DOCUMENT RESUME

ED 173 435 TM 009 572

TITLE Proceedings of the Invitational Conference on Testing Problems (10th, New York, New York, October 30, 1954).

INSTITUTION Educational Testing Service, Princeton, N.J.

PUB DATE 30 Oct 54

NOTE 130p.

EDRS PRICE MF01/PC06 Plus Postage.

DESCRIPTORS *Academically Gifted; *Accelerated Programs; Advanced Placement Programs; Analysis of Variance; Decision Making; Educational Testing; Elementary Secondary Education; Factor Analysis; Group Dynamics; *Group Relations; Higher Education; Individual Differences; *Information Theory; Job Placement; Manpower Utilization; Predictor Variables; Psychological Testing; *Psychometrics; Social Relations; *Testing Problems

IDENTIFIERS Latent Structure Analysis; Russell Sage Social Relations Test

ABSTRACT

Four major topics were discussed at the conference: application of information theory to testing; recent advances in psychometric methods; evaluating group interaction; and new developments in the education of abler students. Papers delivered were: Multiple Assignment of Persons to Jobs, by Paul S. Dwyer; New Light on Test Strategy from Decision Theory, by Lee J. Cronbach; The Relation Between Uncertainty and Variance, by William J. McGill; Some Recent Results in Latent Structure Analysis, by T. W. Anderson; The "Moderator Variable" as a Useful Tool in Prediction, by David R. Saunders; A Method of Factoring Without Communalities, by L. L. Thurstone; A New Technique for Measuring Individual Differences in Conformity to Group Judgment, by Richard S. Crutchfield; The Russell Sage Social Relations Test: A Measure of Group Problem-Solving Skills in Elementary School Children, by Dora E. Damrin; Description of Group Characteristics, by John K. Hemphill; Acceleration: Basic Principles and Recent Research, by Sidney L. Pressey; College Admission with Advanced Standing, by William H. Cornog; and Special Treatment for Abler Students and Its Relation to National Manpower, by Dael L. Wolfle. (MH)

1954

20 NASSAU STREET, PRINCETON, N. J.
PRINTED IN THE UNITED STATES OF AMERICA
Library of Congress Catalog Number: 47-11220

INVITATIONAL CONFERENCE

ON

TESTING PROBLEMS

OCTOBER 30, 1954

EDWARD E. CURETON, *Chairman*

Application of Information Theory to Testing

Recent Advances in Psychometric Methods

Evaluating Group Interaction

New Developments in the Education of Abler Students

EDUCATIONAL TESTING SERVICE

PRINCETON, NEW JERSEY LOS ANGELES, CALIFORNIA

FOREWORD

The 1954 Invitational Conference on Testing Problems, the tenth of these meetings and the seventh sponsored by Educational Testing Service, was one of the most successful. Speakers and invited guests from near and far participated in a program of wide scope and significance. This published record of the proceedings will, we hope, carry the words and wisdom of the participants to an even greater audience.

To Edward E. Cureton, Chairman of the 1954 Conference, goes full credit for a job well done in spite of an arduous professional schedule. His imagination, energy, and attention to detail were primarily responsible for the well-planned and well-conducted program. His successful efforts are deeply appreciated by the many who enjoyed the 1954 Conference.

HENRY CHAUNCEY
President

PREFACE

The 1954 Invitational Conference on Testing Problems, sponsored by Educational Testing Service, was held in New York City at the Roosevelt Hotel as in past years, on October 30. This volume provides a permanent record of the papers and discussions.

In arranging the Conference program, it was necessary to try to balance two pairs of conflicting considerations. In recent years the attendance has risen almost to convention proportions. More than 1000 invitations are sent out, and almost half of those invited are usually present. The first problem, then, was to try to preserve something of the original flavor of an invitational conference, and at the same time try to arrange a program which would appeal to some 500-odd people of diverse interests and backgrounds.

At the technical level, measurement theory is undergoing rapid revisions under the impact of communication theory, information theory, statistical decision theory, latent structure analysis, factor analysis, and scale theory, to name only a few. These theories are all highly mathematical, and the branches of mathematics which they use are not well known to most workers in educational and psychological measurement. In the broader field of evaluation, moreover, clinical psychologists and social psychologists are developing new methods of assessing such things as creative talents, personality traits, the dimensions of group interaction, the nature and quality of leadership in various settings, and the processes of human judgment. These methods are in many cases quite different from those employed in assessing cognitive aptitudes and school achievement. It was felt that the new theories and methods should be brought to the attention of measurement workers despite the rather considerable obstacles to effective communication. The second problem was to try to reduce these obstacles as much as possible, and to find speakers who could present some of the new theories and methods in terms which measurement workers could understand.

We were fortunate in securing as our luncheon speaker Dr. Daniel Starch, whose address, ". . . And Have Not Wisdom," recalled forcibly the need to teach students how to make ethical value judgments, and the need to develop methods for measuring the attainment of this vital educational objective.

The rest of the program consisted of a first morning session on some applications of information theory to testing problems, two parallel sessions later in the morning, one on recent advances in psychometric methods and one on the evaluation of group interaction, and an afternoon session on new developments in the education of abler students. We hope that this program achieved the balances implied by its objectives.

The Chairman welcomes this opportunity to express his sincere appreciation to all the speakers for their contributions, to Educational Testing Service for sponsoring the Conference, to Jack K. Rimalover for his unfailing support, assistance, and counsel, and to Mrs. Catherine G. Sharp for her assistance in making such excellent local arrangements.

EDWARD E. CURETON, *Chairman*
1954 Conference

CONTENTS

GENERAL MEETING

Application of Information Theory to Testing

Multiple Assignments of Persons to Jobs*

PAUL S. DWYER

1. Introduction. I want to talk to you about a problem which arises when men are to be assigned to jobs in the most efficient manner. In the time at my disposal I can give you only a brief outline of the nature of the problem, of the general methods proposed for its solution, and of my recent work on the solution of the problem with the use of transformations. However, I am giving you supplementary material which you may examine later in more detail if you desire to obtain a more comprehensive view of the problem and the methods of solution.

2. The Nature of the Problem. A simple illustration may serve to give you some idea of the nature of the problem. A corporation hires 4 college graduates to fill 4 vacancies without determining which individual is to be placed on which job. These graduates, though they have differing abilities as indicated by their records, are all hired for the same salary. The problem of the corporation is then to place these 4 men on the 4 jobs in such a way that the corporation will obtain maximum value from their services.

The corporation may do this by estimating the worth to the corporation in thousands of dollars per year of each individual if he were to be placed in each one of the 4 jobs. Such a set of estimates is shown in Table I. The entries in the table show the values, denoted by c_{ij}, which indicate the estimated contribution to the total effort (in units of $1000) which individual i will make if he is placed on job j. Thus, individual 1 is most valuable on job 1 but so is individual 2 and individual 4. The problem is to place all 4 individuals on all 4 jobs in such a way that the sum of the assigned c_{ij} values is as large as possible.

Now since each individual can fill one and but one job and since each job must be filled by some one individual, it follows that any assignment of the 4 individuals to the 4 jobs must involve one and but one selection from each row and from each column. Hence the problem

* This research was supported in part by the United States Air Force under Contract No. AF18(600)—1050 monitored by Director, Detachment 4 (Crew Research Laboratory), Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

becomes one of making selections of c_{ij} values, one from each row and one from each column, so that the sum is as large as possible.

A more general statement of the problem results from the display of Table II where the c_{ij} values of N jobs and N individuals are indicated. The problem is to select N values of c_{ij} so that the sum of the c_{ij} values so selected is as large as possible. See references A.

3. Alternate Forms of the Problems: The form of Table I and Table II is that of a square array of c_{ij} values featuring N rows and N columns. As a result of grouping the number of columns may be reduced, in some problems, so that the array takes on rectangular form with height greater than width. For example the problem of Table I may be so reduced since the values of c_{ij} in column 3 are identical with the c_{ij} values of column 4. In so far as the solution is concerned, there is no difference between job 3 and job 4 so that the two jobs may be grouped together in a common job category. If we denote the number of such categories by m, and the number of jobs in job category j by q_j, we have $m = 3$, $q_1 = q_2 = 1$, $q_3 = 2$ for the problem of Table I. The values q_j, which indicate the numbers of individuals to be assigned to the respective job categories, are called quotas. This form of the problem which features these job categories and quotas is sometimes known as the quota form.

The quota form of the problem of Table I is shown in Table III. The quota form of the general problem of Table II is shown in Table IV.

Rows may also be grouped to form personnel categories when the c_{ij} values in different rows are identical or approximately so. When personnel categories and job categories are both used, we have a two-way grouped distribution which takes on the form sometimes called the frequency form. The number of personnel categories is taken as n, and the frequency in personnel category i is indicated by f_i. Of course the sum of the f_i values equals the sum of the q_j values which is N. This frequency form of the problem is illustrated in Table V.

The form of Table I and Table II, since it features the nongrouped values for both individuals and jobs, is sometimes called the individual form.

4. Equivalent Problems. This problem is essentially the equivalent of problems in other fields. For example the Hitchcock transportation problem is the mathematical equivalent of the personnel classification problem though it calls for the selection of the c_{ij} values from each row and column so as to minimize, rather than maximize, the sum. There

is no time here for a discussion of these equivalent problems but references B are provided for those who are interested.

5. Methods of Solution Previously Used. Many methods have been proposed for the solution of this problem. Monographs giving surveys of these different methods are mentioned in references C. Of these methods I should call your attention to the method of all possible assignments, the simplex method, and the method of optimal regions.

In the method of all possible assignments, every possible alternative assignment is made. Now there are N! alternative assignments though some of the assignment sums may be equal. Thus with $N = 4$ in Table I, the computation of the $4! = 24$ possible sums shows us the maximum sum of $23,000. But the method of all possible assignments is impractical for larger values of N since N! increases very rapidly.

The simplex method was designed for more general problems in linear programming and game theory. Dantzig and Votaw have applied it to this problem (see D). It is my thesis that the machinery of the simplex method is unnecessarily complex for this problem and that simpler methods, described below, are preferable.

The method of optimal regions is very useful in solving the classification problem especially when the number of job categories is small (see E). But the method I wish to discuss with you today is the method of transformations by which the array of c_{ij} values can be transformed to a new array from which the solution can be obtained by the selection of zero terms.

6. Method of Transformations. A solution of the problem, and there may be more than one, consists in the assignment of each individual i to some job j so that the sum of the c_{ij} values is to be maximized. This maximum sum is not the solution, though it can be calculated once the solution is known. A solution consists in the assignments, i.e., the pairs of values of i and j. Thus in Table I a solution consists in the assignment of man 1 to job 1, man 2 to job 3, man 3 to job 4, and man 4 to job 2. If we indicate the job assigned to individual i by J_i and consider the men in the order 1, 2, 3, 4 we can write this solution compactly by $J_i = 1, 3, 4, 2$. The solution sum is $6 + 6 + 5 + 6 = 23$ units but this is not the solution. The solution is simply the set of elements (i, J_i).

I can now state an important relation which serves as the basis of the method of transformations. Any constant may be subtracted from every element in any row or column without changing the solution. The solution sum is decreased by the amount subtracted but the solution is not changed. Hence we may subtract simultaneously constants

c_i from every row i and constants c_j from every row j without changing the solution. The purpose of the method of transformations is to make use of successive subtractions from rows and columns until an array results from which the solution is immediately obtainable. Specific directions follow.

The first step in the solution of a maximization problem is the subtraction of the largest element in each row from each element in that row. The results of these subtractions $c_{ij}^{(1)}$, are either 0 or negative. The process is illustrated in Table VI where the maximum values for the rows of Table I are shown at the right of the first array and the values of $c_{ij}^{(1)}$ in the second array. Now the $c_{ij}^{(1)}$ array, since it contains only non-positive terms, cannot have a solution with a sum greater than zero. We cannot locate a solution with sum 0 as long as any column has all non-zero terms. So we subtract the largest element in each column as indicated at the bottom of the array. The resulting $c_{ij}^{(2)}$ array has at least one zero in each row and each column. See the third array of Table VI. Sometimes a solution can be obtained from this array by using only 0 elements. This is not possible in Table VI. An additional transformation is in order.

Before indicating the nature of this transformation, we note that the second and third arrays of Table VI feature negative signs. These could be eliminated if, in the first array, the elements were subtracted from the maximum values rather than vice-versa. Then the subtraction of the smallest element in each column of the second array is indicated. This process is illustrated in the first three arrays of Table VII. The positive elements of the second and third of these arrays are identical, aside from sign, with those of Table VI.

We are unable to find a solution using the 0 elements of the third array of Table VII since the 0 terms in columns 2, 3, 4 are all in row 3. However if we subtract -1 from this row, we can then also subtract 1 (the smallest non-zero value in columns 2, 3 and 4) from each element of columns 2, 3 and 4 to form the $c_{ij}^{(3)}$ array. The net sum of the subtracted values is $1 + 1 + 1 - 1 = 2$ which is placed in the lower right corner. In general any such transformation in which the sum of the subtracted constants is positive and which does not result in negative terms, may constitute the next step of the solution.

