PSYCHOMETRIKA—VOL. 15, NO. 3
SEPTEMBER, 1950

THE PROBLEM OF CLASSIFICATION OF PERSONNEL*

ROBERT L. THORNDIKE
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

The personnel classification problem arises in its pure form when all job applicants must be used, being divided among a number of job categories. The use of tests for classification involves problems of two types: (1) problems concerning the design, choice, and weighting of tests into a battery, and (2) problems of establishing the optimum administrative procedure of using test results for assignment. A consideration of the first problem emphasizes the desirability of using simple, factorially pure tests which may be expected to have a wide range of validities for different job categories. In the use of test results for assignment, an initial problem is that of expressing predictions of success in different jobs in comparable score units. These units should take account of predictor validity and of job importance. Procedures are described for handling assignment either in terms of daily quotas or in terms of a stable predicted yield.

The past decade, and particularly the war years, have witnessed a great concern about the classification of personnel and a vast expenditure of effort presumably directed towards this end. In all branches of the military establishment were found "general classification" tests or test batteries planned to serve a classification function. Since the war the number of published test batteries designed for differential prediction has rapidly multiplied. It seems timely, therefore, to look into the problem of the classification of personnel to see what the concept means, what issues it raises with respect to the theory of measurement, and what problems it presents with respect to the practical operation of a testing program.

It must be indicated that much of the present discussion represents an examination of concepts, a raising of questions, and an offering of intuitive suggestions, rather than a presentation of mathematically established answers. The defining of questions represents a first step in answering them. It is hoped that clarification of the problems and issues in the following pages may stimulate others to solve them.

Personnel classification, as the term is used here, is best de-

*Address of the President of the Division on Evaluation and Measurement of the American Psychological Association, delivered at Denver, Colorado, September 9, 1949.

fined by contrasting it with personnel selection. In the pure case of personnel *selection* we deal with a single job category, we have a limited number of job vacancies and a surplus of job applicants, and our problem is to pick the most promising of the applicants to fill the vacancies. In the pure *classification* program, by contrast, we are concerned simultaneously with a number of job categories, we have no more workers than there are jobs to be filled, and our problem is to decide *which* job shall be done by *which* individual. An example of a pure *selection* situation is that faced by a medical school which has 1000 applicants for admission and wishes to pick from this group the 100 who may be expected to succeed best in the curriculum of the school or in the professional duties of a doctor. The pure *classification* situation is most nearly approached in the military establishment, where a large flow of untrained youths continually pours into the organization and must be channelled into dozens of different types of specialized training and work, and where everyone who meets minimum screening standards must be used in some capacity. The simplest paradigm of the *selection* situation is a follows: We have a job vacancy X and applicants A, B, and C. Which individual should get the job? Reduced to the simplest form, the *classification* problem may be expressed: We have a vacancy in each of three jobs, X, Y, and Z and we have three applicants A, B, and C. Which applicant should be put to work in which job?

In practice, of course, selection and classification occur not only in pure form, but also mixed. Thus if we have three jobs, X, Y, and Z, and add to the three applicants A, B, and C a fourth applicant D, our problem now involves elements of both classification and selection. We must not only sort out our applicants into the several job categories, but also reject the least promising. The mixed case will be fairly commonly found in practical personnel situations. The emphasis will vary from one situation to another, with now selection and now classification dominating. Before we can hope to understand the mixed case, however, we must try to understand the relatively simpler pure case. For that reason, most of what is said here will deal with the problems involved in a pure case of personnel classification.

In the past fifty years, the procedures and statistical rationale for selection have been fairly fully worked out. The cornerstone of that rationale is the multiple regression equation, through which a series of scores for the individual may be combined in a linear expression which will yield a maximally accurate prediction of some criterion of job success. Though plenty of problems remain with respect to the detailed application of multiple regression techniques to the development of a battery of prediction tests, choice from among them, and combination of the tests into a battery, the main outlines for obtaining maximum prediction of a single job criterion are clear and are familiar to every student of tests. In the case of the classification problem, however, no such mathematically best solution has been formulated. (It is, of course, possible that no solution exists.)

Before we can hope for a solution to the problem of classification, we must see to it that the problem is precisely formulated and clearly stated. What, in its essence, is the problem? Reduced to its basic elements and to its pure form, it may be stated as follows:

Given: A set of k job categories with N vacancies to be filled ($N \geqq k$), and N individuals to be used in filling them.

Required: To assign the individuals to the jobs in such a way that the average success* of all the individuals in all the jobs to which they are assigned will be a maximum.

