(B.15) $$\sum_c \sum_{g:c} w_{pgc}(\hat{\theta}_{gcj})_d = \sum_c \sum_{g:c} w_{pgc} \sum_q w_{qgc} \sum_k a_{kq}(\phi_{kj})_d \qquad \text{(for } d \neq c\text{)}.$$

Define:

(B.16) $$\hat{\theta}_{gcj} = \sum_d (\hat{\theta}_{gcj})_d + (\theta_{gcj})_c \qquad \text{(for } d \neq c\text{)},$$

(B.17) $$\phi_{kj} = \sum_d (\phi_{kj})_d + (\phi_{kj})_c \qquad \text{(for } d \neq c\text{)}.$$

Summing (B.14) over the colleges d, adding (B.9), and substituting from (B.16) and (B.17) yields

(B.18) $$\sum_j a_{jp}\hat{\theta}_{gcj} = \sum_j a_{jp} \sum_q w_{qgc} \sum_k a_{kq}\phi_{kj} \,.$$

Similarly, summing (B.15) over d, adding (B.12), and substituting from (B.16) and (B.17) yields

(B.19) $$\sum_c \sum_{g:c} w_{pgc}\hat{\theta}_{gcj} = \sum_c \sum_{g:c} w_{pgc} \sum_q w_{qgc} \sum_k a_{kq}\phi_{kj} \,.$$

By letting $\hat{\theta}$ and ϕ be matrices containing as elements $\hat{\theta}_{gcj}$ and ϕ_{kj}, respectively, (B.18) and (B.19) can be written in matrix form.

(B.20) $$\hat{\theta}A = W'A'\phi A,$$

(B.21) $$W\hat{\theta} = WW'A'\phi.$$

Because of the freedom of transformation outlined in equations (89) through (92), the matrices may be defined, for the present purposes, so that

(B.22) $$A'\phi A = N_{..}I,$$

(B.23) $$WW' = N_{..}^{-1}\Lambda,$$

where Λ is an unknown diagonal matrix. Equations (B.20) and (B.21) become

(B.24) $$\hat{\theta}A = W'N_{..}\,,$$

(B.25) $$W\hat{\theta} = N_{..}^{-1}\Lambda A'\phi.$$

From the definition of the elements of ϕ, the ϕ_{kj}, in (B.17) and of $(\phi_{kj})_d$ in (B.8), the matrix ϕ may be identified with the matrix product $\xi'\xi$ and a series of steps identical with those of equations (A.13)–(A.20) and (A.23) may be performed.

(B.26) $$\phi = GG',$$

(B.27) $$|G'G| \neq 0,$$

(B.28) $$A = G(G'G)^{-1}U,$$

where U is a matrix such as to satisfy (B.28) and have as many rows as the rank of ϕ and as many columns as does the matrix A. Substitutions from (B.26) and (B.28) into (B.24) and (B.25) yield

(B.29) $$\hat{\theta}G(G'G)^{-1}U = W'N_{..}\,,$$

(B.30) $$W\hat{\theta} = N_{..}^{-1}\Lambda U'G'.$$

Postmultiplication of (B.30) by $G(G'G)^{-1}$ gives

(B.31) $$W\hat{\theta}G(G'G)^{-1} = N_{..}^{-1}\Lambda U'.$$

Substitution from (B.31) into (B.29) gives

(B.32) $$\hat{\theta}G(G'G)^{-2}G'\hat{\theta}'W' = W'\Lambda.$$

This equation is in standard form for the diagonal elements of Λ to be the characteristic roots of the matrix $\hat{\theta}G(G'G)^{-2}G'\hat{\theta}'$ and the columns of W' to be the corresponding characteristic vectors. The n_p largest characteristic roots and vectors are to be used. Note that the columns of W' are scaled in accordance with (B.23). The matrix A can be found by substituting from (B.31) into (B.28).

(B.33) $$A = G(G'G)^{-2}G'\hat{\theta}'W'\Lambda^{-1}N_{..}\,.$$

The crux of the iterative procedure based on (B.32) and (B.33) is the determination of the elements $(\hat{\theta}_{gci})_d$ for $c \neq d$ defined in (B.13) and the subsequent elements $\hat{\theta}_{gci}$ defined in (B.16). One possible iterative procedure is to start with some estimates of the weights in the W and A matrices, use these in obtaining estimates of the $(\hat{\theta}_{gci})_d$ for $c \neq d$ in (B.13), and to continue through to a solution for new matrices of weights W and A from (B.32) and (B.33). The new weights could be used to start a new cycle. These cycles could be repeated until convergence is obtained.