In this problem the solution can be obtained from the 0 terms of the $c_{ij}^{(3)}$ array. The 0 terms indicating the solution are marked with asterisks in the $c_{ij}^{(3)}$ array. The corresponding terms in the c_{ij} array are also marked with asterisks. The sum of these terms is the solution sum, 23 units. This may be checked by subtracting from 26, the lower right entry in the first array, the sum of the lower right entries in the two arrays below it. In more complex problems additional transformations of this type may be necessary. But we can prove that it is always possible to make transformations of this type so that the solution may be based eventually on 0 elements. These transformations are fairly easily discovered in practice since they are based on patterns of zeros.

The method is applied to a problem in the frequency form, previously used by Votaw to illustrate the simplex method, in Table VIII. The $c_{ij}^{(2)}$ array results from the subtraction from the maximum row values followed by the subtraction of the minimum column values. But, considering frequencies, there are not enough 0 terms in row 2 and row 3 of the third array to satisfy the frequencies. However, the subtraction of 1 from row 2 and row 3, with the subtraction of -1 from column 3 and column 4 (and a net sum of $30 + 35 - 25 - 28 = 12$) leads to the $c_{ij}^{(3)}$ array with very many 0's. One of the many solutions immediately evident from this array is indicated by superscripts.

A final illustration (Table IX) uses a modification of Brogden's quota form problem with 100 men and 4 job categories. The solution follows the steps outlined above. Subtractions are made from the $c_{ij}^{(1)}$ so as to meet the quotas for column 3 and corresponding subtractions are made from the rows so as to keep one 0 term in each row. A corresponding treatment of the first column of the $c_{ij}^{(2)}$ matrix leads to the $c_{ij}^{(3)}$ matrix with enough 0 terms to reveal the solution indicated in the column headed J_i.

The solution sum 708 units can be obtained by adding the c_{iJ_i} values of the first array indicated by the solution. It can be checked by forming $723 - (6 + 9)$.

7. Concluding Remarks. The method of transformations just described gives a solution to the classification problem which is as simple as one can expect. It can be programmed for machines but, except for the most complex problems, hand methods are quite satisfactory.

We now appear to have a good solution for problems with true c_{ij} values. When true c_{ij} values are unavailable, as they commonly are, questions arise as to the estimation of the values, as to the validity and sampling errors of the estimates, as to the resulting effect on the formulation of the problem, etc. The study of this general area, involving possible alternative procedures using the information available, is very important.

A narrower and more immediate practical problem also commonly confronts us. How can we, with our present knowledge and information, calculate any useful estimates of the c_{ij} to which we can apply the available techniques? (See F.) We might use standard scores of variables correlated with success on the job, as in the Brogden illustration, but, as the vocational counselor knows, such single predictors are not commonly immediately observable. And even combinations of observable predictors, such as those obtained by regression, are not enough for this problem, even if valid; since they must be transformed to estimates of the contribution to the common effort. Some sort of a weight must be given to each particular job since a measure of the importance of the job, as well as the proficiency of the individual on the job, is needed for estimating the contribution of the individual to the common effort. How are we to determine these job weights? Aside from the matter of the validity of the predictors we are forced, for the most part, to rely on the estimates of experts or to use hypothetical weights. In conclusion, I would like to pose this question for future research: How can we use available information in obtaining practical objective measures of those job weights whose determination is prerequisite to any useful solution of the personnel classification problem?

Table I: A Simple Problem

	c_{ij} in $1000 units			
i \ j	1	2	3	4
1	6	3	4	4
2	7	4	6	6
3	3	4	5	5
4	8	6	4	4

Table II: The General Problem

	c_{ij}						
i \ j	1	2	3	...	j	...	N
1	c_{11}	c_{12}	c_{13}	...	c_{1j}	...	c_{1N}
2	c_{21}	c_{22}	c_{23}	...	c_{2j}	...	c_{2N}
3	c_{31}	c_{32}	c_{33}	...	c_{3j}	...	c_{3N}
...	...	...	...	...	...	...	...
i	c_{i1}	c_{i2}	c_{i3}	...	c_{ij}	...	c_{iN}
...	...	...	...	...	...	...	...
N	c_{N1}	c_{N2}	c_{N3}	...	c_{Nj}	...	c_{NN}

Table III: Quota for Problem of Table I

	c_{ij} in \$1000 units		
i \\ q_j	1	1	2
1	6	3	4
2	7	4	6
3	3	4	5
4	8	6	4

Table IV: Quota Form for General Problem

	c_{ij}						
i \\ q_j	q_1	q_2	q_3	...	q_j	...	q_m
1	c_{11}	c_{12}	c_{13}	...	c_{1j}	...	c_{1m}
2	c_{21}	c_{22}	c_{23}	...	c_{2j}	...	c_{2m}
3	c_{31}	c_{32}	c_{33}	...	c_{3j}	...	c_{3m}
...	...	...	...	...	...	...	...
i	c_{i1}	c_{i2}	c_{i3}	...	c_{ij}	...	c_{im}
...	...	...	...	...	...	...	...
N	c_{N1}	c_{N2}	c_{N3}	...	c_{Nj}	...	c_{Nm}

Table V: Frequency Form for General Problem

	c_{ij}						
f_i \\ q_j	q_1	q_2	q_3	...	q_j	...	q_m
f_1	c_{11}	c_{12}	c_{13}	...	c_{1j}	...	c_{1m}
f_2	c_{21}	c_{22}	c_{23}	...	c_{2j}	...	c_{2m}
f_3	c_{31}	c_{32}	c_{33}	...	c_{3j}	...	c_{3m}
...	...	...	...	...	...	...	...
f_i	c_{i1}	c_{i2}	c_{i3}	...	c_{ij}	...	c_{im}
...	...	...	...	...	...	...	...
f_n	c_{n1}	c_{n2}	c_{n3}	...	c_{nj}	...	c_{nm}

Table VI: Successive Transformations for Problem of Table I

	c_{ij}				
i \ j	1	2	3	4	Max.
1	6	3	4	4	6
2	7	4	6	6	7
3	3	4	5	5	5
4	8	6	4	4	8

	$c_{ij}^{(1)}$			
i \ j	1	2	3	4
1	0	−3	−2	−2
2	0	−3	−1	−1
3	−2	−1	0	0
4	0	−2	−4	−4
Max.	0	−1	0	0

	$c_{ij}^{(2)}$			
i \ j	1	2	3	4
1	0	−2	−2	−2
2	0	−2	−1	−1
3	−2	0	0	0
4	0	−1	−4	−4

Table VII: Solution of the Problem of Table I

i	c_{ij}				Max.
1	6*	3	4	4	6
2	7	4	6*	6	7
3	3	4	5	5*	5
4	8	6*	4*	4	8
					26

i	$c_{ij}^{(1)}$				
1	0	3	2	2	
2	0	3	1	1	
3	2	1	0	0	
4	0	2	4	4	
	Min:	1			1

i	$c_{ij}^{(2)}$				
1	0	2	2	2	
2	0	2	1	1	
3	2	0	0	0	−1
4	0	1	4	4	
		1	1	1	2

i	$c_{ij}^{(3)}$				J_i
1	0*	1	1	1	1
2	0	1	0*	0	3
3	3	0	0	0*	4
4	0	0*	3	3	2
					23

Table VIII: Solution of Votaw Problem

f_i \ q_j	c_{ij} 15	20	25	28	12	Max.
12	7	8	6	7	8	8
30	6	5	8	6	7	8
35	5	4	5	7	6	7
23	8	9	8	7	9	9
						788

f_i \ q_j	$c_{ij}^{(1)}$ 15	20	25	28	12	
12	1	0	2	1	0	
30	2	3	0	2	1	
35	2	3	2	0	1	
23	1	0	1	2	0	
Min.	1					15

f_i \ q_j	$c_{ij}^{(2)}$ 15	20	25	28	12	
12	0	0	2	1	0	
30	1	3	0	2	1	1
35	1	3	2	0	1	1
23	0	0	1	2	0	
			−1	−1		12

Table VIII (*Continued*)

f_i \ q_j	$c_{ij}^{(3)}$ 15	20	25	28	12	
12	0	0^{12}	3	2	0	
30	0	2	0^{25}	2	0^{5}	
35	0	2	2	0^{28}	0^{7}	
23	0^{15}	0^{8}	2	3	0	
						761

Table IX: Solution of Brogden Problem

i \ q_j	c_{ij} 25	25	25	25	Max.	i \ q_j	c_{ij} 25	25	25	25	Max.
1	−8	−24	1	−4	1	51	6	0	−1	−16	6
2	−3	−19	3	7	7	52	0	0	2	−5	2
3	−4	−18	1	3	3	53	3	1	4	−4	4
4	−11	−17	6	11	11	54	5	1	1	8	8
5	−1	−16	−8	−10	−1	55	7	1	0	6	7
6	−5	−15	7	−3	7	56	3	1	4	7	7
7	−7	−14	−1	−13	−1	57	−3	2	−3	−11	2
8	−6	−13	2	4	4	58	−6	2	−7	−3	2
9	−1	−13	−6	−4	−1	59	−1	2	−8	7	7
10	−2	−12	−9	−11	−2	60	4	2	−5	−12	4
11	−7	−12	3	1	3	61	−5	3	4	10	10
12	−9	−11	−7	−14	−7	62	−1	3	3	0	3
13	−3	−11	1	−9	1	63	−3	3	−4	−18	3
14	−8	−10	−3	5	5	64	1	3	2	−9	3
15	−4	−10	6	0	6	65	0	4	−4	−2	4

Table IX continued pages 22–28

Table IX (*Continued*)

16	−3	−10	−5	3	3	66	6	4	1	4	6
17	1	−9	−10	24	24	[illegible]7	10	4	0	18	18
18	−5	−9	8	9	[illegible]	[illegible]	4	4	0	0	4
19	2	−9	5	−5	[illegible]	69	6	5	−7	−8	6
20	−2	−8	2	12	12	70	4	5	11	13	13
21	−7	−8	6	9	9	71	−2	5	1	−8	5
22	6	−8	−2	9	9	72	9	5	0	−6	9
23	3	−7	2	−1	3	73	−1	6	0	12	12
24	7	−7	−6	2	7	74	0	6	9	19	19
25	−2	−7	3	6	6	75	1	6	4	4	6
26	2	−6	2	8	8	76	8	7	6	−24	8
27	−9	−6	1	−3	1	77	1	7	−4	−1	7
28	−1	−6	−2	−5	−1	78	0	7	1	5	7
29	−6	−5	−11	−17	−5	79	1	8	3	−10	10
30	5	−5	−5	−9	5	80	0	8	−5	−3	8
31	2	−5	−1	2	2	81	1	8	−1	0	8
32	−2	−5	0	10	10	82	8	9	0	−1	9
33	−3	−4	3	−7	3	83	−2	9	−2	2	9
34	−6	−4	4	15	15	84	3	9	−6	5	9
35	1	−4	5	3	5	85	0	10	−6	−13	10
36	4	−4	7	1	7	86	2	10	−4	5	10
37	−5	−3	−1	−6	−1	87	5	10	−1	−2	10
38	−4	−3	0	4	4	88	4	11	−2	−12	11
39	7	−3	−1	6	7	89	3	11	−3	−4	11
40	1	−3	−3	−10	1	90	0	12	8	13	13
41	−2	−2	4	17	17	91	9	12	0	11	12
42	−3	−2	3	8	8	92	0	13	−2	−15	13
43	−10	−2	2	−2	2	93	11	13	−1	−1	13
44	−1	−2	7	−7	7	94	1	14	−3	−19	14
45	−1	−1	−3	−2	−1	95	−4	15	−3	3	15
46	2	−1	−9	−6	2	96	2	16	−2	−8	16
47	−1	−1	5	−10	5	97	0	17	−4	2	17
48	−4	−1	10	−5	10	98	3	18	1	14	18
49	2	0	−2	−7	2	99	0	19	1	1	19
50	3	0	5	16	16	100	5	24	0	1	24
					21						723

Table IX (*Continued*)

	$c_{ij}^{(1)}$						$c_{ij}^{(1)}$				
i \ q_j	25	25	25	25		i \ q_j	25	25	25	25	
1	9	25	0	5	−2	51	0	6	7	22	
2	10	26	4	0		52	2	2	0	7	−2
3	7	21	2	0		53	1	3	0	8	−2
4	22	28	5	0		54	3	7	7	0	
5	0	15	7	9		55	0	6	7	1	
6	12	22	0	10	−2	56	4	6	3	0	
7	6	13	0	12	−2	57	5	0	5	13	
8	10	17	2	0		58	8	0	9	5	
9	0	12	5	3		59	8	5	15	10	
10	0	10	7	9		60	0	2	9	16	
11	10	15	0	2	−2	61	15	7	6	0	
12	2	4	0	7	−2	62	4	0	0	3	−2
13	4	12	0	10	−2	63	6	0	7	21	
14	13	15	8	0		64	2	0	1	12	−1
15	10	16	0	6	−2	65	4	0	8	6	
16	6	13	8	0		66	0	2	5	2	
17	23	33	34	0		67	8	14	18	0	
18	14	18	1	0	−1	68	0	0	4	4	
19	3	14	0	10	−2	69	0	1	13	14	
20	14	20	10	0		70	9	8	2	0	
21	16	17	3	0		71	7	0	4	13	
22	3	17	11	0		72	0	4	9	15	
23	0	10	1	4	−1	73	13	6	12	0	
24	0	14	13	5		74	19	13	10	0	
25	8	13	3	0		75	5	0	2	2	