TABLE 1
Aptitudes of Three Individuals for Three Jobs

Individual	*Job A*	*Job B*	*Job C*
I	55	60	65
II	50	50	55
III	45	50	45

We can illustrate this by the example of three men and three jobs presented in Table 1. Suppose that those represent perfectly valid measures of aptitudes of the three men for the three jobs, all scores being expressed in standard scores of some reference population. There are, of course, six permutations of assignment of the men to the jobs, and examination of these quickly shows that the sum of the aptitude scores is a maximum when I is assigned to Job *C*, II to Job *A*, and III to Job *B*. The sum is then 165. It is not necessary that the jobs be equally weighted. Thus, if it were especially important to have the best talent in Job *B*, for example, that job might be given triple weight. The maximum of $A + 3B + C$ is obtained by assigning individual I to Job *B*, II to *C*, and III to *A*.

*The term "success", as it is used in this paper, may be interpreted quite broadly to include measures of job satisfaction as well as ratings of performance or measures of production.

This miniature example shows many of the essential characteristics of the classification problem. The problem is one of simultaneous assignment of all individuals, because the assigning of one man can only be made with reference to that of others. The separate decisions are not independent. Typically, some individuals must be assigned to jobs other than the ones for which they have the most aptitude. (In this illustration, Individual II had in one case to be assigned to Job *A* rather than Job *C*.) *Differences* in level of aptitude within the individual emerge as a dominant factor in assignment. Thus, in one case Individual III was assigned to job *A*, although his aptitude for that job was lower than that of either of the other men, because he had no higher aptitude for any other job.

There are, as has been indicated, a finite number of permutations in the assignment of men to jobs. When the classification problem as formulated above was presented to a mathematician, he pointed to this fact and said that from the point of view of the mathematician there was no problem. Since the number of permutations was finite, one had only to try them all and choose the best. He dismissed the problem at that point. This is rather cold comfort to the psychologist, however, when one considers that only ten men and ten jobs mean over three and a half million permutations. Trying out all the permutations may be a mathematical solution to the problem, it is not a practical solution.

The classification enterprise involves two quite distinct groups of problems. One group centers around the development and choice of the tests which are to comprise the classification battery. The other group concerns procedures for *using* the test scores in classification, given a particular battery of predictors. This matter of finite number of permutations, referred to in the previous paragraph, has to do with the second group of problems — those of using the information from a set of aptitude predictors. We will return to this part of the problem a little later. For the present, however, let us turn our attention to the problems involved in developing a battery of tests for use in classification.

Selection of Tests for a Classification Battery

We may ask: What attributes should a test have if it is to be a useful member of a classification test battery? What should be the joint characteristics of a set of tests which are to form a battery to be used for classification?

Perhaps the best way to approach these questions is to consider the pseudo-classification problem of dividing men between two jobs. This is termed a pseudo-classification problem because the limitation to two job categories permits us to direct our attention to differences in aptitude for the two jobs and thus to reduce the problem to one of selection. That is, we are able to *select* men on the basis of a single score representing difference in aptitude. When there are three or more jobs, we have a number of difference scores and are thrown back upon the true classification problem. However, the solution of the two-category problem serves to point up for us the qualities which we shall need to look for in tests and in test batteries for the more general classification problem with three or more jobs.

For each of two job categories, A and B, we can make our best prediction, in the least squares sense, of further success on the job. These may be expressed as

$$\tilde{y}_A = \beta_{1A} z_1 + \beta_{2A} z_2 + \cdots + \beta_{KA} z_K ,$$

and (1)

$$\tilde{y}_B = \beta_{1B} z_1 + \beta_{2B} z_2 + \cdots + \beta_{KB} z_K ,$$

where β_{iA} is the weight to be applied to standard scores in variable i is predicting success in Job A and β_{iB} is the weight for Job B. Now what we are currently interested in is whether the individual is likely to be more successful in Job A or in Job B. That is, we are interested in a prediction of difference in success. This is, of course, the simple difference* between our two predictions. Let us call this $\tilde{\Delta}$, where

$$\tilde{\Delta} = \tilde{y}_A - \tilde{y}_B . \tag{2}$$

For purposes of assignment, individuals could be arranged in rank order with respect to $\tilde{\Delta}$, and a dividing line set which would assign the required number to each job.

Referring to Equations (1), and performing the subtraction indicated in Equation (2), we get

$$\tilde{\Delta} = (\beta_{1A} - \beta_{1B}) z_1 + (\beta_{2A} - \beta_{2B}) z_2 + \cdots + (\beta_{KA} - \beta_{KB}) z_K . \tag{3}$$

We see, then, that the weight which a predictor receives in predicting this difference score is the difference between its weights for the

*For the present, we are considering each of the two jobs to be equally important, and are dealing with the simple difference. It would also be possible to deal with a weighted difference, thereby attaching greater importance to one of the jobs than to the other.

two separate job categories. Under what circumstances will it receive a large differential weighting?