A variant to the foregoing procedure is to start with an estimate of the matrix of predictor variable weights in matrix A, determine the corresponding scores in matrix F for the predictive composites by (81), determine the entries in regression weights matrix W by application of (B.5) for each college, estimate the $(\hat{\theta}_{gci})_d$ for $c \neq d$ by (B.13), and to continue through to determination of a new matrix A of predictor variable weights from (B.33). This new matrix A could be used to start a new cycle. These cycles could be continued until convergence is obtained. In the second procedure, determination of the predictive composite scores for the individuals could be avoided by working with matrices ϕ_c of the $(\phi_{kj})_c$ and θ_c of the $(\theta_{gcj})_c$, there being one of each of these matrices for each college. From (B.7), (B.8), and (83):

(B.34) $$F_c'F_c = A'\phi_c A,$$

(B.35) $$F_c'Y_c = A'\theta_c .$$

An advantage of the second procedure is the precise determination of the regression weights in matrix W for the given matrix A of predictor variable weights so as to satisfy the partial derivative (B.2). The matrix W used at the beginning of the cycles of the first procedure are determined toward the end of the preceding trial and involve the previous estimates of the $(\hat{\theta}_{gci})_d$. As a consequence of the determination of new weights in W from (B.5) the second procedure should converge in fewer trials.

Establishment of a first trial matrix A of predictor variable weights poses a problem. In the illustrative problem the experimenter used his judgment in selecting the A_p vectors for the first and fourth within-college canonical variates. The B_p vectors for the first canonical variate gave high weights to the liberal arts college, and the B_p vector for the fourth canonical variate gave high weights to the technical colleges. An alternative procedure might have been to establish a trial matrix W with one row having high weights for the liberal arts colleges and a second row having high weights for the technical colleges. A corresponding matrix A might have been obtained by solving the simultaneous equations implied by (B.11). In other studies, the colleges might be classified into groups in a similar manner dependent on the judgment of the experimenter and corresponding matrices W might be established. Dependence of the solution on the initial groupings of the colleges could be investigated by using several different initial groupings for several solutions. If the several different starting groups led to the same final solution, this would yield strong evidence for the uniqueness of the solution.

Problems that remain to be discussed are the interpretations of the characteristic roots of the matrix

$$\hat{\theta}G(G'G)^{-2}G'\hat{\theta}'$$

and the convergence of the proposed procedures. Consider that a matrix Y were available with each student having grades in every category of grades at every college. Equation (B.1) could be rewritten without restricting the individuals to those attending college c. Setting partial derivatives with respect to the weights a_{jp} and w_{pgc} equal to zero would lead directly to (B.18) and (B.19), and (B.32) and (B.33) would yield the solutions for the two types of weights. For this case, the characteristic roots of the matrix

$$\hat{\theta}G(G'G)^{-2}G'\hat{\theta}'$$

are an analysis of the variance of the predicted criterion scores into orthogonal components such that the contribution of a predictive composite to the variance of the predicted criterion scores is the corresponding characteristic root. This statement is verified by writing the formula for the variance of the predicted scores $\hat{y}_{ig_c}$ of (84) and using the restrictions of (B.22) and (B.23). The variance of the predicted criterion scores, then, is the sum of the characteristic roots for the predictive composites used. The error variance of prediction is the difference between the variance of the original criterion scores and the variance of the predicted criterion scores. As many predictive composites could be used as there were substantially large characteristic roots. The predictive composites with near-zero roots could be ignored.