Table IX (*Continued*)

26	6	14	6	0		76	0	1	2	32	
27	10	7	0	4	−2	77	6	0	11	8	
28	0	5	1	4	−1	78	7	0	6	2	
29	1	0	6	12		79	9	2	7	0	
30	0	10	10	14		80	8	0	13	11	
31	10	7	3	0		81	7	0	9	8	
32	12	15	10	0		82	1	0	9	10	
33	6	7	0	10	−2	83	11	0	11	7	
34	21	19	11	0		84	6	0	15	4	
35	4	9	0	2	−2	85	10	0	16	23	
36	3	11	0	6	−2	86	8	0	14	5	
37	4	2	0	5	−2	87	5	0	11	12	
38	8	7	4	0		88	7	0	13	23	
39	0	10	8	1		89	8	0	14	15	
40	0	4	4	11		90	13	1	5	0	
41	19	19	13	0		91	3	0	12	1	
42	11	10	5	0		92	13	0	15	28	
43	12	4	0	4	−2	93	2	0	14	14	
44	8	9	0	14	−2	94	13	0	17	33	
45	0	0	2	1		95	19	0	18	12	
46	0	3	11	8		96	14	0	18	24	
47	6	6	0	15	−2	97	17	0	21	15	
48	14	11	0	15	−2	98	15	0	17	4	
49	0	2	4	9		99	19	0	18	18	
50	13	16	11	0		100	19	0	24	23	
			2						2		6

Table IX (*Continued*)

q_j / i	$c_{ij}^{(2)}$ 25	25	25	25		q_j / i	$c_{ij}^{(2)}$ 25	25	25	25	
1	11	27	0	7		51	0	6	5	22	−3
2	10	26	2	0		52	4	4	0	9	
3	7	21	0	0		53	3	5	0	10	
4	22	28	3	0		54	3	7	5	0	
5	0	15	5	9	−3	55	0	6	5	1	−3
6	14	24	0	12		56	4	6	1	0	
7	8	15	[illegible]	14		57	5	0	3	13	
8	10	17	0	0		58	8	0	7	5	
9	0	12	3	3	−3	59	8	5	3	0	
10	0	10	5	9	−3	60	0	2	7	16	−3
11	12	17	0	4		61	15	7	4	0	
12	4	6	0	9		62	6	2	0	5	
13	6	14	0	12		63	6	0	5	21	
14	13	15	6	0		64	3	1	0	13	
15	12	18	0	8		65	4	0	6	6	
16	6	13	6	0		66	0	2	3	2	−3
17	23	33	32	0		67	8	14	16	0	
18	15	19	0	1		68	0	0	2	4	−3
19	5	16	0	12		69	0	1	11	14	−3
20	14	20	8	0		70	9	8	0	0	
21	16	17	1	0		71	7	0	2	13	
22	3	17	9	0		72	0	4	7	15	−3
23	1	11	0	5	−2	73	13	6	10	0	
24	0	14	11	5	−3	74	19	13	8	0	
25	8	13	1	0		75	5	0	0	2	

Table IX (*Continued*)

26	6	14	4	0		76	0	1	0	32	−3
27	12	9	0	6		77	6	0	9	8	
28	1	6	0	5	−2	78	7	0	4	2	
29	1	0	4	12	−2	79	9	2	5	0	
30	0	10	8	14	−3	80	8	0	11	11	
31	0	7	1	0	−3	81	7	0	7	8	
32	12	15	8	0		82	1	0	7	10	−2
33	8	9	0	12		83	11	0	9	7	
34	21	19	9	0		84	6	0	13	4	
35	6	11	0	4		85	10	0	14	23	
36	5	13	0	8		86	8	0	12	5	
37	6	4	0	7		87	5	0	9	12	
38	8	7	2	0		88	7	0	11	23	
39	0	10	6	1	−3	89	8	0	12	15	
40	0	4	2	11	−3	90	13	1	3	0	
41	19	19	11	0		91	3	0	10	1	
42	11	10	3	0		92	13	0	13	28	
43	14	6	0	6		93	2	0	12	14	−1
44	10	11	0	16		94	13	0	15	33	
45	0	0	0	1	−3	95	19	0	16	12	
46	0	3	9	8	−3	96	14	0	16	24	
47	8	8	0	17		97	17	0	19	15	
48	16	13	0	17		98	15	0	15	4	
49	0	2	2	9	−3	99	19	0	16	18	
50	13	16	9	0		100	19	0	22	23	
	3						3				9

Table IX (*Continued*)

i \ q_j	$c_{ij}^{(3)}$ 25	25	25	25	j	J_i	i \ q_j	$c_{ij}^{(3)}$ 25	25	25	25	j	J_i
1	8	27	0	7	3	3	51	0	9	8	25	1	1
2	7	26	2	0	4	4	52	1	4	0	9	3	3
3	4	21	0	0	3,4	3	53	0	5	0	10	1,3	3
4	19	28	3	0	4	4	54	0	7	5	0	1,4	4
5	0	18	8	12	1	1	55	0	9	8	4	1	1
6	11	24	0	12	3	3	56	1	6	1	0	4	4
7	5	15	0	14	3	3	57	2	0	3	13	2	2
8	7	17	0	0	3,4	3	58	5	0	7	5	2	2
9	0	15	6	6	1	1	59	5	5	13	0	4	4
10	0	13	8	12	1	1	60	0	5	10	19	1	1
11	9	17	0	4	3	3	61	12	7	4	0	4	4
12	1	6	0	9	3	3	62	3	2	0	5	3	3
13	3	14	0	12	3	3	63	3	0	5	21	2	2
14	10	15	6	0	4	4	64	0	1	0	13	1,3	3
15	9	18	0	8	3	3	65	1	0	6	6	2	2
16	3	13	6	0	4	4	66	0	5	6	5	1	1
17	20	33	32	0	4	4	67	5	14	16	0	4	4
18	12	19	0	1	3	3	68	0	3	5	7	1	1
19	2	16	0	12	3	3	69	0	4	14	17	1	1
20	11	20	8	0	4	4	70	6	8	0	0	3,4	3
21	13	17	1	0	4	4	71	4	0	2	13	2	2
22	0	17	9	0	1,4	4	72	0	7	10	18	1	1
23	0	13	2	7	1	1	73	10	6	10	0	4	4
24	0	17	14	8	1	1	74	16	13	8	0	4	4
25	5	13	1	0	4	4	75	2	0	0	2	2,3	2
26	3	14	4	0	4	4	76	0	4	3	35	1	1
27	9	9	0	6	3	3	77	3	0	9	8	2	2
28	0	8	2	7	1	1	78	4	0	4	2	2	2
29	0	2	6	14	1	1	79	6	2	5	0	4	4
30	0	13	11	17	1	1	80	5	0	11	11	2	2

TABLE IX (*Continued*)

31	0	10	4	3	1	1	81	4	0	7	8	2	2
32	9	15	8	0	4	4	82	0	2	0	12	1	1
33	5	9	0	12	3	3	83	8	0	9	7	2	2
34	18	19	9	0	4	4	84	3	0	13	4	2	2
35	3	11	0	4	3	3	85	7	0	14	23	2	2
36	2	13	0	8	3	3	86	5	0	12	5	2	2
37	3	4	0	7	3	3	87	2	0	9	12	2	2
38	5	7	2	0	4	4	88	4	0	11	23	2	2
39	0	13	9	4	1	1	89	5	0	12	15	2	2
40	0	7	5	14	1	1	90	10	1	3	0	4	4
41	16	19	11	0	4	4	91	0	0	10	1	1,2	1
42	8	10	3	0	4	4	92	10	0	13	28	2	2
43	11	6	0	6	3	3	93	0	1	13	15	1	1
44	7	11	0	16	3	3	94	10	0	15	33	2	2
45	0	3	3	4	1	1	95	16	0	16	12	2	2
46	0	6	12	11	1	1	96	11	0	16	24	2	2
47	5	8	0	17	3	3	97	14	0	19	15	2	2
48	13	13	0	17	3	3	98	12	0	15	4	2	2
49	0	5	5	12	1	1	99	16	0	16	18	2	2
50	10	16	9	0	4	4	100	16	0	22	23	2	2
													708

REFERENCES

A. To the Nature of the Problem.

1. Brogden, Hubert E. An approach to the problem of differential prediction. *Psychometrika*, 1946, 11, 139-154.
2. Thorndike, Robert L. The problem of classification of personnel. *Psychometrika*, 1950, 15, 215-235.
3. Votaw, D. F., Jr. Methods of solving some personnel-classification problems. *Psychometrika*, 1952, 17, 255-266.
4. Dwyer, P. S. Basic theory and methods in optimal classification of personnel. Personnel Research Branch, Department of the Army, 1953.

B. To Equivalent Problems.

5. Dantzig, George B. Application of the simplex method to a transportation problem. Chapter XXIII. *Activity analysis of production and allocation.* Cowles Commission Monograph No. 13. New York: Wiley, 1951.

6. Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

7. von Neumann, John. A certain zero-sum two-person game equivalent to the optimal assignment problem. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the theory of games*, Vol. II. *Annals of Mathematics Study No. 28.* Princeton: Princeton Univer. Press, 1953.

8. Roby, Thornton B. Prerequisites for pair-scores to be used for assembling small work groups. Research Bulletin AFPTRC-TR-54-13 Air Force Personnel and Training Research Center, Lackland Air Force Base, December 1953.

C. To Surveys of Methods of Solution.

9. Smith, Robert B. Hand computational methods for the classification problem. Air Training Command, Human Resources Research Center, September 1951.

10. Votaw, D. F., Jr. and Dailey, J. T. Assignment of personnel to jobs. Research Bulletin 52-24, Air Training Command, Human Resources Research Center, Lackland Air Force Base, August, 1952.

11. Dwyer, P. S. Selected computational methods and related theory in the optimal classification of personnel. Personnel Research Branch, Department of the Army, 1954.

See also (4).

D. To the Simplex Method.

See (3), (4), (5), (10).

E. To the Method of Optimal Regions and Related Theory.

12. Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. *Psychometrika*, 1954, 19, 11-26.

13. Lord, Frederick M. Notes on a problem of multiple classification. *Psychometrika*, 1952, 17, 297-304.

See also (1), (2), (6), (11).

F. To the Determination of c_{ij} Values.

14. Dwyer, P. S. Selection and linear combination of tests in relation to multiple criteria and differential classification. PRB Research Note 7, Department of the Army, January 1953.

See also (1), (2).

New Light on Test Strategy from Decision Theory*

LEE J. CRONBACH

In every practical use of tests, our aim is to make decisions. This is obvious in personnel selection and in Dr. Dwyer's assignment problem, but it is also true of testing in the classroom and in the clinic. The teacher uses tests because he has to make decisions about appropriate instructional methods. The clinician uses tests as an aid in deciding on therapeutic tactics. Sometimes, as in vocational guidance, the decisions are made not by the tester but by the person tested. Test theory should indicate how to reach the best possible decision in any of these situations.

We use the word *strategy* to refer to the process by which an individual arrives at a decision. A strategy may be very simple: "I shall examine the applicant's grade average, and if it is B or better I shall accept him." The strategy may instead be complex, stating what tests if any will be given, what decision will be made for any particular pattern of results, and what further steps will be taken to decide on borderline cases. Choosing among alternative strategies is the essential problem of test theory.

There are two questions in choosing a strategy. First, with any given procedure for gathering information, what is the best procedure for translating this information into final decisions? Dr. Dwyer has just shown us the solution for one problem of that type. The second, but logically prior, question is: Among several alternative procedures for gathering information, which is most profitable?

In order to compare two strategies, we have to determine how much benefit we gain from either one. Most of the problems of decision theory therefore reduce to determining just how much benefit is gained from a particular decision-making procedure.

Since this morning's program is intended to deal primarily with insights from some of these newer points of view, I shall not dwell on the mathematics of decision problems. There is available a large amount of relevant theory in the work of economists on utility, in the theory of games, and in the statistical decision theory of Abraham

* Based on work conducted under Contract N6ori-07146 with the Office of Naval Research.

Wald. Decision problems can be attacked in many ways, but we have confined ourselves to strategies which maximize expected utility. This is reasonable only if we are dealing with a stable and familiar situation.

