CHAPTER XII

COMBINING THE SUB-TESTS IN A BATTERY

SELECTION OF SUB-TESTS

The validities of the single tests and the correlations between the tests, determine the effect of the composite score which is derived from five tests. The higher the validities of the single tests and the lower the inter-correla-tions of the tests, the higher is the effect of the composite score.

Bingham[1] suggests,

> In selecting tests to be included in a battery, it is advisable to include those which correlate most closely with the criterion, and least closely with each other. By avoiding duplication in this way, each test used contributes maximally to the forecast.

In chapter, VI, of this treatise, it is discussed in details, how these five tests were constructed and finally selected for the purpose of measuring aptitude for teaching. In table No. 68, the validity coefficients of the single tests and the inter-correlations of these tests are given. It can be seen in the table that the validity coefficients of the single tests are reasonably larger in size than the inter-correlations

---

1 Bingham, W. V. D., "Aptitudes and Aptitude Testing", Harper & Brothers Publishers, New York, 1942, p.220.

of the tests. It can, thus, be said that Bingham's suggestion regarding inclusion of tests in a battery is satisfied to some extent.

THE MEANING OF A TEST-BATTERY

There are various kinds of test-batteries in common use.

One of the most common notions of a battery is an assemblage of several tests, each constructed by a different author at a different time for a somewhat different purpose but brought together by the test user for his own immediate ends. Such a battery is usually assembled "ad hoc" to meet some immediate, specific practical need.

According to Mosier,[1] the term "battery" is conventionally applied to a set of separate tests to be administered to the same group of individuals in order to meet a single measurement objective, or a closely interrelated set of such objectives.

The present test consists of five sub-tests, each sub-test designed to measure a factor indicative of an aptitude for teaching. All the sub-tests are to be administered to the same group of individuals to measure their aptitude for teaching - a single measurement objective. All the sub-tests can, therefore, be assembled to form a battery as defined by Mosier.

---

1 Lindquist, E.F., (Editor), "Educational Measurement", American Council on Education, Washington, D.C., 1955, p.764.

All the components in the battery are reduced to a single frame of reference for the reporting and interpretation of scores.

INEVITABILITY OF COMBINING TEST-
SCORES AND OF WEIGHTING

In almost every situation in which mental measurement is applied, more than one measurement is involved.

Most types of jobs call for a number of specific aptitudes. The teaching profession similarly calls for a specific aptitude. The factors of attitude towards children, mental ability, interest in profession, adaptability, professional information and knowledge of subject-matter enter in a complex combination to determine how successful a teacher will be. It is hardly possible to assess this complex pattern of traits with a single type of test material. Different types of testing techniques are required to assess the several traits making for success in teaching. The introduction of the multiple measures of the same individual or set of individuals, makes it necessary to combine the scores from a number of tests in order to get the best prediction of job success.

The scores, as they have been obtained, cannot be added. Wherever statistical combination is possible, the scores should be so combined as to give the best prediction of the criterion.

Each test as it is revealed by the validity coefficients of the single tests, does not contribute equally in -

predicting the aptitude for teaching. The contributions of different tests are in different proportions. Each test-score should, therefore, be weighted in proportion to its contribution in predicting the criterion score. Thus the assumption to be made is that the criterion score to be predicted is proportional to some weighted sum of the scores on the tests which are combined.

The choice here, as in all instances of multiple measures, is not whether to combine or not, but whether to combine statistically or intuitively. The statistical method of combining test-scores is a scientific one, and to that extent it is accurate and gives better results. The studies thus far made indicate that statistical methods of combination have a distinct advantage in most situations over the informed judgment of a single individual. A statistical method of combining tests is, therefore, applied in the present study. A brief description of different methods of combining tests is given before the method applied here is described in details.

METHODS OF COMBINING TEST-SCORES

Thorndike[1] lists the following four methods of combining the test-scores.

(1) Non-linear function method.

Thorndike, R. L., "Personnel Selection" - Test and Measurement Techniques, John Wiley & Sons, Inc., New York, 1949, pp. 185-201.

(2) Multiple-cut off method.

(3) Clinical method.

(4) Multiple-regression method.

NON-LINEAR FUNCTION METHOD

In this method, the test scores may be combined according to some algebraic function. The algebraic function involved may not be a simple linear function of the single test-scores. The composite score may involve the square or cube of test scores, or the product of tests 1 and 2, or some more complex function. The tests are combined algebraically but not as a simple sum.