The definition of the elements $(\hat{\theta}_{gci})_d$ in (B.13) corresponds in essence to defining the grade for each individual in each category of grades at colleges he did not attend as the predicted grade for him in each of these categories of grades. The approximations of the $(\hat{\theta}_{gci})_d$ in the two iterative procedures correspond to approximating the predicted grades. The characteristic roots are the contributions of the new predictive composites in predicting the criterion score matrix including these approximations. The convergence of the procedures can be argued as due to a reduction in the error of estimate variance at every step. Given a matrix of criterion scores including approximations for grades at colleges not attended by students, the solution for A and W by (B.32) and (B.33) yields a minimum error of estimate variance for any given number of predictive composites. Replacement of the approximations to grades at colleges not attended by each student so that these approximations conform to the just obtained predictive composites reduces the corresponding errors of estimate and the error variance of estimate. Re-solution for the predictive composites reduces the error variance further until a minimum is obtained. At this time, the existing errors of estimate are associated with the observed college grades and not with the inserted estimates of grades at colleges the students did not attend.

As a consequence of the preceding propositions, the characteristic roots of trial matrices

$$\hat{\theta}G(G'G)^{-2}G'\hat{\theta}'$$

give some evidence as to the propriety of the number of predictive composites being used. If the number of sizeable roots is less than the number of predictive composites, the number of predictive composites could be reduced without materially increasing the error of estimate variance. If the number of sizeable roots is greater than the number of predictive composites, one or more additional predictive composites are indicated. These can be obtained by using more characteristic roots and vectors from (B.32). As the trials progress, the roots corresponding to predictive composites being used should increase in size and the other roots should reduce in size. One hope is that these other roots would approach zero.

When a solution for the weights in A is obtained, a transformed matrix of weights that yield the deviation composite scores from the total means, when applied to the original scattered data matrix X, can be obtained from

(44). Applications of these transformed weights to the column sums of the scattered data matrix X for the students at each college yields the sums of the predictive composite scores for these students. The additive constants for the regression equations for each college can be obtained in the usual manner from the sums of the predictive composite scores, the sums of the grades, and the number of students in the study at the college in combination with the regression weights for that college in the matrix W. The square of the multiple correlation for predicting the grades in each category of grades at each college is obtained by dividing the value of $(W_{g_c}F_c'F_cW_{g_c})$ by the variance of the grades in the category, where W_{g_c} is a column vector of regression weights for category g_c of grades.

APPENDIX C

Computing Notes

These notes on computing procedures for the several models will be organized around the computations for the illustrative study. The form of most of the steps should not change for larger studies; however, increased allowances will be necessary for larger computed numbers, more schools, more colleges, more tests, and more grades for each school and college as appropriate. Insofar as these extensions involve only increased capacity for data and can be planned in detail for particular studies, only the forms of the extensions will be indicated for those few steps where the nature of the extensions are not obvious.

For a majority of the steps for the illustrative study an IBM 650 was found to be effective. In only two steps, the determination of characteristic roots and vectors for two matrices, each 41 × 41, was a more powerful machine required. For a more extensive study, however, a more powerful machine would be appropriate for much of the analysis.

In the present case, four special programs were written in addition to the several library programs already available for matrix manipulation. Two of these programs concerned the original summarization of the data into forms utilized in the analysis for the three models. The two other programs were written for the predictive model. Solutions for the two canonical correlation models used library matrix manipulation programs. The following notes are organized into a section on the original summarization, a section on the computations for the canonical correlation models, and a section on the computations for the predictive model.

Original Summarization of Data

A punched card containing the following information was prepared for each individual. All variables were scaled so that every grade and score was positive.

Individual Number:	i,
School Code:	s,
College Code:	c,
School Grade:	x_{isci_s} ,
Test Scores* (x_{isci_t})	
Verbal:	x_{iscv} ,
Mathematical:	x_{iscm} ,
Dummy Variable:	x_{iscJ_t} ,
College Grade:	y_{iscg_c} .

Two dummy variables, each having values of unity for each individual, were left implied, one will be designated x_{iscJ_s} for the school and the other will be designated y_{iscG_c} for the college.

In extending this format for the data for a more extensive study, additional card fields may be required for school grades, tests and test scores, and college grades. One field was used for each variable listed above for the illustrative study.

Assignment of index numbers during calculations involved, for the predictor variables, giving numbers 1–19 to the dummy variable for the schools, numbers 20–38 to the grades at the schools, numbers 39 and 40 to the test scores, and 41 to the test dummy variable; for the criterion variables, numbers 1–6 were given to the grades at the colleges and numbers 7–12 to the college dummy variables.

Computer program 1 was used to obtain the sums of grades and scores, sums of squares, and sums of cross-products for each college, separately. The data cards were sorted by college and, within college, by school. The resulting values were output on cards separately for each college as outlined below.