Let us assume that we have only two predictors, 1 and 2. In this case, we find that the differences in weight are given by the formulas

$$\beta_{1A}-\beta_{1B}=\frac{(r_{1A}-r_{1B})-r_{12}(r_{2A}-r_{2B})}{1-r_{12}^2}$$

and

$$\beta_{2A}-\beta_{2B}=\frac{(r_{2A}-r_{2B})-r_{12}(r_{1A}-r_{1B})}{1-r_{12}^2}. \qquad (4)$$

An inspection of these formulas reveals that, ordinarily, a test will receive a substantial differential weight if it (a) has a substantial difference in validity for the two criteria in question and (b) does not have a high positive correlation with other tests which differentiate in the same direction or a high negative correlation with other tests which differentiate in the reverse direction. Occasionally a test may be found which receives a substantial differential weight in spite of having the same validity for both jobs, because it has a substantial correlation with another variable which *does* have a sharp validity difference between the two job categories. This is analogous to the suppression variable in the usual selection problem, in which a variable with near zero validity may receive a substantial negative weight if it has a high correlation with some variable or variables with high validity.

TABLE 2
Relation of Differential Weights and Differential Validity to Test Validities and Intercorrelation

	Example					
	1	*2*	*3*	*4*	*5*	*6*
r_{1A}*	.30	.30	.30	.30	.30	.20
r_{1B}	.30	.10	.10	.10	.00	.20
r_{2A}	.40	.15	.15	.15	.00	.20
r_{2B}	.40	.45	.45	.45	.45	.60
r_{12}	.50	.80	.40	.00	.00	.60
r_{AB}	.00	.00	.00	.00	.00	.00
$\beta_{1(A-B)}$†	.00	1.22	.39	.20	.30	.38
$\beta_{2(A-B)}$	.00	—1.28	—.46	—.30	—.45	—.62
$R_{(A-B)\widetilde{(A-B)}}$‡	.00	.54	.33	.26	.38	.45

*r_{1A}, r_{1B}, r_{2A}, r_{2B} are correlations of Tests 1 and 2 with job criteria *A* and *B* respectively.
†$\beta_{1(A-B)}$ and $\beta_{2(A-B)}$ are the regression weights for differential prediction.
‡$R_{(A-B)\widetilde{(A-B)}}$ is the correlation between actual aptitude difference and predicted aptitude difference.

Several examples of the weights which variables may receive as differential predictors are given in Table 2, and serve to illustrate the interaction of validities and intercorrelations.

The first example is a pair of tests both of which have the same validities for both criteria. In this case, of course, neither test has any differential validity, neither receives any weight as a differential predictor, and the validity of the differential prediction is exactly zero. The last row of the table is a row of validities for differential prediction. We will turn to the formula for this value shortly.

Examples 2, 3, and 4 all involve the same set of validities, but different correlations between the two tests. Here Test 1, which is more valid for Criterion *A*, receives a positive weight; while Test 2, which is less valid for *A*, gets a negative weight. With a high positive correlation between the two tests, the weights are high and the differential validity of the pair high. This corresponds to the (highly improbable) situation in a selection battery of finding a pair of tests, one with positive and one with negative validity, which have a high positive correlation. As the correlation between the tests drops, their differential weights drop; and their validity for differential prediction drops also.

The comparison of Examples 4 and 5 brings out the importance for differential prediction of getting tests which, while valid for one variable, have zero or negative validity for another. The drop in validity of Test 1 for Criterion *B* from .10 to .00 and of Test 2 for Criterion *A* from .15 to .00 raises the validity of the pair of tests for differential prediction from .25 to .38 with the existing pattern of test and criterion correlations.

Example 6 illustrates the suppression variable in the context of differential prediction. Test 1, which has equal validity for both criteria, receives a positive weight in differential prediction, because of its substantial correlation with Test 2, which is much less valid for Criterion *A* than for Criterion *B*.

Table 2 probably directs undue attention to the correlation between the test variables, because of the way the illustrations were selected. Usually it is likely to be the difference in validity which is the critical matter. The table suggests that in appraising a test for addition to a *classification* battery we should be as vitally concerned that it have vanishing validity for some job categories as that it have high validity for others.

We might turn next to a consideration of the expression for the validity of $\tilde{\Delta}$, our differential prediction. This can be obtained from the familiar formula for the correlation of sums and differences. We let

$\tilde{A}$ stand for composite score predicting success in Job A,
A stand for the actual success in Job A,
$\tilde{B}$ stand for composite score predicting success in Job B, and
B stand for actual success in Job B.

Then $(\tilde{A} - \tilde{B}) = \tilde{\Delta}$ is the predicted difference in success on the two jobs, and $\underline{(A - B)}$ is the actual difference in success. The validity of the differential prediction is

$$R_{(A-B)(\widetilde{A-B})} = \sqrt{\frac{R^2_{A\tilde{A}} + R^2_{B\tilde{B}} - 2R_{A\tilde{A}}R_{B\tilde{B}}r_{\tilde{A}\tilde{B}}}{2(1 - r_{AB})}}. \qquad (5)^*$$

This is the formula which was used for computing the values in the last row of Table 2.