MULTIPLE CUT OFF METHOD

A minimum qualifying score, on each one of the tests, is set by the test-constructor. A testee's performance on the test is judged by comparing the score obtained by him on the test with the minimum qualifying score fixed on the test. In this procedure the individual receives no composite score. His test record is merely inspected to see whether he qualifies on all tests, and he is placed into one of two categories, the qualified or the disqualified.

CLINICAL METHOD

This method is based on a clinical evaluation of a set of scores. The clinician makes a judgment of a testee on the basis of the test scores along with other available information about him. The test results are used in a non-mathematical manner, and to that extent it is subjective.

MULTIPLE-REGRESSION METHOD

Except for simple summation or averaging, the statistical method of combining scores most common in use is that of multiple regression. The regression may be linear or non-linear.

The most frequently used approach to the combination of measures for the prediction of an outside criterion is that of multiple linear regression.

The use of multiple regression techniques assumes that the criterion score of an individual can be predicted from a weighted sum of the predictor tests included in the battery. This means that for each test some constant is determined. The constants are in general different for each test, but the same constants are used for every individual. The composite score for each individual is the sum of his test scores, after each test score has been multiplied by its appropriate constant. The algebraic sum of all these weighted scores is equal to the predictor score. Thus an equation is obtained by which a predicted score is obtained if the weighted test scores are available. This equation is called a multiple regression equation. The problem is, then, how to determine these constant multipliers or the regression weights as they are generally called.

Before the different methods to derive the multiple regression equation are discussed, the term 'multiple correlation' should be explained. The correlation between the best weighted composite test scores and the criterion variable is known as the multiple correlation. The multiple correlation is always taken as positive. The multiple correlation serves as an index of the degree to which a test battery is being successful in predicting a criterion. The square of the correlation gives the percentage of variance in criterion score which is predicted by the test battery.

MULTIPLE REGRESSION METHOD APPLIED IN COMBINING PRESENT TEST

The situation, here, is a multiple predictor and a single criterion measure. The usual procedure, therefore, is to derive a multiple-regression equation, with weights that will maximise the correlation between predicted and obtained criterion measures.

The regression equation[1] which expresses the relationship between a single variable, X_1, and any number of independent variables, X_2, X_3, X_4 ------ X_n may be written in score form as follows:

$$\bar{X}_1 = b_{12.34....n}X_2 + b_{13.24....n}X_3 + ...+b_{1n.23..(n-1)}X_n + K.$$

---

1 Garrett, H. E., "Statistics in Psychology and Education", Longmans, Green & Co., New York, 1958, p. 412.

The regression coefficients $b_{12.34....n}$, $b_{13.24...n}$, etc., give the weights to be attached to the scores in each of the independent variables when X_1 is to be estimated from all these in combination.

The multiple regression equation is used mainly for two purposes: (1) analysis and (2) prediction. In analysis, the purpose is to determine the "weight" of each of a number of variables in contributing to the performance, called a criterion

DIFFERENT METHODS USED TO DERIVE THE MULTIPLE-REGRESSION EQUATION

(1) The method based on partial correlation:

According to Garrett,[1] when the independent variables are four or less than four, the multiple-regression equation is set up by way of partial correlation and its accuracy as a predicting instrument is given by the coefficient of multiple correlation.

(2) The method of least squares:

This method has widespread application in experimental science. Thorndike[2] suggests the use of this method in combining test-scores. He says, "

> The above k equations, the "normal" equations, serve to define those values of the B's which give the best linear

---

1 Garrett, H. E., Op.Cit., p.404 and p.410.

2 Thorndike, R.L., Op.Cit., pp.186-190 and Appendix A.

> combination of the separate test-scores - best in the least squares sense, in that the sum of the squares of the differences between predicted success and actual measure of success will be a minimum.

The same author describes the use of the following two methods for solving the set of simultaneous equations.

(1) The first is a compact version of the Doolittle procedure for solving normal equations.

(2) The second is an iterative procedure based on the work of Kelley and Salisbury.