The decision model requires us to specify three aspects of any decision. One is the proposed strategy or decision rule. For example, the strategy might be to give two tests, combine scores by a regression formula, and accept everyone above a given cutoff. Second, we consider the adequacy of the information to be used. The usual contingency matrix or scatter diagram relating test scores and criterion scores deals with this question. Dr. McGill's work with information analysis is primarily concerned with studying validity matrices. The third necessary element is an evaluation matrix. This, sometimes called a payoff matrix, states specifically just what benefit or detriment accompanies each possible decision. Dr. Dwyer's Tables I and II are evaluation matrices (but his problem is so stated that the validity of the predictors also affects his entries). Once these three aspects of a problem have been described, we are ready to compute the payoff a person can expect if he bases decisions upon this information and this strategy.

Dr. Dwyer properly drew attention to the fact that it will be difficult and at times impossible to write the evaluation matrix for a particular situation. To let this difficulty deter us from using decision theory, however, would be to deny the possibility of sound test theory. Test effectiveness simply *cannot* be evaluated without an evaluation matrix. Even the conventional procedures of test analysis assume certain payoffs covertly, and the reasonableness of some of these hidden assumptions is open to question. In the future, testers may wish to determine utilities by a sort of cost accounting in any specific practical situation, in order to arrive at the best decisions. Our project is proceeding along different lines. We are working with hypothetical (but we hope realistic) decision problems. By assuming that the evaluation matrix has some characteristic form, we are able to judge the utility of different types of tests and strategies. Such an approach can be no better than our assumptions. We hope nonetheless to arrive at general principles of testing which will illuminate many real situations.

Let me turn to some of the concepts a decision approach brings to our attention. I shall cover four such points.

1. Our model suggests that the value of test information should be judged by how much it improves decisions over the best possible decisions made without the test, whereas the conventional validity coefficient reports how much better test decisions are than *chance* de-

cisions. In the majority of situations where tests could be used, a substantial amount of information is already available, and if no test were given the decision would still be considerably better than chance.

Our most valid tests are essentially work samples of their criteria. Where such a work sample might be used, evidence of past performance is also a valid basis for decisions, and such evidence is often readily available. In predicting school marks, for example, a scholastic aptitude test is not greatly more valid than past school records. The contribution of this test to decision making is much smaller than its zero-order validity coefficient would indicate, because better-than-chance decisions could be made without it.

A similar conception applies to classroom testing. The basic knowledge and skill objectives can be assessed with considerable accuracy from day-to-day assignments; a test can add only a small increment to the soundness of decisions. On the other hand, a teacher has rather little basis for judging which pupils have problems of adjustment. The teacher may therefore gain more useful knowledge from a test of adjustment which has limited validity, than from an achievement test which largely duplicates data already available. There are serious weaknesses in our tests for such educational outcomes as creativity, reasoning habits, attitudes, and application of knowledge to problem situations. They are markedly inferior in validity to tests of general intelligence or factual knowledge. But the factors that make testing difficult also prevent valid non-test decisions about these objectives. It may therefore be wiser to use imperfect tests of important objectives that are hard to measure, than to use highly valid tests that merely supplement non-test data.

Utility analysis leads us to examine the value of adapting to individual differences in either selection or placement. This can best be considered in terms of a placement problem, such as assigning students to sections of freshman English according to their initial ability. We might think of the various levels as predetermined, and of the test as assigning persons to each category using fixed cutting scores. It is sounder to see the test, the curricula, and the cutting scores as interlocked. We can increase or decrease the demands instruction in any section makes, to fit it to the ability of the persons assigned. Under this procedure we benefit more from testing than when we leave the treatment fixed.

Certain simple assumptions lead to interesting conclusions. If a sample is divided into groups, using fixed cutting scores, the extent to which treatment for the groups should be differentiated depends on the validity of the placement test. If the information has zero validity, utility is maximized when we teach all sections in the manner suited to the average of the population. As validity increases, the treatments given the sections may differ more; but no matter how valid the test, there is an optimum degree of differentiation of treatment. If treatment is differentiated beyond this point, the benefit from sectioning declines. Indeed, it is possible to differentiate treatments so radically that a *loss* in utility results from sectioning even though the test used has considerable validity.

This analysis raises serious question as to whether we are right when we urge teachers to adapt to individual differences. If the teacher has a standard plan, well fitted to the average of the group, he should hesitate to depart from it. Marked alteration of the plan to fit individuals appears to be advisable only when individual differences are validly assessed and their implications for treatment clear.

3. We turn, now, to another suggestion encountered in decision theory. It is customary to look at a test as a unit, and to use it just for one terminal decision. At any point in testing, however, we can make a terminal decision or can continue to gather information. New frontiers open for us when we view testing as a multi-stage, or sequential, operation.

Suppose, in a simple selection problem, we have several short aptitude tests, which together might constitute a selection battery. We give the first short test; some men can be rejected or accepted at once, but less clear-cut cases are retained for further testing. After the second test is given to these men, we can make more final decisions, and only a borderline group goes on to the third test. This process terminates when the benefit from information to be gained at any stage is outweighed by the cost of testing. Considering cost of testing, the sequential method is more profitable than giving the same test to everyone. If testing is expensive, one reaches the final decision for a surprisingly large proportion of men, after only the first short test. One paper on this line of attack has been published by Arbous and Sichel (1), but our detailed results will differ from theirs in important respects.

A sequential plan would require new ways of organizing testing. I shall discuss one procedure for possible use in vocational guidance, clinical diagnosis, or evaluation of classroom learning. Here it can be described in terms of the job assignment problem. For different jobs, so many abilities are relevant that we cannot hope to measure them all accurately in a reasonable period. With brief tests, however, one could crudely measure as many as 50 variables in a half day. Such a survey will indicate some jobs for which the man is an unlikely prospect, and a second group of jobs for which the tests show possible high aptitude. For the man's second test session, perhaps only an hour after the first test, we would assemble a set of booklets to test him more thoroughly on those promising aptitudes. This progressive narrowing would be continued. When the final job assignment is made, we would have a highly reliable measure of the man's aptitude for that job, and also a good measure for the other jobs seriously considered as alternatives. But we would have wasted little time in getting an accurate measure of his space ability or his dexterity, if these areas were not among his better aptitudes on the first survey.

We might actually develop different sorts of tests for the earlier and later screens. The Strong Vocational Interest Blank might be replaced by a brief questionnaire, perhaps one page long. This seems likely to identify the important interest groups for a given man. There might then be a separate, longer interest blank for each of these interest groups, to provide more precise differentiation between related occupations than is now possible.

4. Perhaps our most far-reaching conclusion is that we should take a more favorable view of tests with low validity. Traditionally, if a score has low validity, we conclude that it should not be used. But such tests become valuable when selection ratios are low (as Taylor and Russell noted), when they give even a little new evidence on an important decision, and particularly when they are used as a preliminary survey.

The survey is especially important when many decisions are to be made. Sometimes, as in vocational choice, the decisions are interrelated and lead to one final course of action. The decisions may, on the other hand, be quite independent, as when one diagnoses many persons. The problem in testing is ordinarily to select information-getting devices which will yield greatest benefit for the time available. If we spend a lot of time to get an accurate answer to one question, we must answer other questions without added information. In this situation it may be much wiser to use several tests of limited validity, so that every decision is made with some wisdom, than to get highly accurate information for just one of the decisions.

The difference between validity and utility is clear when we compare group and individual tests. An individual mental test measures one person, with essentially the same expenditure of effort by the tester as the group test measuring one hundred. If the two tests have the same validity, the group test gives us 100 times as much information, and bears on 100 decisions while the individual test bears on one. Hence the improvement of decisions is vastly greater when the group test is used. If the individual test has much higher validity than the group test, which is best to use would depend on the specific decision problem.

This conception permits a favorable view of interviews and other clinical procedures which cover many aspects of the personality. These methods, though undependable, are well suited to a wide-casting survey, gathering a little information on each of hundreds of questions. Such a preliminary scanning draws attention to the critical areas where further information should be gathered prior to any decision. The traditional, narrowly-focussed measuring device is ideal when we know in advance exactly what question needs to be answered. But in deciding whether a man will make a good executive, or in locating a patient's chief conflict areas, no such focus is possible. The first task of the assessor is to discover which critical variables will dictate the proper decision about the individual; in different cases, different variables will be critical.

Personnel workers regard the interview as indispensable, and clinicians have considerable faith in projective methods and qualitative analysis of intelligence test protocols. In my opinion, this faith has developed largely because of rewarding experience with these techniques in their survey function, i.e., as the first stage in a sequential assessment. If it is true that these multi-dimensional techniques have a unique place in assessment, we should judge how well they do that job, and should not demand that they be good measuring instruments—which they are not. On the other side of the picture, if their proper function is to make preliminary surveys so that more intensive examination can follow, one should not rest final judgments on these fallible instruments.

Taken as a whole, decision theory is a mathématical system which permits us to examine the problems we face in developing tests, choosing between tests, and interpreting tests. Whenever we can specify any particular decision problem in the detail Professor Dwyer's problem required, then decision theory can tell us just what to do. By studying common type-problems, decision theory can also offer general recommendations regarding testing strategy.

Conventional test theory assumes that we use tests to obtain numerical measures on an interval scale, as in the physical sciences. That is rarely or never true. The function of psychological and educational tests is to aid in making discrete decisions. The greatest contribution of decision theory is to help testers see this function more clearly.

(See next page for references)

REFERENCES

1. ARBOUS, A. G., AND SICHEL, H. S. On the economies of a pre-screening technique for aptitude test batteries. *Psychometrika*, 1952, 17, 331-346.

2. BLACKWELL, DAVID, AND GIRSHICK, M. A. *Theory of games and statistical decisions*. New York: Wiley, 1954.

3. BROGDEN, H. E., AND TAYLOR, E. K. The dollar criterion—applying the cost accounting concept to criterion construction. *Personnel Psychol.*, 1950, 3, 133-154.

4. BROSS, IRWIN D. J. *Design for decision*. New York: MacMillan, 1953.

5. CRONBACH, LEE J. The counselor's problems from the perspective of communication theory. In Vivian H. Hewer (Ed.), *New perspectives in counseling*, Minnesota Studies in Personnel Work, No. 7. Minneapolis: Univer. of Minnesota Press, 1955.

6. EDWARDS, WARD. The theory of decision making. *Psychol. Bull.*, 1954, 51, 380-418.

7. GOODMAN, LEO A. The use and validity of a prediction instrument. *Amer. J. of Sociol.* 1953, 58, 503-512.

8. THRALL, R. M., COOMBS, C. H., AND DAVIS, R. L., Editors, *Decision Processes*, New York: Wiley, 1954.

The Relation Between Uncertainty and Variance

WILLIAM J. McGILL

There has been a great deal of hue and cry about information theory in psychological circles. In the midst of this hue and cry, it is easy to become confused about where information theory is supposed to make contact with psychology. What, if anything, should information theory contribute to psychology?

The theory deals with transmitting symbols from one system to another. That certainly sounds familiar. Many questions that come up in learning, in perception, psycho-physics, testing (just to mention a few areas), take on new significance in the terms of information theory. Consequently when we read presentations of information theory and run across words like "message," "noise," "channel capacity" our nostrils begin to quiver and we sniff a familiar scent in the air. At once we try to find analogues for the classical problems of psychology in the theorems presented to us by the communication engineers. This is one large zone of contact between information theory and psychology. Information theory is a source of analogies and ideas that might not have occurred to us, if we thought about our problems in another way. Perhaps the analogies are helpful, perhaps not. I would rather not discuss the merits of using information theory for this purpose. I bring it up only because I want to hammer at the distinction between information *theory* and information *measures*. The theory is concerned with transmitting symbols despite noise. Information *measures* are concerned with the arithmetic of mean-log-probability.

Mean-log-probability is something like variance. It measures the amount of spread in a discrete probability distribution. The formula for mean-log-probability is usually written as follows:

$$U(y) = - \sum_{k=1}^{k=r} p(k) \log_2 p(k)$$

In this formula, y is a variable that can assume any one of *r* discrete values. Each of these values has some probability, p(k). The negative sign before the summation insures that U(y) is positive. The interesting thing about mean-log-probability is that any numbers or measurements which the various categories of y might represent do *not* appear in the formula for U(y). In other words U(y) is non-metric.

Mean-log-probability can be calculated for a variable like "methods of parental discipline" (just to choose an example), where no numbers can be attached to any of the methods. One method is as good or as bad as another. All you can say about them is that they are different. Obviously, you can also compute mean-log-probability for metric variables like IQ, but in so doing you sacrifice whatever you might gain from knowing that an IQ of 100 is very close to an IQ of 105 and very far from an IQ of 150.

In the formula for mean-log-probability, logs are taken to the base 2. This is done in order to provide a simple unit—the "bit." When U(y) is measured in bits, it turns out that U(y) is the average number of binary, or two-category decisions required in order to identify one of the values of y exactly. It is not really necessary to measure mean-log-probability in bits, but almost everyone does.

You can apply mean-log-probability in many situations where information theory has nothing whatever to say, and where in fact an attempt to apply information theory might even look a little ridiculous.