1. Two cards resulted for each school for the college, each card being identified as to college, school, and card number. The information contained in each card is shown in Table 14. Note that the last entry for each card is identical with the first entry for that card due to the identity of the school and college dummy variables. In practice, the last entry was not punched into the cards, but the data were picked up from the first entry.

In more extensive studies than the present one, the inclusion of more school grades per school would increase the number of cards similar to Card 2, one card for each category of grades. Further, the number of entries on each card similar to the second entry would increase, correspondingly. Extensions in the number of test variables and college grade variables would increase the number of entries on each card.

2. Seven cards were produced for each test variable x_{iscv} , x_{iscm} , and

*If observed scores were not available, zeros were punched into the card in the test variable fields, including the dummy variable. When test scores were available, they were punched into the score fields and unity was punched into the field for the dummy variable.

TABLE 14

Information Contained in Each Card

Card 1	Card 2
$\sum_{i:sc} x^2_{i_s c J_s} = N_{sc}$	$\sum_{i:sc} x_{i_s c i_s} x_{i_s c J_s}$
$\sum_{i:sc} x_{i_s c J_s} x_{i_s c i_s} = \sum_{i:sc} x_{i_s c i_s}$	$\sum_{i:sc} x^2_{i_s c i_s}$
$\sum_{i:sc} x_{i_s c J_s} x_{i_s c v} = \sum_{i:sc} x_{i_s c v}$	$\sum_{i:sc} x_{i_s c i_s} x_{i_s c v}$
$\sum_{i:sc} x_{i_s c J_s} x_{i_s c m} = \sum_{i:sc} x_{i_s c m}$	$\sum_{i:sc} x_{i_s c i_s} x_{i_s c m}$
$\sum_{i:sc} x_{i_s c J_s} x_{i_s c J_t} = \sum_{i:sc} x_{i_s c J_t}$	$\sum_{i:sc} x_{i_s c i_s} x_{i_s c J_t}$
$\sum_{i:sc} x_{i_s c J_s} y_{i_s c g_c} = \sum_{i:sc} y_{i_s c g_c}$	$\sum_{i:sc} x_{i_s c i_s} y_{i_s c g_c}$
$\sum_{i:sc} x_{i_s c J_s} y_{i_s c G_c} = N_{sc}$	$\sum_{i:sc} x_{i_s c i_s} y_{i_s c G_c}$

$x_{i_s c J_t}$; for the college grade variable $y_{i_s c g_c}$; and for the college dummy variable $y_{i_s c G_c}$. Let $X_{i_s c}$ designate any of the five preceding variables. The first three cards for each variable contained the $\sum_{i:sc} X_{i_s c}$ for the several schools. Note that the entries for the college dummy variable $y_{i_s c G_c}$ are the N_{sc}. The next three cards for each variable contained the $\sum_{i:sc} X_{i_s c} x_{i_s c i_s}$ for the several schools. Note that the entries for the college dummy variable $y_{i_s c G_c}$ are the $\sum_{i:sc} x_{i_s c i_s}$. The seventh card for each variable contained $\sum_s \sum_{i:sc} X_{i_s c} Y_{i_s c}$, where $Y_{i_s c}$ is used as an alternative designation to $X_{i_s c}$ for the five variables, and for each $X_{i_s c}$, $Y_{i_s c}$ ranges over the five variables. Note that when $X_{i_s c}$ represents the college dummy variable, $y_{i_s c G_c}$,

$$\sum_s \sum_{i:sc} X_{i_s c} Y_{i_s c} = \sum_s \sum_{i:sc} y_{i_s c G_c} Y_{i_s c} = \sum_s \sum_{i:sc} Y_{i_s c} .$$

Also note that when both $X_{i_s c}$ and $Y_{i_s c}$ represent the college dummy variable

$$\sum_s \sum_{i:sc} X_{i_s c} Y_{i_s c} = \sum_s \sum_{i:sc} y^2_{i_s c G_c} = N_{.c} .$$

Increases in the size of studies would involve systematic increases in the number of cards of each type indicated in the preceding paragraph.