The relationships involved in Formula (5) are brought out more clearly in the special case in which $R_{A\tilde{A}} = R_{B\tilde{B}}$. The formula then becomes

$$R_{(A-B)(\widetilde{A-B})} = \sqrt{\frac{R^2_{A\tilde{A}}(1 - r_{\tilde{B}\tilde{B}})}{(1 - r_{AB})}}. \qquad (6)$$

We can see that differential validity depends upon three considerations. Other things being equal, high validity of differential prediction will result when (1) the validity of prediction for the separate jobs is high, (2) the correlation between the weighted composite scores for predicting success in the two jobs is low or, better still, negative, and (3) the correlation of true success on the two jobs is high.

Only the first and second of the above considerations relate to the qualities of our test battery; the third represents a constant when we are working with a particular pair of jobs. If valid differential prediction is to be achieved, the two equally important considerations are that separate validities be high and that the intercorrelations of the separate score composites be low. Neglecting the case of the suppression test, which is rarely met in practice, we can say that the first condition will be satisfied insofar as for each of the jobs there are one or more tests which have high validity, and insofar as

*This formula, which replaces an erroneous formula included in the original paper, was derived by William G. Mollenkopf. Its development is presented in Research Bulletin 50-9, Educational Testing Service, Princeton, N. J.

the correlations between the tests are low. The second condition will be achieved insofar as the tests which have a high weight for one job have low or zero weights for others, and insofar as the correlations between tests are low. Our objective, then, is a battery of tests in which each test has high validity for one or two jobs but has near zero validity for the others, and in which the intercorrelations of the separate tests are low.

Under what circumstances is a test likely to approach such a validity pattern? One would judge that maximum simplicity, purity, and univocality of the function measured is the governing condition. If we conceive of abilities as having primarily either positive or zero validity for a job, then the fewer abilities a test taps the greater is the number of job categories for which it may have zero validity and the greater is the value it can have for purposes of classification. The classification situation seems to be one in which the simple, factorially pure test comes into its own. In the creative and research activity underlying our test construction, then, efforts should be directed to measuring as many attributes of human behavior in as pure form as possible.

Suppose now that we have a large pool of tests, from which we wish to select a limited number to constitute a classification battery. How shall we go about selecting the optimum pool of tests? Here again, for a simple selection program, the enterprise should be relatively straightforward. We should undertake simply to select those with the largest regression weights. In the case of a classification program, our problem is a good deal more complex. Our objective is to be able to differentiate aptitude for any one of the jobs or job families into which we are classifying individuals, from aptitude for any other one. In our thinking we can perhaps best approach this in terms of the hypothetical pure factors of factor analysis. If we had knowledge of the factor composition of job success for each of the jobs in which we are interested, we could specify what we would like to have in the tests of our battery. We would ideally like a highly reliable pure test of each of the factors which appears in certain job criteria but not in others. Of secondary desirability would be tests of factors appearing with very different loadings in the different job criteria. Such a battery would permit maximally valid differential prediction of success in the different jobs.

Though it may be relatively easy to recognize what would constitute the ideal for a classification program, it is much more difficult to provide guidance for selection from among a pool of non-ideal existing tests. Where the tests possess varying degrees of factorial purity, varying levels of validity, varying degrees of correspondence in factorial pattern, any rigorous rules for selecting one particular set of tests will be very difficult to formulate, and no such rules are offered at the present time. The problem is further complicated by the fact that in practice we can rarely limit our activities to *pure* classification, because an element of selection is often included. It will often be necessary to compromise between tests which are outstanding in differential validity and tests which are high in general validity for a wide range of jobs. In application, then, the element of practical judgment will continue to bulk large in our choice of tests to constitute a battery. We can only hope that that judgment may be provided with a better set of guiding principles in the future than it has had in the past.

Use of Test Scores to Accomplish Classification

Now let us assume that the battery of tests which we are going to use in our classification program is fixed, and inquire how we shall use the battery of tests to accomplish the assignment of men.

First of all, we will want to combine single test scores into weighted composite scores, one for each job. Except in the special case of two job categories which we considered earlier, in which it is possible to direct our efforts toward the prediction of a single *difference* in aptitude between the two jobs, I see no escape from the step of predicting success in each of the several jobs singly.* Thus, for each job category we will need just the same weighted composite of test scores which we would require if our problem were merely to *select* persons for that job on the basis of the battery of tests which we have assembled.

Our next need is to express these score composites for the different jobs in units which will make it most convenient to compare the individual's probable success in the different jobs. Clearly we need some single type of standard score scale for all of the job specialties.