For example, suppose you want to express the relation between anxiety level in children and several methods of parental discipline. The relation might be demonstrated very easily by using mean-log-probability. But our understanding of the relation would not be enhanced if we considered the disciplinary methods as messages and the anxiety levels as the versions of these messages received at the end of the channel. The "capacity" of such a "channel" is of little interest to us. The fact that we can with effort construct information-theory-type interpretations of relations like this one, merely demonstrates that communication engineers and psychologists measure things with roughly the same kind of arithmetic. It does not mean that information *theory* has anything significant to contribute to our understanding of the relation between discipline and anxiety. Consequently we can be interested in the information measures entirely apart from their significance in information theory. To fortify this distinction, let us now call these information measures by the name "uncertainty" measures. Uncertainty means mean-log-probability. It has no necessary connection with information *theory*.

In this paper I want to show that uncertainty is essentially non-metric variance. This can be a very hollow claim revealing little more than a superficial similarity unless we are prepared to outline in detail the properties of variance that uncertainty possesses.

What are the principal properties of variance?

1. Variance is a measure of spread or variability—so is uncertainty.

2. Variances that are independent are additive—uncertainties that are independent are also additive.

3. Variances can be partitioned into components that reflect contributions to the total variance from a number of predicting variables taken one at a time and in combination; uncertainty has precisely the same property.

4. Variances are seriously affected by the metric or scale properties of the data from which they are derived. On the other hand uncertainty is unaffected by the metric or scale properties of the data. You can chop a distribution into parts and rearrange all the parts without changing uncertainty—but the variance *will* change radically.

It is obvious that uncertainty and variance are closely related but it is equally obvious that the two measures are not identical. There is no simple equation that takes you from uncertainty to variance. Instead, the parallel is in terms of structure or operations. To make this point clearer I have prepared two tables, Tables 1 and 1A in the handout. Table 1 shows the symbols, formulas, and definitions used in a double-classification analysis of variance. Table 1A shows the same arrangement for a double-classification uncertainty-analysis. You can see by glancing across the tables that the parallelism is very complete. In both cases the predictor variables do not have to be metric. They can be pure classifications like our example of methods of parental discipline. In the variance analysis the dependent variable is metric. It is something that can be measured numerically. I have pictured the dependent variable also as having discrete classifications but this is just a convenience of notation.

You can see the relation between the two analyses very clearly when you look at this pair of equations:

$$U(y) = U_{wx}(y) + U(y{:}w,x) \quad (1)$$
$$V(y) = V_{wx}(y) + V(y{:}w,x) \quad (1a)$$

The definitions of the terms in these equations are explained in Tables 1 and 1A. The equations state that the variability of the criterion or dependent variable can be analyzed into predictable and unpredictable parts. Furthermore the predictable variability may be decomposed as follows:

$$U(y{:}w,x) = U(y{:}w) + U(y{:}x) + U(y{:}\overline{wx}) \quad (2)$$
$$V(y{:}w,x) = V(y{:}w) + V(y{:}x) + V(y{:}\overline{wx}) \quad (2a)$$

where again U stands for uncertainty and V stands for variance. These equations state that the total predictable variability can be broken down into a part predictable from w, a part predictable from x, and a part predictable from unique combinations of w and x.

Sometimes the uncertainty interaction term, $U(y{:}\overline{wx})$, can be negative. What is not generally realized is that $V(y{:}\overline{wx})$, a component of *variance*, can also be negative if it is defined as it is in Table 1. The reason is that this term contains a hidden cross product, a correlation term, as well as a sum of squares. Furthermore the correlation term has a negative sign and can sometimes be greater in magnitude than the interaction sum of squares. This means that the interaction term in the variance-analysis can sometimes be pushed negative by the negative correlation term.

When does this happen? It happens when the predictor variables are not independent. In ordinary analysis of variance this situation never comes up because we are careful to construct the analysis so that there are equal numbers of observations in every cell. This automatically insures that the predictor variables are independent. We can guarantee that things behave properly in both uncertainty-analysis and variance-analysis by making the predicting variables orthogonal, i.e. independent. When the predictor variables have some sort of metric this is equivalent to requiring that the correlation between the predictors is zero. In fact, all the arguments I have been making for the similarity between uncertainty-analysis and analysis of variance are equally true of multiple regression analysis, i.e. the case in which the variances in Table 1 are computed around a regression plane. In multiple regression any interaction computed by the rule shown in the last line of Table 1 is bound to be due to non-orthogonality.

We see that multiple correlation, analysis of variance and uncertainty analysis are all very closely related. The different predictive methods are necessitated because sometimes the predicting variables have metrics and sometimes not, sometimes the dependent variable has a metric and sometimes not.

What remains stable in each case is the *structure* of the statistical process of prediction, the operations involved in making the prediction.

One important practical consequence of this invariance of structure in statistical prediction is that we can now analyze non-metric data in the manner of analysis of variance without resorting to the distortion of giving the data an artificial metric. The non-metric analysis is carried out with uncertainty measures and mean-log-probability. Another very important consequence is that all the literature on experimental design in variance analysis is directly applicable to the uncertainty analysis.

Table 1

Symbols, formulas, and definitions used in double classification analysis of variance. The criterion variable, y, is assumed to be metric. The predictor variables, w and x, are categorized but not necessarily metric.*

Symbol	Formula	Definition
1. $V(y)$	$\sum_{k} \frac{n_{..k}}{n}(y_k - \bar{y})^2$	*Total Variance:* The variance of the criterion variable y.
2. $V_w(y)$	$\sum_{i} \frac{n_{i..}}{n} \sum_{k} \frac{n_{i.k}}{n_{i..}}(y_k - \bar{y}_{i..})^2$	*Conditional Variance:* The variance of y when one predictor variable, w, is held constant.
3. $V_{wx}(y)$	$\sum_{i,j} \frac{n_{ij}}{n} \sum_{k} \frac{n_{ijk}}{n_{ij}}(y_k - \bar{y}_{ij})^2$	*Error Variance:* The variance remaining in y when both predictor variables, w and x, are held constant.
4. $V(y{:}w)$	$V(y) - V_w(y)$	*Main Effect:* The variance of y due to the predictor variable w.
5. $V_w(y{:}x)$	$V_w(y) - V_{wx}(y)$	*Partial Main Effect:* The variance of y due to the predictor variable x, when the predictor variable, w, is held constant.
6. $V(y{:}w,x)$	$V(y) - V_{wx}(y)$	*Total predictable Variance:* The variance of y due to the joint influence of the predictor variables w and x.
7. $V(y{:}\overline{wx})$	$\begin{cases} -V(y) + V_w(y) + V_x(y) - V_{wx}(y) \\ V_w(y{:}x) - V(y{:}x) \end{cases}$	*Interaction variance:* The variance of y due to unique combinations of the predictor variables, w and x.

* The criterion variable is indicated by y. It can assume any value y_k. The two predictor variables are w and x, and these two variables can assume values w_i or x_j. Thus the data arrays are three dimensional. In the three dimensional data matrix, n_{ijk} refers to the n[illegible] cases occurring in a single cell. The dot notation indicates that n_{ijk} has been added over the missing subscripts.

Table 1A

Symbols, formulas, and definitions used in three-variable uncertainty analysis. The criterion variable y is assumed to be non-metric. The predictor variables, w and x, are categorized and may or may not be metric variables.*

Symbol	Formula	Definition
1. $U(y)$	$-\sum_{k}\frac{n_{..k}}{n}\log_2\frac{n_{..k}}{n}$	*Total uncertainty:* The amount of uncertainty in the criterion variable y.
2. $U_w(y)$	$-\sum_{i}\frac{n_{i..}}{n}\sum_{k}\frac{n_{i.k}}{n_{i..}}\log_2\frac{n_{i.k}}{n_{i..}}$	*Conditional uncertainty:* The amount of uncertainty in y when one predictor variable, w, is held constant.
3. $U_{wx}(y)$	$-\sum_{i,j}\frac{n_{ij.}}{n}\sum_{k}\frac{n_{ijk}}{n_{ij.}}\log_2\frac{n_{ijk}}{n_{ij.}}$	*Error uncertainty:* The amount of uncertainty remaining when both predictor variables, w and x, are held constant.
4. $U(y{:}w)$	$U(y) - U_w(y)$	*Contingent uncertainty:* The uncertainty in y due to the predictor variable, w.
5. $U_w(y{:}x)$	$U_w(y) - U_{wx}(y)$	*Partial Contingent Uncertainty:* The uncertainty in y due to the joint influence of the predictor variables w and x.
6. $U(y{:}w,x)$	$U(y) - U_{wx}(y)$	*Multiple Contingent Uncertainty:* The uncertainty in y due to the joint influence of the predictor variables, w and x.
7. $U(y{:}\overline{wx})$	$-U(y) + U_w(y) + U_x(y) - U_{wx}(y) - U_w(y{:}x) - U(y{:}x)$	*Interaction uncertainty:* The uncertainty of y due to unique combinations of the predictor variables, w and x.

DISCUSSION

PARTICIPANTS

LEE J. CRONBACH, EDWARD E. CURETON, PAUL S. DWYER, DOROTHEA EWERS, M. M. KOSTICK, WILLIAM J. MCGILL, VICTOR H. NOLL, JOSEPH ZUBIN

DR. NOLL: This is by way of comment rather than criticism on a point that Dr. Cronbach made in his paper. I have reference to the second point in relation to use of tests in providing for individual differences.

It seems to me that the implication there was a little unrealistic in that we do not ordinarily, in the classroom, attempt to provide for individual differences on such a mechanical basis. We do not say, "Your IQ is 117; consequently, you do this," and "Your IQ is 90, and you do this."

It seems to me it is more a matter of providing a variety of materials from which the student then, by a process of selection, determines what is suitable for him. In other words, it is more a curricular problem than it is a problem of measurement. Perhaps I am putting an implication on Dr. Cronbach's statement which he did not intend, but I think it is important that we think of it in terms of providing a variety of experiences from which students of different types of ability and interest then choose rather than a mechanical process of determining, on the basis of a test, just what a student is going to do.

DR. CRONBACH: I agree with Dr. Noll's views on the curriculum. The generalization of the paper is this: our system of analysis forces us to reconsider much of the doctrine we have had regarding individual differences. This doctrine takes a variety of forms. For instance, experiments on homogeneous groups have not led to conclusive results. In these studies, pupils have been given different treatments, treatments varied rather mechanically, plus some additional flexibilities such as Dr. Noll mentioned. The results to be expected from X-Y-Z sectioning will differ greatly, depending upon the extent to which X pupils truly differ from Y's and Y's differ from Z's. I know of nothing done in these research programs to make sure that the adaptation to individual differences applied in each section was optimal.

If treatment is overdifferentiated or underdifferentiated, they certainly do not get the full value of homogeneous grouping.

Our result also bears on the assessment literature. Thus it has been said that there must be something wrong with counselors because their standard deviation of estimates of ability is lower than the measured standard deviation. But the evidence is that counselors are making mistakes chiefly because they are *over*differentiating, taking the tests too seriously.

Obviously we have been encouraging teachers to look at the students' personalities and to handle some of them differently from others, on the basis of judgment that one needs encouragement and another needs stern treatment.

The answer lies in this direction: it is fine to differentiate treatment, so long as you do it on a tentative basis, and allow for trial and error. If you make irreversible decisions, such as "you cannot go to college," then your information has to be extremely good.

DR. CURETON: It seems to me we might consider one illustration while we are on this topic, because it happens to be fairly common in American universities and to rest on pretty definite information of a non-metric type. This is the matter of freshman English. I think probably in colleges and universities you will find more specific sectioning in English than in almost any other area. The reason is that there seem to be some fairly definite cutoffs which imply specifically different types of instruction. For example, if a student can write good English sentences quite regularly, and can also handle them in paragraphs and larger units with acceptable style, many universities will excuse them from freshman English at the outset. Secondly, if he can write acceptable grammatical sentences, but does not do too well in organizing them into paragraphs and larger units and produces poor style, he is likely to be assigned to a regular section. And finally, if he is unable to write grammatical sentences and handle properly the basic elements of grammatical structure, then he is likely to be assigned to a special section in English in which he will get the type of drill that is not needed by those whose secondary preparation is somewhat better.

DR. CRONBACH: Our analysis suggests research on the performance of students of differing initial ability under various treatments along this continuum, from very routine English training to the highest level of English. The highest level might involve going into matters of style, for instance, well above the usual considerations of clarity.

DR. KOSTICK: This is for Dr. Cronbach: I was wondering about using the sequential method in entrance examinations.

I carried out a study of the effectiveness of entrance examinations in our Massachusetts State Teachers Colleges. We found that by the end of the sophomore year, of those people who were admitted by certification, only 6% "flunked out" or dropped out with low grades. We also found that for those people who were admitted with an entrance examination mark of over 160, only 18% were dropped. Of course, the mark of 160 or any other mark doesn't have meaning unless you know what it stands for. We used a combination of the score on the psychological examination and the score on the English Examination.

What I am particularly interested in, is the lowest group. We have a range between 130 and 140 which is questionable. From that group, 48% "flunked out."