Computer program 2 developed the entries in the matrices $\xi'\xi$, $\zeta'\xi$, and $\zeta'\zeta$ separately for the basis matrices ξ^t and ζ^t for deviation composite scores from the means for all individuals and for the basis matrices ξ^c and ζ^c for the within-college deviation composite scores. Let the entries in $\xi'\xi$ be designated by ϕ_{kj} in conformity to the definitions in (B.8) and (B.17); let the entries in $\zeta'\xi$ be designated θ_{gj} in conformity to the definitions in (B.7) and (B.16); and, by similarity, let ψ_{hg} designate the entries in $\zeta'\zeta$. The superscripts t and c may be attached to these symbols to distinguish between coefficients for the deviation composite scores from the means for all individuals and coefficients for the within-college deviation composite scores.

In the first step of program 2, the cards for the college dummy variable were used as a source of the sums for each college of the scores on the variables. The following coefficients were computed, saved in the machine memory, and punched into output cards. The sums were also saved in the machine memory.

1. For each college:

$$N_{.c}^{-1} \sum_{s} \sum_{i:sc} x_{isc j} , \quad \text{and} \quad N_{.c}^{-1} \sum_{s} \sum_{i:sc} y_{iscg} ,$$

where j ranges over all predictor variables, including school dummy variables, school grades, test scores, and test dummy variable; and g ranges over all criterion variables, including college grades and college dummy variables. Note that the second coefficients involving the y_{iscg} are zero when g refers to either the grades or the dummy variables for colleges other than the reference college.

2. For sums over all colleges:

$$N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} x_{iscj} \quad \text{and} \quad N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} y_{iscg} ,$$

where j ranges over all predictor variables and g ranges over all criterion variables.

The cards output from program 1 were input for the second section of program 2 after being sorted into the following order.

School Cards 1 in order by school and, within school, by college;
School Cards 2 in order by school and, within school, by college;
Cards for test variables x_{iscv} , x_{iscm} , and x_{iscJ_c} in order by test variable and, within test variable, by college;
Cards for college variables y_{iscg_c} and y_{iscG_c} in order by variable and, within variable, by college.

Note the second use of the cards for the college dummy variable y_{iscG_c} . The pack of cards resulting from this sorting contained 53 groups of cards: a group of School Cards 1 for each of the 19 schools, a group of School Cards 2 for each of the 19 schools, a group for each of the 3 test variables, a group for the college grades variable y_{iscg_c} for each of the 6 colleges, and a group for the college dummy variable y_{iscG_c} for each of the 6 colleges. Each of the groups for the predictor variables (school variables and the test variables) contained cards for all colleges. Each of the groups for the criterion variables (college variables) contained cards for only one college.

A unit of computations was performed for each group of cards, these computations being similar for the various units. In the computation for each unit a series of 53 memory locations was used to store the data from the group of cards. Each of these locations corresponded to a predictor variable or a criterion variable, the first 19 corresponding to the school dummy variables, the next 19 corresponding to the school grades variables, the following 3 corresponding to the test variables, the following 6 to the college grades variables, and the last 6 corresponding to the college dummy variables. At the beginning of the computations for a unit, each of the foregoing memory locations was set at zero. As the cards in a group were read into the machine, the data on the cards were added into the appropriate memory locations of the series. Using the sums and coefficients stored in memory in the first section of the program and the contents of the series of memory locations after all cards in a group had been read the following values were computed.

1. For each group of cards for a predictor variable (let k designate the predictor variable), the following values were computed.

$$\phi_{kj}^{c} = \sum_{c} \sum_{i} x_{ik} x_{ij} - \sum_{c} (N_{.c}^{-1} \sum_{s} \sum_{i:sc} x_{isck})(\sum_{s} \sum_{i:sc} x_{iscj}),$$

$$\phi_{kj}^{t} = \sum_{c} \sum_{i} x_{ik} x_{ij} - (N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} x_{isck})(\sum_{c} \sum_{s} \sum_{i:sc} x_{iscj}),$$

$$\theta_{kg}^{\prime c} = \sum_{c} \sum_{i} x_{ik} y_{ig} - \sum_{c} (N_{.c}^{-1} \sum_{s} \sum_{i:sc} x_{isck})(\sum_{s} \sum_{i:sc} y_{iscg}),$$

$$\theta_{kg}^{\prime t} = \sum_{c} \sum_{i} x_{ik} y_{ig} - (N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} x_{isck})(\sum_{c} \sum_{s} \sum_{i:sc} y_{iscg}),$$

where $\sum_{c} \sum_{i} x_{ik} x_{ij}$ is the contents of the jth location in the portion of the series for the predictor variables, and $\sum_{c} \sum_{i} x_{ik} y_{ig}$. is the contents of the gth location in the portion of the series for the criterion variables. The symbol θ' is used to designate the entries in the transpose of the matrix containing the entries θ. Note that the θ'^{c}_{kg} are zero for g equal to the college dummy variables G_c. Punched cards are output with the results for each group of input cards before the cards for a new group are read.