(3) THE WHERRY-DOOLITTLE TEST SELECTION METHOD

When there are 'n' predictor variables, each predictor measuring a different aspect of criterion or introduced as a suppression variable, and one criterion measure, the use of multiple-linear-regression and the Wherry-Doolittle technique to derive it are suggested by Mosier.[1] There are five predictor variables and one criterion variable in the present problem, and therefore, following Mosier's suggestion, the Wherry-Doolittle test selection method is applied here.

In this method, the most efficient team of tests is selected one at a time from a large number. It selects the tests of the battery analytically and adds them one at a time until a maximum multiple-correlation is obtained. The Wherry-

---

1 Lindquist, E. F., Op.Cit., Chapter No.18, Table No. 15, and page No. 790.

Doolittle technique also provides a method of solving certain types of multiple correlation problems with a minimum of statistical labour.

In applying this technique to the present problem, the procedural steps suggested by Garrett[1] were followed in toto. The computational details are not given. Only the tables showing different results obtained are given in the following pages.

The Wherry-Doolittle method was applied to serve the following three purposes at a time.

(i) To select those tests which yield a maximum R with the criterion and discard the rest,

(2) To calculate the multiple R after the addition of each test, stopping the process when R no longer increases,

and (3) To compute a multiple regression equation from which the criterion can be predicted with the highest precision which the given list of tests is capable of.

---

1 Garrett, H. E., Op.Cit., pp. 426-437.

TABLE NO. 68

INTER-CORRELATIONS OF FIVE TESTS AND A CRITERION

List of Tests (N = 358)

C = Criterion

I = Attitude towards children

II = Mental Ability

III = Interest in Profession

IV = Adaptability

V = Professional Information

	I	II	III	IV	V
C	0.380	0.383	0.100	0.316	0.237
I	-	0.1852	0.1222	0.2923	0.2054
II	-	-	0.0739	0.2042	0.1811
III	-	-	-	0.1534	0.0691
IV	-	-	-	-	0.1459
V	-	-	-	-	-

TABLE NO. 69

VALUES OF 'V' FOR DIFFERENT TESTS

	Tests				
	I	II	III	IV	V
V_1	-0.3800	-0.3830	-0.1000	-0,3160	-0.2370
V_2	-0.3091	〃	-0.0717	-0.2378	-0.1676
V_3	〃	〃	-0.0370	-0.1564	-0.1126
V_4	〃	〃	-0.0178	〃	-0.1014
V_5	〃	〃	-0.0147	〃	〃

TABLE NO. 70

VALUES OF 'Z' FOR DIFFERENT TESTS

	Tests				
	I	II	III	IV	V
Z_1	1.0000	1.0000	1.0000	1.0000	1.0000
Z_2	0.9657	〃	0.9945	0.9583	0.9672
Z_3	〃	〃	0.9823	0.8912	0.9366
Z_4	〃	〃	0.9688	〃	0.9321
Z_5	〃	〃	0.9679	〃	〃

COMPUTING THE VALUES OF $\frac{Vm^2}{Zm}$ FOR DIFFERENT TESTS

(i) $\frac{V_1^2}{Z_1} = \frac{(-0.3830)^2}{1.0000} = 0.1467$

(ii) $\frac{V_2^2}{Z_2} = \frac{(-0.3091)^2}{0.9657} = 0.0989.$

(iii) $\frac{V_3^2}{Z_3} = \frac{(-0.1564)^2}{0.8912} = 0.0274.$

(iv) $\frac{V_4^2}{Z_4} = \frac{(-0.1014)^2}{0.9321} = 0.0110.$

(v) $\frac{V_5^2}{Z_5} = \frac{(-0.0147)^2}{0.9679} = 0.0002.$

TABLE NO. 71

SHOWING VALUES OF "SHRUNKEN" MULTIPLE CORRELATION COEFFICIENT FOR EACH TEST AND ALSO SHOWING WHETHER THE TEST IS RETAINED

a	b	c	d	e	f	g	h
m	$\frac{V_m^2}{Z_m}$	K^2	$\frac{N-1}{N-m}$	$\bar{K}^2$	$\bar{R}^2$	$\bar{R}$	Test # (no. of the selected test)
0	-	1.0000	N = 358	-	-	-	-
I	0.1467	0.8533	1.0000	0.8533	0.1467	0.3830	II
II	0.0989	0.7544	1.0030	0.7567	0.2433	0.4933	I
III	0.0274	0.7270	1.0060	0.7314	0.2686	0.5183	IV
IV	0.0110	0.7160	1.0080	0.7217	0.2783	0.5276	V
V	0.0002	0.7158	1.0110	0.7237	0.2763	0.5256	III *

* Since test III does not increase the multiple correlation coefficient, it is not selected.