One possibility is to take the composite aptitude scores for each job, just as they stand, and transform them into standard scores with the same mean and standard deviation. Quite possibly we may wish to normalize the distribution at the same time. In the interests of concreteness, let us suppose that we are going to use a mean of

*Dr. P. J. Rulon has recently reported informally on the development of procedures for computing a multiple discriminant function which may eliminate the need for predicting success for separate job categories. The full report of this method will be awaited with interest.

50 and a standard deviation of 10. Then for the standard reference population, the distribution of prediction scores for each job category will have a mean of 50 and a standard deviation of 10. This general type of score is very familiar, and has been widely used in testing programs of all sorts. When each test score or score composite is to be used in connection with a number of jobs in the same job family, it may represent as serviceable a type of score as it is possible to prepare. When, however, each of several score composites is to be used to select men for *one specific job*, the usual standard score units seem to have two limitations. First, they do not take any account of the differences in validity of the different score composites. Second, they do not take any account of differences in importance of the several jobs. Let us examine these two points.

Suppose, as an extreme example and for purposes of clarification, we have score composites for Jobs *A* and *B* which have validities (against equally relevant and reliable criteria of success) of .20 and .80 respectively. Suppose we have a job applicant who is one standard deviation above the mean on each of these score composites. What does this mean with respect to his probable success in the two jobs? In the first job, where the predictor has a validity of .20, our best estimate is that the applicant will be two-tenths of a standard deviation above the mean in job success. In the second job, by contrast, we may expect he will be eight-tenths of a standard deviation above the mean. The standard score of 60 will have very different significance in these two cases. It will stand for sharply different levels of expected success.

We may think of true or actual achievement as consisting of two components, one which is predicted by our score composite and one which is not. Similarly, variance in true achievement consists of two parts, predictable variance and non-predictable variance. The predictable variance will be R^2 times the total variance, where R is the validity of the test composite. It is proposed that more meaningful comparability of score units for different job categories is achieved when the variance is made equal for scores representing actual achievement in the various jobs. The variance of the composite predictor scores will then be proportional to R^2, and the standard deviations proportional to R. In the example, where Composite *A* had a validity of .20 and Composite *B* one of .80, the two score distributions should not have equal standard deviations, but their standard deviations should stand in the ratio of .20 to .80. Thus, if the standard deviation of true achievement were set equal to 10 for each job, the standard deviation of composite scores would be 2 for

Job A and 8 for Job B. Thus, our applicant who fell one standard deviation above the mean in both aptitude composites would receive a converted score of 52 for Job A and one of 58 in Job B. These scores would correspond to our best prediction that he would fall two-tenths of a standard deviation above the mean in success on Job A and eight-tenths of a standard deviation above on Job B.

This type of score conversion seems to facilitate direct comparison of probabilities of success in different jobs. The same numerical value now signifies the same probable status in the group in each job. The highest numerical value for an individual corresponds to the job in which we should predict highest success for him relative to his fellows. Where differences in battery validity for different jobs are substantial, tempering our expectation of individual job performance by what we know of the validity of the predictor for each job should represent a significant improvement in interpretation of individual aptitudes.

A second adjustment of score composites which should perhaps enter in before the composites are used as a basis for assignment is weighting them in accordance with the importance of the job. Thus, let us assume that we have a flow of incoming recruits who must be assigned either to electronic technicians school, to aviation mechanics school, or to cooks and bakers school. Let us assume that current evaluation of the needs of the Service weights electronic technicians 5, aviation mechanics 2, and cooks and bakers 1. A score which is to be used for classification might well incorporate these weights as multiplying factors, to insure that even slight superiority in a high-priority job resulted in assignment to that job. In some cases, of course, there may not be enough difference in the importance of dif-

TABLE 3
Illustration of Standard Scores Weighted for Validity and Job Importance

	Electronic Technician	*Aviation Mechanic*	*Cook-Baker*
Importance of factor	5	2	1
Validity of test composit	.60	.70	.40
S. D. — equally weighted true criterion scores	10	10	10
S. D. — adjusted for validity (predicted criterion scores)	6	7	4
S. D. — adjusted for validity and importance	30	14	4
Mean — all score distributions	50	50	50

ferent jobs or not enough may be known about the importance of different jobs to make make this type of weighting fruitful.

Let us carry our illustration further, and combine these two types of weighting factors. Suppose that for the three jobs which we have just considered, the validity coefficients against comparably good criterion measures are respectively. 60, .70, and .40. Then, the standard deviation of converted scores for Job *A* (electronic technician) might be $10 \times 5 \times \underline{.60} = 30$. For aviation mechanic it would be 14; and for cooks and bakers, 4. The picture is summarized in Table 3. The final converted scores, all with the same mean, but with standard deviations adjusted both for validity and for job importance, represent a type of score which appears to permit the simplest and most direct comparison, with view to personnel assignment.* We must now consider how these scores, or scores derived by some other system, might be used in the operation of assignment.