2. For each group of cards for a criterion variable (let h designate the criterion variable), the following values were computed.

$$\theta_{hj}^{c} = \sum_{c} \sum_{i} y_{ih} x_{ij} - \sum_{c} (N_{.c}^{-1} \sum_{s} \sum_{i:sc} y_{isch})(\sum_{s} \sum_{i:sc} x_{iscj}),$$

$$\theta^{t} = \sum_{c} \sum_{i} y_{ih} x_{ij} - (N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} y_{isch})(\sum_{c} \sum_{s} \sum_{i:sc} x_{iscj}),$$

$$\psi_{hg}^{c} = \sum_{c} \sum_{i} y_{ih} y_{ig} - \sum_{c} (N_{.c}^{-1} \sum_{s} \sum_{i:sc} y_{isch})(\sum_{s} \sum_{i:sc} y_{iscg}),$$

$$\psi_{hg}^{t} = \sum_{c} \sum_{i} y_{ih} y_{ig} - (N_{..}^{-1} \sum_{c} \sum_{s} \sum_{i:sc} y_{isch})(\sum_{c} \sum_{s} \sum_{i:sc} y_{iscg}),$$

where $\sum_c \sum_i y_{ih} x_{ij}$ is the contents of the jth location in the portion of the series for the predictor variables, and $\sum_c \sum_i y_{ih} y_{ig}$ is the contents of the gth location in the portion of the series for the criterion variables. Note that all values of θ^c_{hg} and ψ^c_{hg} are zero when either h or g equals a college dummy variable G_c. Also note that the ψ'^c_{kg} are zero when g does not equal h; thus, the matrix ψ^c of ψ^c_{hg} is diagonal. Both the matrices θ^c and ψ^c may be reduced in size by eliminating the college dummy variables. Punched cards are output with the results for each group of input cards before the cards for a new group are read.

The remaining computations for original summarization of the data were accomplished using library programs for matrix manipulations. The matrices ϕ^c and ϕ^t output by program 2 were factored to the matrices G^c and G^t by determination of the characteristic roots and vectors as outlined in (A.15)–(A.20) and the corresponding matrix products $G(G'G)^{-1}$ were obtained. An IBM 704 computer was used to determine the characteristic roots and vectors of the ϕ matrices. All other computations were performed with an IBM 650 computer.

The matrix ψ^t was factored to the matrix H^t by determination of its characteristic roots and vectors and subsequent calculations similar to those for factoring the ϕ matrices, and the matrix $H^t(H^{t\prime}H^t)^{-1}$ was obtained. Since the matrix ψ^c was diagonal, the matrix of characteristic vectors was known to be an identity matrix and ψ^c was known to contain the characteristic roots. As a consequence

$$H^c(H^{c\prime}H^c)^{-1} = (\psi^c)^{-1/2}.$$

Computations for the Canonical Correlation Models

Essentially parallel computations are involved for the two canonical correlation models, and the general procedure will be outlined here. The superscripts c and t, correspondingly, will be omitted. Each of the matrices and coefficients can have the appropriate superscript attached to specialize the procedure to one or the other canonical correlation models. Following are the computational steps starting from the matrices determined in the original summarization of the data.

1. Determine the matrix M of (A.29):

$$M = (G'G)^{-1}G'\theta'H(H'H)^{-1}.$$

2. Compute the matrix product $M'M$ and determine the characteristic roots and vectors of the product matrix. Each characteristic root is the square of a canonical correlation, r_p^2, and the corresponding characteristic vector is the vector V_p, as per (A.32).