I was wondering if it would be a good suggestion to see whether we could use some sort of sequential testing of these people, instead of going to the expense of putting them through the freshman or sophomore year and then having practically 50% of them drop out.

DR. CRONBACH: In sequential testing you often can get more thorough measures than you could afford to give to everyone. There is no compelling reason why borderline cases could not sit for a two-day examination even though you wouldn't think of doing this with every entrant. However, you are likely to make a great deal more gain if you find some new sources of tests that will supplement your prediction rather than just extending the old one.

The statement that was made focuses on the importance of having very clearly in mind your value assumptions before you try to make decisions about test strategy. We can each make our own value assumption. I would be hesitant in this particular example to accept the implication that if a person flunks out after two years, this result has negative utility. There may be positive utility in those two years of college, particularly when judged from the viewpoint of the individual, rather than from the values of the institution. I think we fall too easily into seeing the problem only from the institutional side.

DR. EWERS: I simply want to add to Dr. Cureton's statement that we do the same thing for freshmen in high school, except we have one more category, that is, if they cannot read, we put them in a fourth group. But I wonder really how well we do teach them in these various levels. If we could do a better job there, we might not have so much trouble at the college level.

DR. ZUBIN: Dr. McGill, I am a little puzzled by your analysis, especially by the terms you used. They evoke memories of previous knowledge and I wonder whether you would like to connect up for us the previous meaning attached to these times with the meaning you have given them. Analysis of variance gives the components of the total variance in a systematic fashion, and, of course, leaves a little bit for the uncontrolled variance as an error term. Ordinarily we refer to it as the standard error, something left with which to measure the significance of the single components. As a matter of fact, what is the relationship between uncertainty and these older ideas of standard error? Are they at all related, or are they two completely independent things?

Dr. Dwyer's method, of course, applies only to situations where you already know the value of each job and you know the value of the person, and so on. Now, could that method be applied to situations where we still do not know these values? In the clinical field, for example, we do not know the amount of money we could save by a certain procedure. Would it be possible to set up a contingency table with unknown c_{ij}'s and then solve for them under certain conditions? Rather than assume you know your c_{ij}'s, could you put in specified c_{ij}'s and see what happens?

DR. MCGILL: The answer is that it was my intention to try to evoke associations from as many people as possible, because I believe that if you have the old associations firmly in hand, it is extremely easy to manipulate these so-called new concepts. They are not new at all.

The analogy of uncertainty with error variance is an example. The "error" uncertainty contains error plus everything else you forgot to analyze. If you have constructed an experiment properly (which almost nobody ever does), the error will be what the model claims it is.

The interesting thing about uncertainty analysis is its technique for testing null hypotheses. In this analysis you do not test the predictable components against error; you test the error against zero and the predictable components against zero. It is a tricky little switch, but the interpretation of the components is identical with variance analysis.

DR. DWYER: You can't work the problem unless you have some values of c_{ij}. There isn't any reason why you can't have a whole series of hypothetical c_{ij}'s and maybe the collective problem would be some sort of answer. This formulation of the problem demands it, but we could have a whole series of hypothetical values, which might be interesting.

SECTION 1

Recent Advances in Psychometric Methods

Some Recent Results in Latent Structure Analysis

T. W. ANDERSON

Latent structure analysis may be thought of as an analysis of discrete data that is analogous to factor analysis of continuous data. The usual model for factor analysis is

$$x_i = \sum_{\nu=1}^{m} \lambda_{i\nu} f_\nu + \mu_i + u_i, \qquad i = 1, \ldots, p, \tag{1}$$

where x_i is the i-th test score, f_ν is the ν-th factor score, μ_i is a constant and u_i is the sum of the i-th specific factor and the error of measurement. The test scores are observed; the quantities on the right hand side are unobserved or latent.

For convenience of discussion, we shall assume that the f_ν and u_i are normally distributed. The key assumptions of this model are that the set u_i are statistically independent of (uncorrelated with) the set f_ν and the u_i are statistically mutually independent (mutually uncorrelated). This means that if we take the subpopulation defined by the requirement that the factor scores are given values, then in this subpopulation the test scores are statistically independent; that is, given the factor scores the predictability of one test score from another is zero. More specifically, given the values of f_ν, the test scores are normally and independently distributed with means $\Sigma_\nu \lambda_{i\nu} f_\nu + \mu_i$ and variance σ_i^2. These assumptions are reflected in the formulas for the variances and covariances of the test scores. If we assume the factors are uncorrelated and have unit variances and let the mean of u_i be zero and the variance of u_i be $E\, u_i^2 = \sigma_i^2$, then the variance of the i-th test score is

$$E(x_i - \mu_i)^2 = \sum_{\nu=1}^{m} \lambda_{i\nu}^2 + \sigma_i^2, \tag{2}$$

and the covariance of two test scores is

$$E(x_i - \mu_i)(x_j - \mu_j) = \sum_{\nu=1}^{m} \lambda_{i\nu} \lambda_{j\nu}, \qquad i \neq j. \tag{3}$$

We can say that the common factors "explain" the interdependence or correlation of the test scores; the effect of the u_i appear only in the variances.

In considering latent structure analysis, we shall assume that each item is a dichotomy; that is, an item score takes on the value one for a positive response and zero for a negative response. Thus the continuous test scores x_i are replaced by the discrete item scores y_i. The continuous factor scores f_ν are replaced by the latent attributes g_ν, which may be discrete or continuous. In the subpopulation defined by given latent attribute scores, the responses to the p items are statistically independent; in other words, given the latent attribute score the predictability of one item from another is zero. In such a subpopulation let $\pi_i(g)$ be the probability of a positive response to the i-th item; for ease of exposition we shall assume there is only one latent attribute score g. Then, for example, the probability of a positive response on every item is

$$(4) \quad \Pr\left\{ y_1 = 1, y_2 = 1, \ldots, y_p = 1 \mid g \right\} = \pi_1(g)\pi_2(g) \ldots \pi_p(g).$$

For given g, the y_i are a set of independent binomial variables. Next we assume that there is a distribution of the latent attribute, say f (g) over the whole population. The probabilities of $y_1, \ldots, y_p$ or the relative frequencies of various response patterns for the entire population are obtained from those for the subpopulations by averaging with respect to f (g).

To make these ideas more concrete we shall consider two special cases of the latent structure model. One of these can be derived from the factor analysis model. For simplicity we shall assume that there is one common factor f and $\mu_i = 0$; that is,

$$(5) \quad x_i = \lambda_i f + u_i, \qquad i = 1, \ldots, p.$$

Suppose that the continuous score x_i is replaced by a dichotomous variable y_i with $y_i = 1$ if $x_i > a_i$ and $y_i = 0$ if $x_i \leq a_i$. From (5) we know that $x_1, \ldots, x_p$ are normally distributed with means zero, variances $\lambda_i^2 + \sigma_i^2$ and covariances $\lambda_i \lambda_j$, and from this fact we can compute the probability of [illegible] pattern of the dichotomous responses. For example,

$$(6) \quad \Pr\{y_1 = 1, \ldots, y_p = 1\} = \int_{\alpha_p}^{\infty} \ldots \int_{\alpha_1}^{\infty} h(x_1, \ldots, x_p)\, dx_1 \ldots dx_p,$$

where h $(x_1, \ldots, x_p)$ is the density of the normal distribution. When f is fixed, the conditional distribution of x_i is normal with mean $\lambda_i f$ and variance σ_i^2; in this subpopulation the probability of a positive response on the i-th item is

$$(7)\quad \Pr\{y_i = 1 \mid f\} = \pi_i(f) = \int_{\alpha_i}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}(x_i - \lambda_i f)^2/\sigma_i} dx_i = 1 - \Phi\left(\frac{\alpha_i - \lambda_i f}{\sigma_i}\right),$$

where Φ (b) is the probability of a unit normal variable being less than b. From the fact that the u_i are statistically independent it follows that in the subpopulation of a given f the x_i are independent and therefore the y_i are independent. For example,

$$(8)\quad \Pr\{y_1 = 1, \ldots, y_p = 1 \mid f\} = \pi_1(f) \ldots \pi_p(f) = \left[1 - \Phi\left(\frac{\alpha_1 - \lambda_1 f}{\sigma_1}\right)\right] \ldots \left[1 - \Phi\left(\frac{\alpha_p - \lambda_p f}{\sigma_p}\right)\right].$$

The case of one latent attribute (that is, one factor score) has been of particular interest to sociologists and social psychologists. The items may have to do with opinions on various questions, all of which are related to a given attitude. For instance, the items may be questions such as "Would you like to work with a negro?" "Would you like to live in a nonsegregated community?", etc. The underlying attitude is that of racial prejudice. The end purpose of the investigation may be to rate the respondents on a scale of racial prejudice.

In (7) $\pi_i(f)$ is given as a particular function of f. Other functions may be postulated. Lazarsfeld has been studying mathematical and statistical problems that arise when $\pi_i(f) = a_i + b_i f + c_i f^2$ or $\pi_i(f) = a_i + b_i f^{c_i}$ $(0 \leq f \leq 1)$.

Perhaps the simplest of the latent structure models arises when the latent attribute is discrete. Suppose that the attribute can have one of q values, say $f = 1, \ldots, q$. Then $\pi_i(f)$ can be designated as $\pi_i{}^{\alpha}$, where $\alpha = 1, \ldots, q$. The distribution of the latent attribute is specified by $\Pr\{f = \alpha\} = \nu_\alpha$. The probability of drawing an individual from the α-th latent class is ν_α; the probability of drawing such an individual and getting a positive response on the i-th item is $\nu_\alpha \pi_i{}^{\alpha}$; and the probability of drawing some individual and getting a positive response on the i-th item is

(9) $$\pi_i = \sum_{\alpha=1}^{q} \pi_i^\alpha \nu_\alpha.$$

Similarly, the over-all probability of getting positive responses on items i and j is

(10) $$\pi_{ij} = \sum_{\alpha=1}^{q} \pi_i^\alpha \pi_j^\alpha \nu_\alpha, \qquad i \neq j,$$

and the probability of getting positive responses on i, j, and k is

(11) $$\pi_{ijk} = \sum_{\alpha=1}^{q} \pi_i^\alpha \pi_j^\alpha \pi_k^\alpha \nu_\alpha, \qquad i \neq j, j \neq k, i \neq k.$$

There are similar expressions for the probabilities of other sets of responses.

The π_i, π_{ij}, and π_{ijk} are known or are estimated from the data; the π_i^α and ν_α are to be inferred. When the structure is known a respondent can be classified into one of the q latent classes on the basis of his responses to the items.

A method for inferring π_i^α and ν_α, suggested by Bert Green [2], is based on factor analysis methods. Equation (10) can be written

(12) $$\pi_{ij} = \sum_{\alpha=1}^{q} (\pi_i^\alpha \sqrt{\nu_\alpha}) (\pi_j^\alpha \sqrt{\nu_\alpha}) \; ;$$

these are identical with (3) if $\lambda_{i\alpha} = \pi_i^\alpha \sqrt{\nu_\alpha}$. The matrix (π_{ij}) can be factored to determine the matrix $(\lambda_{i\alpha}) = (\pi_i^\alpha \sqrt{\nu_\alpha})$, but there is left the indeterminacy of rotation. To eliminate this, Green suggests also factoring $\Sigma_k \pi_{ijk}$. Generally there will be only one solution that factors both matrices; in the case of fallible or statistical data there will be one factor matrix that gives the best fit. (This description of the method should be taken only as approximate.)

An important difficulty with this method is that π_{ii}, π_{ijj}, etc. are not defined and not observed and these must be approximated in some fashion; this causes particular trouble with the second matrix $(\Sigma_k \pi_{ijk})$ that is factored. Another method [1] which bypasses this difficulty as well as the need for factoring large matrices is to use only part of equations (9), (10), and (11). The part of (10) that is used is a set that is entirely below the main diagonal of the matrix (π_{ij}). In this method $2q^2$ of the observed π_i, π_{ij} and π_{ijk} are used to infer the same number of the latent parameters. The algebra that is involved is the standard theory of characteristic roots and vectors, but we cannot go into details here. This method has been developed further to use more of the data, and a large sample theory of tests and confidence regions has been derived. Neither of the methods mentioned is in principle efficient in using all of the information in the data, but either one makes it possible to analyse data in terms of this model. However, further developments in statistical inference are needed even for the simple case.

REFERENCES

1. ANDERSON, T. W. On estimation of parameters in latent structure analysis. *Psychometrika*, 1954, 19, 1-10.

2. GREEN, BERT, JR. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 1951, 16, 151-166.

3. LAZARSFELD, P. F. A conceptual introduction to latent structure analysis. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. The Free Press, 1954, Pp. 349-387.

4. STOUFFER, SAMUEL, *et al. Measurement and prediction*. Vol. IV. *Studies in social psychology in World War II*. Princeton Univer. Press, 1950. Ch. 10 and 11.