"Shrunken" multiple correlation coefficient, $\bar{R}$, was calculated by the following Wherry shrinkage formula:

$$\bar{R}^2 = 1 - K^2 \frac{(N-1)}{(N-m)}$$

Where $\bar{R}$ is the "shrunken" multiple correlation coefficient.

TABLE NO. 72

A WORK SHEET PREPARED TO AID IN THE SELECTION OF TESTS TO BE ADDED TO OUR BATTERY

	I	II	III	IV	V	-C	Check sum	Test 77
a_1	-	-	-	-	-	-	-	-
b_1	0.1852	1.0000	0.0739	0.2042	0.1811	-0.3830	1.2614	II
c_1	-0.1852	-1.0000	-0.0739	-0.2042	-0.1811	+0.3830	-1.2614	
a_2	1.0000	0.1852	0.1222	0.2923	0.2054	-0.3300	1.4251	
b_2	0.9657	x	0.1085	0.2545	0.1719	-0.3091	1.1915	I
c_2	-1.0000	x	-0.1124	-0.2635	-0.1780	+0.3201	-1.2338	
a_3	0.2923	0.2042	0.1534	1.0000	0.1459	-0.3160	1.4798	
b_3	x	x	0.1097	0.8912	0.0636	-0.1563	0.9083	IV
c_3	x	x	-0.1231	-1.0000	-0.0714	+0.1754	-1.0192	
a_4	0.2054	0.1811	0.0691	0.1459	1.0000	-0.2370	1.3645	
b_4	x	x	0.0286	x	0.9321	-0.1015	0.8591	V
c_4	x	x	-0.0307	x	-1.0000	+0.1089	-0.9218	

CALCULATION OF THE MULTIPLE REGRESSION EQUATION

TABLE NO. 73

'C' ENTRIES FOR THE FOUR SELECTED TESTS
AND FOR THE CRITERION

	Tests - in order of selection				
	II	I	IV	V	-C
C_1	-1.0000	-0.1852	-0.2042	-0.1811	0.3830
C_2	-	-1.0000	-0.2635	-0.1780	0.3201
C_3	-	-	-1.0000	-0.0714	0.1754
C_4	-	-	-	-1.0000	0.1089

When equated to zero, each row in the above table is an equation defining the beta weights. The equations for obtaining β's can, therefore, be written as under:

$$-1.0000\beta_2-0.1852\beta_1-0.2042\beta_4-0.1811\beta_5+0.3830 = 0 \quad \text{..........(A)}$$

$$-1.0000\beta_1-0.2635\beta_4-0.1780\beta_5+0.3201 = 0 \quad \text{..........(B)}$$

$$-1.0000\beta_4-0.0714\beta_5+0.1754 = 0 \quad \text{..........(C)}$$

$$-1.0000\beta_5+0.1089 = 0 \quad \text{..........(D)}$$

From equation, D, we have:

$$\beta_5 = 0.109$$

Substituting this value of β_5, in equation, C, we have:

$$-1.0000\beta_4 - 0.0714 \times 0.109 + 0.1754 = 0$$

$$\therefore \beta_4 = 0.168$$

Again, substituting values of β_5 & β_4, in equation, B, we have:

$$-1.0000\beta_1 - 0.2635 \times 0.168 - 0.1780 \times 0.109 + 0.3201 = 0$$

$$\therefore \beta_1 = 0.256.$$

And lastly, substituting values of β_5, β_4 & β_1 , in equation, A, we have:

$$-1.0000\beta_2 - 0.1852 \times 0.256 - 0.2042 \times 0.168 + 0.1811 \times 0.109 + 0.3830 = 0$$

$$\therefore \beta_2 = 0.282.$$

$\therefore$ The multiple regression equation in σ- score form is:

$$\bar{Z}_c = 0.282Z_2 + 0.256Z_1 + 0.168Z_4 + 0.109Z_5.$$

To enable any test-user to use the raw scores on tests without converting them into standard scores, the above obtained multiple-regression equation in σ- score form is written in (raw) score form. The means and SD's of the criterion and of the tests are given in the following tables. They are, then, used to convert the equation which is in σ-score form to the equation in raw score form.