3. Select the canonical variates to be employed in accordance with the sizes of the canonical correlations.

4. For each selected canonical variate determine the weight vectors A_p and B_p in accordance with (A.30), (A.23), and (A.24).

$$U_p = \frac{1}{r_p} M V_p \, ,$$

$$A_p = G(G'G)^{-1} U_p \, ,$$

$$B_p = H(H'H)^{-1} V_p \, .$$

When the weight vectors have been determined, they may be transformed by (44) to obtain weight vectors A_p^* and B_p^* which produce deviation composite scores from the means for all individuals when these transformed vectors are applied to the original scattered data matrices X and Y.

Computations for the Predictive Model

These notes will outline computing procedures for each trial of the second iterative method discussed in Appendix B. The general strategy of this method will be assumed as discussed in Appendix B, including the establishment of an initial trial matrix A of predictor variable weights and the extension or reduction of trial matrices A from the results of one trial before the use of the new matrix A in the next trial.

Computer programs for the predictive model will be described in terms of the computations for the illustrative problem. These computations involved two predictive composites. For a more extensive study involving more schools, and/or more colleges, and/or more predictive composites, the capacities of the several steps will have to be adjusted accordingly. The following notes include several modifications from the programs used. These changes provide some improvements in the program plans as well as some simplification for presentation. The modifications should not affect results obtained.

Input data for program 3 included a trial matrix A of predictor variable weights and the output cards from program 1, these cards being sorted by college, and, within college, into groups by predictor variables and criterion variables. The order of these card groups was: School Cards 1 for the school dummy variables in order by school, School Cards 2 for the school grades in order by school, test verbal score cards, test mathematical score cards, test dummy variable cards, college grades cards, college dummy variable cards.

In program 3, after initialization and storage of the weights in the matrix A in machine memory, the following computations were performed separately for each college to determine that college's regression weights, w_{pgc}, and groups of coefficients used in the subsequent program 4.

1. Three series of 41 machine memory locations were utilized in the first step. In each series there was a memory location for each predictor

variable. For the first series, which was for temporary storage of data read from input cards, the predictor variables corresponding to the memory locations will be designated by h. Each of the remaining two series will be for coefficients designated u_{jp}, which are sums of products between predictor variable scores and predictive composite scores. Two predictive composites p were used in the illustrative study. The predictor variables corresponding to the memory locations will be designated by the letter j.

2. The memory locations in the temporary storage series were set to zero, and the data on the group of cards for a predictor variable, designated by the letter j, were stored in the appropriate locations in the series. These data are the sums of products between the scores on the predictor variable and scores on other predictor variables and the sum of squares of the scores on the predictor variable. Two coefficients were computed for the predictor variable j, one coefficient for $p = 1$ and the other coefficient for $p = 2$:

$$\sum_h \Big(\sum_{i:c} x_{i_c j} x_{i_c h}\Big) a_{hp} ,$$

where $(\sum_{i:c} x_{i_c j} x_{i_c h})$ is the contents of the hth location in the temporary storage series. The two coefficients were stored in the corresponding locations of the two series of locations for the coefficients μ_{jp}. This step was repeated for each predictor variable in turn as variable j.

3. The group of cards for the college grades variable was read into the temporary storage series and the following two coefficients were computed:

$$\sum_h \Big(\sum_{i:c} y_{i_c g_c} x_{i_c h}\Big) a_{hp} ,$$

where $(\sum_{i:c} y_{i_c g_c} x_{i_c h})$ is the contents of the hth location of the temporary storage series and p takes on the values of 1 and 2. These coefficients were stored in machine memory.

4. The group of cards for the college dummy variable was read into the temporary storage series and the following two coefficients were computed:

$$\sum_h \Big(\sum_{i:c} x_{i_c h}\Big) a_{hp} ,$$

where $(\sum_{i:c} x_{i_c h})$ is the contents of the hth location of the temporary storage series and p takes on the values 1 and 2. These coefficients were stored in machine memory. The values of $\sum_{i:c} y_{i_c g_c}$ and $N_{.c}$ were read from the final card for the college dummy variable and placed in machine memory.