The "Moderator Variable" as a Useful Tool in Prediction

DAVID R. SAUNDERS

This paper is intended to be partly informative and partly persuasive. From the information standpoint, I hope to supply you with answers to three main questions: What is a moderator variable? How do you use a moderator variable? and, thirdly, Why should you, anyway? The persuasive aspect of the paper is obvious in the third question, just as in the title. It is an important aspect, because much of the information I can give you is neither very new nor very complicated.

Moderator variables have been used for many years by our friends in economics and agriculture to help in fitting regression surfaces to their data. In these applications the moderator effect is typically just one of many that are possible with a multivariate curvilinear regression based on a polynomial expansion. Our economic and agricultural friends don't need and don't seem to have any special name for it. Our biological friends might be tempted to suggest the name "synergistic variable," but this is a term that already has a lot of additional scientific connotations that we want to remain neutral on.

So far as I know, Gaylord and Carroll were the first to use a moderator variable in psychology. They called it a "population control variable," and presented a paper on it during the 1948 APA meetings. The term "population control variable" is a good one, because it suggests a very important application. But it is a bad term to the extent that it tends to blind us to a number of other equally important applications, which I will touch on. The term "moderator variable" seems to be general enough in its meaning, and still not to be loaded with too many undesirable connotations.

Well—we have just christened this thing, but have only hinted at what it is. Let's look at some examples. By now I've managed to think of dozens of attractive hypothetical examples, but I'll spare you these and emphasize two examples that have been fully worked out and cross-validated.

The first example is the one that originally led me to think a moderator variable might be an important concept. Frederiksen and Melville, at ETS, had just shown that interests were less predictive of academic success for "compulsive" people than for "non-compulsive" people (2).

For this discussion, we can ignore the measures that were used to define interests, success, and compulsiveness, except to note that they *all* were regarded as continuous variables. Frederiksen's and Melville's experimental design, nevertheless, had to be the analysis of covariance; the total experimental sample was arbitrarily divided near the median compulsiveness score to produce two groups—called "compulsives" and "non-compulsives." The relations between interest and success were compared for these groups, and were found to differ significantly in the slopes of the regression lines. This was done separately for ten different interests as predictors.

This example is typical of many situations in which the use of a moderator variable should be considered. In this particular kind of situation the term "population control variable" would also be apt. Clearly, what we are after is a means of treating compulsiveness as the continuous variable which it is—a means of avoiding the arbitrariness of dichotomizing or otherwise dividing the population into *smaller* pieces—a means of maintaining the integrity of the total population while still maintaining a statistical control on each individual's membership in one of a continuous, infinite series of sub-populations defined by his compulsiveness score. In short, we will allow the meaning of his interest score to be "moderated" by his compulsiveness score.

Mathematically, this turns out to be extremely simple to do. Suppose we start with an ordinary linear regression using several variables. Keeping everything in standard scores, we would write the equation

$$y = \sum_{i=1}^{k} \beta_i x_i,$$

where y is our criterion, and β_i is the beta-weight for predictor x_i. Now suppose that β_i, instead of being a constant, is itself a linear function of a series of moderator variables, z_j's. If we plug this into our equation, and do a little rearranging, we immediately find that we can write

$$y = \sum_{i=1}^{k} a_i x_i + \sum_{j=1}^{l} b_j z_j + \sum_{ij} c_{ij} x_i z_j.$$

This would be just another linear regression if it were not for the last term, involving the products of the x's and the z's. If we want to, we can always choose origins of measurement for the x's and z's that will make everything drop out of this equation *except* the products, and a constant term.

It is evidently these product terms which are inextricably tied up with the use of a moderator variable. And that is all that has happened. So long as there is a clear separation of the x's and z's, we cannot have any use for squared variables, let alone terms of higher power. All we have to do to fit the model to data is to find the appropriate xz product (or products) for each subject, and treat it (or them) as new independent predictors in any standard multiple correlation technique. Of course, we cannot introduce a product variable into a battery unless both of its factors are already there.

There are many interesting mathematical sidelights to this thing, but if I want to persuade you that moderator variables are useful and practical, we'd better go back to our examples.

We took the data that Frederiksen and Melville had collected, and computed the products of the interest predictors with compulsiveness for each individual subject. Then we ran multiple correlations, first adding compulsiveness to the interest predictor, and then adding their product to the battery (3).

For three of the ten interest predictors studied, the simple addition of compulsiveness to the battery gave a significant increase to the multiple R. In these three cases the compulsiveness score happened to act as a suppressor variable. A moderator variable is different from a suppressor variable, though they both typically have zero zero-order relation to the criterion; a moderator variable does not have to have zero-order correlations with even the predictor.

Back to the example: In five of the remaining seven instances, addition of the appropriate product score to each battery of two measures resulted in a further significant increase in the multiple R. The sign of the beta weight for the third term was correctly predicted from the hypothesis in all ten of the ten instances. You will recall that the hypothesis told us at which end of the compulsiveness scale to look for good predictions.

These results looked promising. So we moved the scene of operations from Princeton to Rochester; from a group of self-referred counselees to a larger group tested routinely during Freshman Week; from Strong Interest Blanks scored with weighted responses to scoring with unit weights. The criterion was still freshman grade average for engineering majors, and the moderator was still the Accountant Interest score of the Strong, as a measure of compulsiveness. Insofar as the same or similar interest scores were available to try as predictors, we were able to cross-validate all but one of the statistically significant findings from Princeton.

In this example we have observed that in the more significant instances, the predictive contribution of the moderator effect is just as great as the initial contribution of the interest variable predictor. This may not seem like saying much, but note that Frederiksen and Melville reported correlations ranging all the way from zero to over fifty for sub-groups of the same population, divided using a pair of compulsiveness measures. Even when a moderated regression model does not lead to much increase in the multiple R, it does lead to quite different predictions for some individuals, and should be used if its effect is even statistically significant. These predictions may differ in standard error as well as in the expected value itself.

In the second main example that I want to discuss here, *neither* the moderator *nor* the predictor has been shown to have a significant zero-order correlation with the criterion! In this situation, the predictor and moderator lose their separate identity, and a name like "population control variable" becomes more awkward to use. In this example, there is no significant multiple R until the product term is introduced; then it jumps to values like .45. This example is based on some of Fiedler's recently reported work on the influence of leader-keyman relations on small group effectiveness. Many of you may have seen this written up in *Time* magazine recently, even if you missed Fiedler's APA presentation and haven't yet seen his report to the Office of Naval Research (1).

The gist of it is this. Suppose you are the formally designated leader of a small group or team. There is probably someone in the group who is your right-hand man—your principal subordinate or keyman. If you are the kind of person who generally has warm feelings for most people, for whom similarities among other people are more important than differences, it will pay you and your group for you to maintain a relative aloofness from your group, and especially from your keyman. On the other hand, if you tend to think of others you like as being different from those you dislike, it will pay to do the opposite, namely for you to cultivate strong sociometric ties with your group, and especially with your keyman.

Here, then, are two variables you need to measure to predict a particular leader's effectiveness in a particular group. While neither variable is related to the criterion, their product has a substantial negative correlation with it. The results happen to be psychologically very sensible, and they can probably be used to counsel leaders towards a more effective style of leadership, and to predict what groups will respond best to given styles of leadership.

These two examples have been very different in many respects, but they do have two things in common in addition to their featuring of important moderator effects. For one thing, they both feature non-cognitive variables as predictors—and as moderators, if you still care which is which. It seems to me that we can expect to find examples of this kind relatively more easily, but cognitive examples are still a real possibility. For instance, if we could obtain a meaningful coefficient of reliability for an individual's test score, independently of the group within which he happened to be tested, it would probably moderate any predictions made from the test score itself.

In the second place, both these examples were initially studied by breaking a total sample up into sub-groups. There are many studies that have been carried this far, and then reported in tedious detail for lack of an organizing concept such as the moderator variable provides. For the past three years I have made it a point at the APA meetings to seek out studies of this kind; I have always found my fill without going very far afield. For some reason, studies involving the "Authoritarian Personality Syndrome" seem to be in particular need of something like a moderator variable. But no more so, I would say, than people who do configural scoring or make clinical judgments on the basis of personality tests.

There is one last topic on my agenda. Assuming you have decided to look for moderator effects, what is the best way to go about it? There are at least four methods to consider. 1. You can start with a good hypothesis. This is always a good idea. 2. If you don't have one handy, you can look for one by studying a few cases intensively, and seeking out variables whose interpretation seems to depend on other variables. 3. You can do the same thing with larger groups of cases by looking for sets of sub-groups within which correlations are significantly different. 4. You can do the same thing with items instead of variables, by testing the interaction variance of a pair of items against a criterion. This fourth approach is a whole technique in itself, and I wish there were time to tell you about it. It capitalizes on the electronic computers; it can be generalized beyond pairs of items; and it brings the old idea of keying patterns of response to several items into a new perspective.

REFERENCES

1. Fiedler, Fred E. The influences of leader-keyman relations on combat crew effectiveness. Urbana: Group Effectiveness Research Laboratory, University of Illinois, June 1954. (Technical Report No. 9, Contract No. N6-ori-07135.)

2. Frederiksen, N., and Melville, S. D. Differential predictability in the use of test scores. *Educ. psychol. Measmt.*, **14**, in press.

3. Saunders, D. R. Moderator variables in prediction. Princeton: ETS Research Bulletin 53-23, October 1953.

A Method of Factoring Without Communalities

L. L. THURSTONE

Ever since multiple factor analysis was developed over twenty years ago, the unknown diagonal cells of the correlation matrix have been a serious problem. These are called the communalities and they represent that part of the total variance of a test which it shares with one or more of the other tests in the battery. For each method of factoring it has been necessary to estimate these diagonal values in the correlation matrix and various methods have been devised for doing so.

In a theoretical case one can construct a correlation matrix of order n and rank r where $r < n$. The diagonal entries of such a correlation matrix are then so determined that all minors of order $(r + 1)$ vanish. When the order n is larger than, say, 6 or 8, the diagonal values are usually unique. In dealing with experimentally observed correlation coefficients, the coefficients necessarily have variable errors so that a low rank cannot be found in any exact manner. However, one can usually write another correlation matrix with side entries that are nearly the same as the experimentally observed values and which is of a rank much lower than n so that, for example, $r < n/2$. This is the situation often found in multiple factor studies.

Various methods of factoring the correlation matrix have been devised in which the diagonal elements are first estimated. One of the simplest is to assign to each diagonal cell a value equal to the absolute value of the highest entry in the corresponding column or row. The estimate is revised after each factor has been extracted. This rough method of estimation is quite successful when the order is fairly large, say, about 15 or 20 or higher. Although this procedure is quite successful for many scientific problems with correlation matrices of high order, it is far from satisfactory for theoretical formulations of the factoring problem.

There is an important relation between the communalities and the number of factors that are used for describing a correlation matrix. The larger the number of factors, the higher are the communalities. The communality of any test variable j is the sum of squares in the corresponding row of the factor matrix F. If it is decided that the first

r principal axes are inadequate for the description of the correlation matrix, then the next principal axis may be determined. This gives $(r + 1)$ columns of the factor matrix. The communality of any test j is then augmented by the square of the entry in column $(r + 1)$ in the row j. But it is never known beforehand how many principal axes need to be determined in order to reduce the residual correlation coefficients to values that can be ignored in terms of their sampling errors. Hence we have the anomalous situation of not knowing at the start of the computations how many factors need to be postulated to account for the correlations. The communality estimate should rise with the number of factors that are extracted but the relation depends entirely on the unknown configuration of test vectors in each given problem.

In practice the problem is resolved by adjusting the communalities for each factor that is extracted from the correlations or by repeating the whole factoring process until the estimated communalities at the start agree with the sum of squares of the factor loadings in the rows of the resulting factor matrix. But this process depends on a certain number of factors as determined by the first cycle. The adjustment should be done over again if the investigator decides to increase the number of factors used. When the number of factors is quite large, such as 10 or 15, then the adjustments in the diagonals are ordinarily quite small for each additional factor.

In trying to relate the theory of multiple factor analysis to practical scientific work with large correlation matrices, this situation is evidently quite unsatisfactory, even though the practical compromises have been adequate to resolve most of the scientific problems so far in the isolation of the components of human intelligence which have been called primary mental abilities.

Several months ago I was sitting in an airplane in Helsinki in Finland, waiting for the take-off for Stockholm. It occurred to me suddenly that this awkward situation that we have fought and compromised with for a long time could be resolved in a ridiculously simple way. I shall describe the idea here although we have not yet developed the best computing methods, and I hope that some other students of the multiple factor problem will consider the new method and how it might be reduced to the simplest possible computing procedures.