TABLE NO. 74

MEAN & STANDARD DEVIATION OF THE CRITERION

	N	M	SD
Criterion, 'C' * (Total marks on 'Varsity exam,- Total of Part I & Part II marks)	358	54.72	4.9

* The total marks of each testee from each unit were converted in percentages and the raw (per cent) score was used to calculate M. & SD.

TABLE NO. 75

MEANS & SD'S OF TESTS, CALCULATED FROM RAW SCORES

Sub-Test	N	M	SD
I	358	22.72	2.46
II	358	18.94	4.29
III	358	14.93	3.40
IV	358	10.32	2.87
V	358	12.13	3.30

To write the regression equation in score form the B's are transformed into b's in the following way:

(i) $b_2 = \beta_2 \times \frac{\sigma_c}{\sigma_2} = 0.282 \times \frac{4.9}{4.29} = 0.322$

(ii) $b_1 = \beta_1 \times \frac{\sigma_c}{\sigma_1} = 0.256 \times \frac{4.9}{2.46} = 0.510$

(iii) $b_4 = \beta_4 \times \frac{\sigma_c}{\sigma_4} = 0.168 \times \frac{4.9}{2.87} = 0.287$

(iv) $b_5 = \beta_5 \times \frac{\sigma_c}{\sigma_5} = 0.109 \times \frac{4.9}{3.3} = 0.162$

$\therefore$ The multiple regression equation in terms of raw scores is:

$$\bar{X}_c - 54.72 = 0.322\ (X_2 - 18.94) + 0.510\ (X_1 - 22.72) + 0.287\ (X_4 - 10.32) + 0.162\ (X_5 - 12.13)$$

$$\therefore\ \bar{X}_c = 0.322\ X_2 + 0.510\ X_1 + 0.287\ X_4 + 0.162\ X_5 + 32.107$$

This may be written as:

$$\bar{X}_c = 0.32\ X_2 + 0.51\ X_1 + 0.29\ X_4 + 0.16\ X_5 + 32.11$$

Where $\bar{X}_c$ = predicted criterion score (in percentage),

X_2 = raw score on test II,

X_1 = raw score on test I,

X_4 = raw score on test IV,

X_5 = raw score on test V,

and 32.11 is a constant value.

SE of $\bar{X}_c$

This is given by the following formula:

$$\sigma_{est}\ \bar{X}_c = \sigma_c \sqrt{1 - \bar{R}_c^{\,2}\ (2145)}$$

$$= 4.9 \sqrt{1 - 0.2783}$$

$$= 4.9 \sqrt{0.7217}$$

$$= 4.9 \times .85$$

$$= 4.16$$

The 0.95 confidence interval is $\bar{X}_c + 1.96 \times 4.16$ or $\bar{X}_c \pm 8$.

Mosier[1] says,

> If, and only if, the several predictor scores and the criterion are normally distributed and the regression is linear, then the average discrepancy (predicted-obtained) will be the same throughout the range.

This is the prediction equation finally established to predict an aptitude for teaching possessed by a testee.

The present aptitude-test - battery includes, therefore, the following four tests.

---

1 Lindquist, E. F., Op.Cit., p. 786.

Sr. No.	Test No.	Test	Raw score	Multi-plier	Weighted score
1	I	Attitude Towards Children	X_1	0.51	$0.51X_1$
2	II	Mental Ability	X_2	0.32	$0.32X_2$
3	IV	Adaptability	X_4	0.29	$0.29X_4$
4	V	Professional Information	X_5	0.16	$0.16X_5$

Test III - Interest in Profession - should be discarded from the final test-battery. But the test is not summarily rejected for the reasons discussed in chapter XV. More about this test is discussed in the same chapter.

CHECKING THE β WEIGHTS AND MULTIPLE R

$$R^2 = \beta_2 r_{c2} + \beta_1 r_{c1} + \beta_4 r_{c4} + \beta_5 r_{c5}$$

$$= 0.282 \times 0.383 + 0.256 \times 0.380 + 0.168 \times 0.316 + 0.109 \times 0.237$$

$$= 0.2842$$

$$\therefore R = 0.5331.$$

Also $1 - K^2 = 1 - 0.7160 = 0.2840$

(K^2 is taken from row 4 of column C in table No. 71 on page No. 368).