Three general ways may be proposed in which a set of composite scores could be used for the assignment of personnel in a classification situation. For purposes of identification these may be called the Method of Divine Intuition, the Method of Daily Quotas, and the Method of Predicted Yield. They will be discussed in that order.

The essence of the Method of Divine Intuition is that it is not a method at all — or not describable in any exact terms. The individual responsible for making recommendations or assignments looks at the set of aptitude scores for each individual and, reconciling these scores with what he knows about flow and quotas, comes by unspecified and unspecifiable processes to a decision as to the category into which each man is to be put. The approach is, of course, quite unstandardizable. At its worst, it degenerates into the not unknown routine of assigning the *A*'s to one job, the *B*'s to the next, and so

*Hubert Brogden (An approach to the problem of differential prediction. *Psychometrika*, 1946, 11, 139-154) has approached the problem of a uniform score scale for predictions of success in different jobs from a somewhat different angle. He proposes translating everything into units of dollars saved by assigning Individual I, rather than an average individual, to Job *A*. However, he indicates that it will usually not be feasible to get a direct estimate of this dollar saving, and that one will have to rely upon subjective estimates of the type discussed in the present article. Brogden makes one point which is a desirable supplement to our discussion of a uniform scale for predictor scores. This is that one must take account of the extent to which efficiency varies from person to person. In some jobs the difference in level of performance may be relatively slight between the best individual and the average individual, while in other jobs it may be very great. The weighting factor for any job should take account not only of the importance of the job but also of the extent of individual differences in performance of it. The two judgments might be made separately, or they might be synthesized into a single-compound judgment of the importance, to the success of the total organization, of the differences which are in fact found between individuals in the performance of that particular job.

on through the alphabet. At its best, it becomes the sincere but relatively unguided attempt of the individual personnel worker to reconcile a host of complex aptitude profiles with a varied array of job assignments.

It is as much in this unstandardized situation as anywhere that the adjusted type of standard score which has been described earlier would be helpful. It would be helpful because it would take out of the hands of the individual operating personnel worker the necessity for making judgments about the relative validity of the different score composites and the relative importance of the different jobs. These judgments would be centrally made once and for all by the best talent which the organization was able to bring to bear upon the problem. The first-line operating personnel worker could take the numbers at their face value, and strive simply to get every individual into the job for which he had the highest numerical score.

Difficulties arise when, due to quotas, it is not possible to get all (or even most) of the individuals into the job for which their score is highest. This will be the case when there is a disproportionate need for personnel in certain particular jobs or in certain families of closely related jobs. It is for these situations, and to minimize the subjectivity inherent in assignment by unrestrained individual judgment, that the methods of Daily Quotas and Predicted Yield are proposed.

The Method of Daily Quotas undertakes to take the quota requirements for each day (or other specified period), and make the optimum adjustment of aptitudes to assignments within the limitations of the personnel and quotas available for that period. In approaching this problem of optimum adjustment of individuals to quotas, we encounter the basic dilemma of classification. Due to the interdependence of the assignment of different individuals, the optimum allocation of individuals requires that everybody be assigned simultaneously. That is, in order to decide to which job any *one* person is to be assigned, one needs to know what *other* persons are candidates for these same jobs and what their qualifications for each job are. But there is no practical way in which simultaneous assignment of any substantial number of individuals can be handled. Since simultaneous assignment is impractical, what is required is some way of establishing the best order in which the individuals shall be assigned to jobs. That is, since we must start with some individual and assign him first, with whom shall we start, who shall be assigned next, and in what order shall we proceed from there?

The key to this problem lies in the fact, which we have previously noted, that it is *differences* in aptitude which are important for classification. The person whose assignment has great influence upon the success of our classification enterprise is the person who shows a wide range in potential contribution in different jobs. The person who can make an equal contribution in all jobs can be assigned to any one without gain or loss to the total classification enterprise. We need, then, some index of variability in expected contribution, in terms of which we can arrange the members of our group in order, with a view to assigning first those with the greatest variability.

One may question whether it will be possible to find any one index of variability which can be shown analytically to be the best for use in classification. The procedure which is proposed below is admittedly a rule-of-thumb one, but is perhaps a reasonable way of organizing aptitude estimates for use in personnel classification. It is assumed that we start with aptitude estimates for each man for each job category, expressed in some type of comparable score units. The units which were proposed earlier take account of the validity of the score composites and the importance weighting of the jobs, but the procedure would be essentially the same whether this is done or not. The steps would be essentially as follows:

(1) Determine for each individual a measure of *spread* of his predicted achievement scores. The problem is to determine the particular measure of spread which will most effectively identify those with certain outstanding aptitudes. The measure should be simple, and it should give special weight to spread among the individual's higher aptitude scores. It would be an exceptional situation in which it would be necessary to assign any individual to a job which fell in the lower half of his aptitude ratings, and so we do not need to concern ourselves about how inept the individual is for the jobs for which he is particularly inept. Often it will be a question of choosing from among the two or three highest ratings. As a practical expedient, it is suggested that for each individual the difference be determined between his *highest* predicted achievement score and his *median* predicted achievement score. An alternative might be to determine the difference between his highest and next highest scores.