In the experimentally given correlation matrix, the diagonal cells are unknown. They are certainly not experimentally given. Let us determine the first column of the factor matrix by the method of least squares so that the first factor residuals are a minimum. This is precisely what we do in determining the first principal component by Hotelling's iterative method *with one important exception.* This exception is that we ignore the unknown diagonal cells completely. Since they are unknown, they are not represented in any observation equations. Only the experimentally known correlation coefficients participate in the observation equations for the least squares solution. When the first column of the factor matrix has been determined so as to minimize the residuals for the side correlations, then we find the first factor residuals for the side entries. The diagonal cells are ignored completely. Then we proceed as before for each additional column of the factor matrix until the residuals of the side correlations can be ignored. The communalities for this given number of factors are then simply the sums of squares of the rows of the factor matrix F. In this procedure we determine the communalities after the factoring job has been completed. If, for any reason, we decide to reduce the residuals still further, that can be done by extracting another factor. The new communalities are then merely the sums of the squares of the rows of F with another column added. None of the factoring need be repeated because communality estimates are not involved in the factoring of the correlation matrix.

It was several weeks until I had the opportunity to try the new method in Frankfurt, Germany, where I discussed this method with three of my former students in Chicago. They were Dr. Hans Anger, Dr. Sten Henrysson, and Rolf Bargmann. Bargmann set up three test cases and reported that the solution gets close to the principal axes solution and Dr. Sten Henrysson has reported similar findings with the method in Uppsala. When I return to my laboratory in Chapel Hill, I expect to investigate the method further.

The computational procedure can be tried in a manner analogous to Hotelling's iterative solution but it is likely that one of several other alternatives will be more effective. In one manner of writing the problem we get third degree normal equations which can be solved by successive approximations with additive corrections to the assumed factor loadings.

There may be an interesting geometric twist to this problem in that the customary geometric model may have to be revised but I am not prepared yet to elaborate on the geometric implications of this problem. The slight disturbance of the geometrical model for the correlation coefficient and the factor matrix can be seen with a theoretical case of exact rank r. When this new method is used, the r th factor residuals do not vanish identically. They are small, however. The factoring matrix obtained can still be given the usual geometrical interpretation and the correlation coefficients are closely represented as scalar products of the test vectors that are defined by the factor matrix. It is a question to be investigated whether the disturbance of the true geometrical model is less important than the disturbance implied in all the adjustments of the communalities in practical computing procedures.

This procedure calls to mind Spearman's formula for the best fitting single factor. However, the application of such a formula may lead to trouble if applied to the residuals for subsequent factors after the first factor. The theory of this problem should also be investigated with reference to the possible appearance of the Heywood case. So far we have not encountered it. In several trials we find factorial solutions that are close to the principal axes.

We hope that other students of factorial analysis may be interested to explore this factoring method.

DISCUSSION

PARTICIPANTS

T. W. ANDERSON, EDWARD E. CURETON, FREDERIC LORD, FRANK ROSENBLATT, P. J. RULON, DAVID R. SAUNDERS, ROBERT L. THORNDIKE, L. L. THURSTONE, LEDYARD R TUCKER, JOHN W. TUKEY, JOSEPH ZUBIN

DR. ZUBIN: Dr. Anderson, did you make each probability of response independent, or would you allow for interaction, for instance, between saying yes to one item and no to another? Is the probability of that pattern based on the product of the original probabilities, or might there be some interaction?

DR. ANDERSON: The assumption here is that given the latent structure score, then the responses are independent. If one estimates the latent parameters, and then writes down the right-hand sides of the equations and finds they do not correspond so well to the left-hand sides, that would indicate that the assumption is not borne out very well. Then one can extend the model so that essentially, instead of basing this on one latent score, you have several latent scores. For a particular couplet, let us say, you may have to throw in an extra latent score to account for the kind of interaction you are talking about.

DR. RULON: Are there any other questions? I do not know whether Dr. Anderson made himself so clear that there are no questions, or so obscure that there are no questions.

DR. CURETON: I have a comment that might have a bearing on Dr. Zubin's question. It is to be noted that the number of latent classes is so determined that the responses will be independent within such classes as far as that can be determined statistically. In other words, you simply postulate enough latent classes so that you do have independence within each one of them.

DR. ZUBIN: Is there any way of telling how many subgroups you have in the overall sample?

DR. ANDERSON: One technique for doing that is to carry over what you do in factor analysis, again using the analogy between these equations. In factor analysis, there are techniques to determine how many factors there are, and the same techniques can be used here to determine how many latent classes there are. However, from the point of view of the theoretical statistician, these techniques are not as well worked out, or they do not apply as well, as in the ordinary factor analysis.

DR. CURETON: Dr. Saunders, at the same time that you put in the "moderator variable," the formal product, wouldn't it be equally easy to put in the squares of the X's and Z's, so that the direct regressions are stretched from linear to parabolic?

DR. SAUNDERS: In general, there is no reason why this cannot be done. Again, in general, if you do do this, you are increasing the number of parameters that are going to be determined from the experimental data. In the absence of some kind of evidence to suggest that these square terms are important, I think it would be a cautious approach to steer clear of them. Let us be open-minded, to be sure.

DR. THORNDIKE: Along the same line, isn't a fifth possible way of exploring these moderator variables just routinely to compute all the products and run them in as additional variables in your basic prediction enterprise?

DR. SAUNDERS: This is an approach, and it is a lot of work.

DR. CURETON: You were talking about electronic computers.

DR. ROSENBLATT: If you start with a matrix of variables, you could compute an ordinary linear multiple r for this matrix, and take the coefficient of determination (or r^2). If you compare this coefficient of determination with that of a correlation which includes the product terms, have you given any thought to how the difference between these correlations stands with respect to the usual notion of statistical interaction, coming out of an analysis of variance? It seems to me that the correspondence becomes quite close (i.e., the incremental correlation due to the moderator variable corresponds closely to the interaction variance)

DR. SAUNDERS: I would say that it is quite close. Incidentally, this spells out an approach that you can take in testing the significance of the product term. You need to compare these coefficients of determination with and without the product terms.

DR. THURSTONE: This is exactly the same thing as we get in Hotelling's procedure, except that the sum of the squares of the factor loadings in Hotelling's method is the sum of the squares in all of the factor loadings and the corresponding relations here would include the diagonals. In this case we deal only with the known values. But if you write zero in the diagonals, then this term can be interpreted in an ordinary way. This is the sum of the squares of all but one of the factor loadings.

PERLMAN, Mildred, Department of Personnel, New York City
PERLOFF, Robert, Personnel Research Branch, AGO
PERLOFF, Evelyn, Board of Education, Maryland
PERRY, William D., University of North Carolina
PETERSON, Donald A., Life Insurance Agency Management Association
PHELAN, Robert F., Prudential Insurance Company
PIERSON, George A., Queens College
PINZKA, Charles F., Educational Testing Service
PITCHER, Barbara, Educational Testing Service
PLUMLEE, L. B., Educational Testing Service
POLIN, A. Terrence, U. S. Air Force, Mitchel AFB
POLLACK, Norman C., New York State Department of Civil Service
PRESSEY, Sidney L., Ohio State University
PRESSEY, Mrs., Sidney L., Ohio State University
RABINOWITZ, William, Bank Street College of Education, New York City
RADASCH, John, Department of Education, Concord, New Hampshire
RAPPARLIE, John H., Owens—Illinois
RASKIN, Evelyn, Brooklyn College
RASKIN, Judith, Queens College
RAYMOND, Thomas J., Harvard University
RAYMOND, Vincent R., Newark College of Engineering
READ, Thomas, Peddie School, Hightstown, New Jersey
REGAN, James J., Special Devices Center, ONR
REICHLER, Melvin, Queens College
REMMERS, H. H., Purdue University
REUTER, W. H., Educational Testing Service
RHULE, Warren A., U. S. Armed Forces Institute
RICHARDSON, M. W., Richardson, Bellows, Henry and Company, Incorporated
RICHARDSON, Ruth P. NYSPA, New York City
RICKS, J. H., Jr., Psychological Corporation
RIMALOVER, Jack K., Educational Testing Service
ROBBINS, Irving, Queens College
ROCA, Pablo, Department of Education, Puerto Rico
ROCK, Robert T., Jr., Fordham University
ROSENBLATT, Frank, Cornell University
ROSNER, Benjamin, Columbia University
RULON, P. J., Harvard University
SADACCA, Robert, Educational Testing Service
SAUNDERS, David, R., Educational Testing Service
SAIT, Eward, Rensselaer Polytechnic Institute
SAWIN, E. I., Air Force ROTC Headquarters, Montgomery
SCATES, Alice Y., U. S. Office of Education
SCATES, Douglas E., American Social Hygiene Association
SCHAPIRO, Harold B., Great Neck, New York
SCHRADER, William B., Educational Testing Service
SCOTT, C. Winfield, Vocational Counseling Service, Inc.
SEASHORE, Harold G., Psychological Corporation
SEBALD, Dorothy D., Hunter College
SEBALD, J. F., Worthington Corporation
SEIBEL, Dean W., Harvard University
SFORZA, Richard F., New York State Department of Civil Service
SHARP, Catherine G., Educational Testing Service
SHAYCOFT, Marion F., American Institute for Research
SHIMBERG, Benjamin, Educational Testing Service
SIEGMANN, P. T., American Telephone & Telegraph

SILBERMAN, Harry, City College of New York
SIMPSON, Mrs. Elizabeth A., Illinois Institute of Technology
SMITH, Allan B., University of Connecticut
SMILEY, Marjorie B., Hunter College
SMIT, Jo Anne, Washington, D. C.
SMITH, Denzel D., Office of Naval Research
SMITH, Louise, Educational Testing Service
SMITH, Muriel, Educational Testing Service
SOLOMON, Robert, Educational Testing Service
SORENSON, Garth, University of California
SOUTHER, Mary T., Tower Hill School, Wilmington
SPANEY, Emma, Queens College
SPAULDING, Geraldine, Educational Records Bureau
SPEER, George S., Illinois Institute of Technology
SPENCER, Douglas, New York City
STARCH, Daniel, Daniel Starch & Staff
STARCH, Mrs. Daniel, Mamaroneck, New York
STEPHAN, Frederick F., Princeton University
STERNBERG, Carl, Queens College
STEWART, Mary, New York University
STEWART, Naomi, Educational Testing Service
STICE, Glen, Educational Testing Service
STOKES, Thomas M., Metropolitan Life Insurance Company
STONE, Paul T., Huntingdon College
STUART, William A., Educational Testing Service
STULBAUM, Harold, Metropolitan Life Insurance Co.
SULLIVAN, Daniel S., New York State Department of Civil Service
SULLIVAN, Richard H., Educational Testing Service
SUPER, Donald E., Teachers College, Columbia University
SWINEFORD, Frances, Educational Testing Service
SYMONDS, Percival M., Teachers College, Columbia University
TASSO, Charles A., American Cyanamid
TATSUOKA, Maurice, Harvard University
TAYLOR, John F., Woodmere Academy
TAYLOR, Justine N., Educational Testing Service
TEMPLETON, Hugh, New York State Department of Education
TERRAL, J. E., Educational Testing Service
THORNDIKE, Robert L., Teachers College, Columbia University
THURSTONE, L. L., University of North Carolina
TRAXLER, Arthur E., Educational Records Bureau
TRIGGS, Frances, Commission on Diagnostic Reading Tests
TUCKER, Anthony C., Department of Defense
TUCKER, Ledyard R, Educational Testing Service
TUKEY, John W., Princeton University
TUKEY, Mrs. John W., Princeton, New Jersey
TURNBULL, William W., Educational Testing Service
TWYFORD, Loran C., Special Devices Center, ONR
VICKERY, Verna L., Southeastern Louisiana College
VOSS, Harold A., Special Devices Center, ONR
WADELL, Blandena C., World Book Company
WAHLGREN, Hardy L., State Teachers College, Geneseo, New York
WALKER, Helen M., Teachers College, Columbia University
WALLACE, Wimburn, L., Psychological Corporation
WALSH, B. Thomas, Personnel Department, Philadelphia
WALSH, John J., Boston College
WALTON, Wesley, W., Educational Testing Service

WALTZMAN, Hal, New York State Department of Civil Service
WATSON, Walter S., Cooper Union
WELLCK, A. A., University of New Mexico
WELNA, Cecilia T., University of Connecticut
WENZEL, Bernice M., Barnard College
WESMAN, A. G., Psychological Corporation
WHITE, George L., Silver Burdett Company
WHITLA, Dean K., Harvard University
WHITNEY, Alfred G., Life Insurance Agency Management Association
WILKE, Marguerite M., Board of Education, Greenwich
WILKE, Walter H., New York University
WILKS, S. S., Princeton University
WILLARD, Richard W., Harvard University
WILLIAMS, Robert J., Columbia University
WILLIAMS, Roger K., Morgan State College, Baltimore
WILSON, Kenneth M., Princeton University
WILSON, Phyllis C., Queens College
WINANS, S. David, New Jersey Department of Education
WINGO, Alfred L., Virginia State Department of Education
WINTERBOTTOM, John A., Educational Testing Service
WITTENBORN, J. R., Rutgers University
WOLFLE, Dael, American Association for the Advancement of Science
WOOD, Ray G., Ohio State Department of Education
WRIGHT, Wilbur H., Teachers College, Geneseo, New York
YOGMAN, Harvey, University of Denver
ZALKIND, Sheldon S., Brooklyn College
ZUBIN, Joseph, Columbia University

B15R1.5