Since R^2 and $(1 - k^2)$ are almost equal, the β weights and multiple R tally very closely.

SHRUKEN $\bar{R}$

The "Shruken" $\bar{R}$, adjusted for chance errors, is given by:

$$\bar{R}^2 = \frac{(N - 1)R^2 - (m - 1)}{(N - m)}$$

Where R^2 is square of multiple R calculated above and is 0.2842.

$N = 358$, $m = 4$ = no. of selected tests.

$$\therefore \bar{R}^2 = \frac{357 \times 0.2842 - 3}{354}$$

$$= 0.2781$$

$$\therefore \bar{R} = 0.5274.$$

This value of $\bar{R}$ (0.5274) exactly tallies with that (0.5276) obtained in row 4 of column 'g' in table No. 71 on page No. 368.

In the present problem, shrinkage in multiple R is quite small (0.5331 - 0.5274 = 0.0057) as the sample is fairly large and there are only four tests in the multiple regression equation.

The validity coefficient, r, obtained (in the previous chapter) is 0.5020 and the multiple $\bar{R}$ obtained here is 0.5274. Thus there is an increase in validity of the test by 0.0254

(0.5274 - 0.5020) i.e. about 5.1 per cent when the separate tests are properly weighted and combined in a battery.

> Few investigators have employed multiple and partial correlations. A higher predictive value has been obtained from multiple correlations when a number of the factors condition-ing success in teaching were combined. Hellfritzsch and other Wisconsin studies of Barr and others illustrate this point.[1]

SOME LIMITATIONS TO MULTIPLE-REGRESSION EQUATIONS

According to Guilford,[2] perhaps there are still some psychologists who would prefer to combine predictive measures in an intuitive fashion in predicting a criterion. Studies in which this approach has been used as compared with the multiple regression approach almost invariably show superior-ity for the latter.

The multiple regression approach has its limitations, however. Cureton[3] says,

> It is doubtful that any other statistical techniques have been so generally and widely misused and mis-interpreted in educational research as have those of multiple correlation.

It must be admitted that there are pitfalls in the way of the

---

1 Monroe, W.S. "Encyclopedia of Educational Research", The MacMillan Company, New York, p. 1393.

2 Guilford, J. P., "Psychometric Methods", McGraw-Hill Book Company, Inc., New York, 1954, pp. 403-404.

3 Lindquist, E.F., Op.Cit., p.690.

user of multiple-regression methods. But it should also be admitted that under the appropriate conditions their use can be very effective.

Some of the common limitations are discussed in brief below.

(1) First to be considered is that these methods rest on the assumption of linear regressions ($y = ax_1 + bx_2 + \ldots\ldots\ldots$) among the measures going into the equation.

As Mosier[1] points out, when the entire range of ability is taken into account, the regressions of a criterion on scores of ability are probably quite commonly non-linear. But the same author[2] again says,

> In many extensive investigations, particularly in the armed forces during the war, non-linear relations between test and criterion were sought and seldom found.

Considering the conclusion reached from these extensive investigations and a suggestion from Guilford that even if some regressions are non-linear, it is possible by transformations to reduce them to rectilinear form and then use multiple-regression methods, it can safely be assumed that the present combination is a linear one.

(2) After three or four of the most valid tests have

---

1 Lindquist, E.F., Op.Cit., p. 785.

2 Lindquist, E. F., Op.Cit., p. 786.

been combined in an equation to predict a criterion, adding more tests rarely improves prediction. This is said to be due to the phenomenon of linear restraint.

Cureton[1] recommends that multiple-regression procedures be applied after removing as many of the restraints as possible.

There are only four selected tests in the present battery and so a complete absence of linear restraint can justifiably be assumed.

(3) Perhaps the chief limitation to R, the coefficient of multiple correlation, is the fact that, since it is always positive, variable errors of sampling tend to accumulate and thus make the coefficient too large.

The shrinkage formula for correcting an inflated R, was applied to the R obtained for the present battery and it was found that the difference between R and $\bar{R}$ was quite negligible suggesting that the variable errors of sampling did not affect the present R.

---

1 Lindquist, E. F., Op.Cit., p. 692.