(2) Arrange all the individuals in the order of their measure of spread, from high spread to low spread.

(3) Assign to a job first the individual with the largest spread score, then the one with the next largest, and so on. Insofar as quotas permit, assign each individual to the job for which he has the highest predicted achievement. If assignments are made in this order,

maximum flexibility of assignment will be available for those individuals who show the greatest spread in predicted achievement. The last individuals to be assigned, for whom a number of jobs may be excluded because the quotas are already filled, will be those who show relatively little difference with respect to their predicted achievement in the several jobs.

To explore the effectiveness of this routine, a synthetic set of scores was set up by drawing cards from a deck. The numbers in each denomination were arranged so as to yield an approximately normal distribution of scores. Scores on ten independent "tests" were obtained by drawing cards from the pack. Then composite scores for predicting achievement in seven "jobs" were obtained by combining different groupings of "test" scores. These seven score composites were obtained for a sample of 50 individuals. The composite scores were designed to have population means of 50 and population standard deviations of from 4 to 5.

First, the limits of zero effectiveness of classification and 100% effectiveness of classification were defined by determining the mean aptitude score when men were assigned to jobs at random, on the one hand, and when each man was assigned to the job for which he had the highest aptitude (without regard to quotas) on the other hand. For this sample of 50 cases, the mean scores were 48.34 and 52.16, respectively. The difference of 3.82, almost one standard deviation of the original aptitude score distributions, represents the difference between random assignment and ideal assignment.

Then it was assumed that the quota was 7 for each job except the last, for which the quota was 8. First assignment of the men was made in "alphabetical" order, that is their order in the list of scores. As far as quotas permitted, each man was assigned to the job for which his "predicted achievement" was highest. Following this procedure, the mean of all the aptitude scores was 51.82, constituting 91% of the possible improvement over random assignment. Then the 50 cases were arranged in a new order, on the basis of the measure of *spread of aptitude* for each individual. Assignment was made following this order, again assigning each individual, within quota limits, to the job for which he had the highest aptitude. The mean of all the aptitude scores of men for the jobs to which they were assigned was now 51.92, representing 94% of the possible improvement over random assignment. In this particular illustration, with substantially equal quotas for each job, order of assignment made little difference; and most of the advantages of using aptitude tests could be obtained regardless of the order in which the men were assigned.

By contrast, it was next assumed that the quotas for the several jobs were respectively 20, 10, 5, 5, 5, 3, and 2. Under these circumstances, assignment in "alphabetical" order achieved a mean aptitude score of 50.86, or 66% of the possible improvement over random assignment, while assignment in order of spread of aptitudes achieved a mean aptitude score of 51.58, or 85% of the possible improvement over random assignment. These results support the reasonable conclusion that it is when quotas are out of balance with supply that taking account of the factor of spread makes possible improved classification. In this case, arranging the individuals in order of spread of aptitude scores and assigning them in that order resulted in a substantial improvement in the effectiveness of assignment.

The great limitation of the Method of Daily Quotas as a procedure for assignment is that it is restricted by the requirements of fluctuating quotas. Thus, Monday's quota may call for 100 aviation mechanics to fill the roster of a class entering training, and then there may be no further call for aviation mechanics for the rest of the week. If assignment operates on a short-time basis and such wide swings are typical, classification is reduced to a fraction of its possible effectiveness. It is only when quotas are reasonably stable from day to day, or when personnel can be pooled for a long enough time to permit them to become stable, that the Method of Daily Quotas can be an effective classification routine.

One way of stabilizing assignment procedures in the face of fluctuating daily quotas is to adopt some variation of the Method of Predicted Yield. This means, in general, forecasting the numbers required in each job category for a period of time, and then working out a standard routine for assignment which will, on the average, yield the required proportions in each job. The procedure should be that which will, for the average of all persons assigned, achieve the maximum predicted achievement of each man in the job to which he is assigned. (The type of weighted standard score described in a previous section might well be used for expressing predicted achievement in different job categories.)

By way of illustration, let us suppose that we wish to divide job applicants among maintenance, clerical, and sales positions. Let us suppose, for simplicity, that the relative importance and validity factors are such that we may use standard scores with the same standard deviation for all three job categories. Let us further assume that estimates of the flow needed into the three job areas are in the proportions 50, 30, 20. We have for each individual three standard scores which predict his probable achievement relative to his fellows in each of the three areas. Our problem is to prepare some set of rules which will result in the assignment of the required proportions to the three jobs, and will at the same time maximize predicted achievement. How shall we set the dividing lines which will separate those who should be assigned to Job A from those who should be assigned to Job B or Job C?