5. A coefficient μ_{jp} was computed from each coefficient obtained in step 2:

$$\mu_{jp} = \sum_h \Big(\sum_{i:c} x_{i_c j} x_{i_c h}\Big) a_{hp} - N_{.c}^{-1} \Big(\sum_{i:c} x_{i_c j}\Big) \sum_h \Big(\sum_{i:c} x_{i_c h}\Big) a_{hp} ,$$

where the contents in the temporary storage series are interpreted as the

$(\sum_{i:c} x_{i_c j})$. Each of these coefficients was stored in the location from which the $\sum_h (\sum_{i:c} x_{i_c j} x_{i_c h}) a_{hp}$ was obtained.

6. A coefficient γ_{pg_c} was computed for each value of p (1 and 2).

$$\gamma_{pg_c} = \sum_h (\sum_{i:c} y_{i_c g_c} x_{i_c h}) a_{hp} - N_{.c}^{-1} (\sum_{i:c} y_{i_c g_c}) \sum_h (\sum_{i:c} x_{i_c h}) a_{hp} .$$

7. A coefficient ν_{pq} was computed for each of the three combinations of pq (11, 12, and 22), q being an alternate index to p to designate predictive composite:

$$\nu_{pq} = \sum_h \mu_{jp} a_{jp} .$$

8. The regression weights w_{pg_c} were computed by the following formula:

$$w_{pg_c} = \frac{(\gamma_{pg_c} \nu_{qp} - \gamma_{qg_c} \nu_{pq})}{(\nu_{pp} \nu_{qp} - \nu_{pq}^2)} ,$$

where, when $p = 1$, $q = 2$, and when $p = 2$, $q = 1$.

9. Compute for each predictor variable j:

$$\sum_p \mu_{jp} w_{pg_c} .$$

10. Information output on punched cards from program 3 was the coefficients μ_{jp} of step 5, the γ_{pg_c} of step 6, the ν_{pq} of step 7, the w_{pg_c} of step 8, and the $\sum_p \mu_{jp} w_{pg_c}$ of step 9.

Program 4 computed the new matrix A by the following steps.

1. Compute the entries $\hat{\theta}_{g_c j}$, defined in (B.16), of the matrix $\hat{\theta}$ by the formula:

$$\hat{\theta}_{g_c j} = (\theta_{g_c j})_c - \sum_p \mu_{jp} w_{pg_c} + \sum_p \sum_d (\mu_{jp})_d w_{pg_c} ,$$

where $(\theta_{g_c j})_c$ are the entries in the matrix θ^c obtained in program 2, $\sum_p \mu_{jp} w_{pg_c}$ was obtained in program 3 for college c, $(\mu_{jp})_d$ was obtained in program 3 for college d (in the present case, d ranges over all colleges including college c).

2. Input the matrix $G(G'G)^{-1}$ for the within-college deviation composite scores and compute the matrices:

$$\hat{\theta} G(G'G)^{-1} = M ,$$

$$MM' = Q .$$

3. Determine the characteristic roots Λ and unit vectors V of Q, see (B.32).

4. Compute the new matrix A by the formula:

$$A = G(G'G)^{-1} M' V' \Lambda^{-1/2} N_{..}^{1/2} .$$

This equation is obtained from (B.33) and the scaling requirement of (B.23). Note that there are as many columns in the computed matrix A as there are colleges in the study.

Output from this program were the roots Λ and vectors V of the matrix Q; also output was the matrix A. The series of roots were inspected to determine which columns of A should be used in the next trial. For the illustrative study, the first and third roots were the largest on every trial so that the first and third columns of the output A matrix were taken as the weights for the principal predictive composites and were used in the subsequent trials. In successive trials, the two largest roots increased in size while the other four roots decreased in size. None of these four other roots was large enough to indicate a need for an additional predictive composite.

When a final weights matrix A has been obtained, the column vectors may be transformed by (44) to obtain weight vectors A_p^* which produce deviation composite scores from the means for all individuals when these transformed weight vectors are applied to the original scattered data matrix X.

REFERENCES

Bloom, B. S. and Peters, F. R. *The use of academic prediction scales for counseling and selecting college entrants*. New York: Free Press of Glencoe, 1961.

Burnham, P. S. The evaluation of academic ability. *College Admissions*. New York: College Entrance Examination Board, 1954, pp. 76–92.

Kendall, M. G. *The advanced theory of statistics*. New York: Hafner, 1951.

Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

Toops, H. A. The transmutation of marks. *Ohio College Association Bulletin No. 88*, 1933. (Mimeographed Report)