If we are willing to assume that the various regressions of success on aptitude composite are linear, to specify that we will use straight lines (or planes or hyperplanes) to separate our sub-groups for assignment, and to make the further assumption that the various aptitudes with which we are concerned are normally distributed, the problem appears to be susceptible to mathematical solution. A mathematical solution for the two-variable case has been proposed by Dr. T. F. Cope of Queens College.* The solution in this case supports the earlier discussion, in which it was pointed out that the two-variable problem can be treated as a single set of difference scores. The mathematical solution is of the form

$$Z_A - Z_B = \text{constant}\,,$$

where Z_A and Z_B are standard scores, of the type we have discussed, for predicting success in Jobs A and B.

Thus, to decide whether a person should be assigned to Job A or Job B, one merely subtracts the B score from the A score, each score being weighted if that seems desirable. If the difference is greater than a specified constant, the individual is assigned to Job A; if equal to or less than that constant, to Job B. The size of the constant is determined by the numbers required in the two jobs and by the correlation between the scores for predicted aptitude in the two jobs. The value of the constant k is given by an equation of the form

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{x+k} z\, dy = p\,.$$

The numerical value of k can be approximated in any given case by preparing a distribution of difference scores, and counting off the required number of cases.

If some cases are rejected for both jobs, i.e., if there is a minimum score to qualify for each job, the value of k will be determined

*Personal communication to the author.

by the two minimum scores to qualify. These qualifying scores, in turn, will be determined by the number of individuals required for each job. Brogden has shown that the constant k will, in this case, be equal to the difference in the two cut-off scores.

The solution of the form $Z_A - Z_B = k$ will apparently generalize to three or more dimensions. In this case, one will have $\frac{n(n-1)}{2}$ hyperplanes which will divide the multivariate frequency distribution into a number of sub-regions. It is anticipated that these hyperplanes will all be of the form

$$\begin{aligned} Z_A - Z_B &= k_1 \\ Z_A - Z_C &= k_2 \\ Z_B - Z_C &= k_3 \\ &\text{etc.} \end{aligned}$$

In this case, one would have to deal merely with a set of simple differences between predictor scores. To assign an individual, one could proceed by taking the individual's A score and subtracting from it his B score. The difference would tell whether he would be better assigned to A or B. Taking the winner from this comparison, one would subtract from it the C score to determine whether C was a better assignment than the previous choice. There would be, for each individual, $n - 1$ subtractions and decisions, where n is the number of job categories among which placement is being made. Assignment would be made to the job category which survived as the winner in all these comparisons. The routine work for this task could be arranged so that it would proceed quite rapidly.

The basic problem in this case is to define the constants k_1, k_2, k_3, etc. in such a way as to yield the required proportion of cases in each of the job categories. This problem is also believed susceptible to an analytical mathematical solution, though the solution is not available at this time. Even without a mathematical solution, however, it does not appear too laborious to determine approximate values of the constants, based upon a set of empirical data, by an iterative procedure. The procedure can be illustrated by data for three variables and 162 individuals, which were analyzed by the writer. The three variables may be thought of as predictor scores. (As a matter of fact, they were final course grades and certain component grades from a measurement course.) Identifying them as variables A, B, and C, the correlations were as follows: $r_{AB} = .38$; $r_{AC} = .66$; $r_{BC} = .90$. For each of the variables, scores were converted to standard scores with means of 50 and standard deviations of 10.

The first problem was to determine values of k_1, k_2, and k_3 to give a yield in the proportion 50 *A*'s, 30 *B*'s and 20 *C*'s. The initial estimates of the *k*'s were made by taking the variables by pairs and using the procedure described earlier. That is, variables *A* and *B* were considered, with the required proportion being 50 to 30, and k_1 was estimated to be

$$-0.32\ (10) \sqrt{1-(.38)^2} = -3.0\ .$$

Similarly, k_2 was estimated to be —4.1 and k_3 to be —1.1. For the first empirical trial, therefore, the values —3, —5, and —2 were used as the *k*'s. The rules for assignment became: "Assign persons to Job *A* unless *B* score is 3 points higher than *A* score or *C* score is 5 points higher than *A* score; Assign to Job *B* unless *C* score is 2 points higher than *B* score. This procedure led to the following assignments:

To Job *A*: 90 cases, or 56%
To Job *B*: 57 cases, or 35%
To Job *C*: 15 cases, or 9% .

Adjustments were then made to yield more cases in Job *C*, at the expense of both Jobs *A* and *B*. That is, the constants were changed so that the *C* scores would not have to be as much above the *A* and *B* scores. On the fourth trial, at the end of about half an hour, the following rules for assignment were arrived at: "Assign to Job *A* unless *B* or *C* score is 3 points higher than *A* score; assign to Job *B* unless *C* score equals or exceeds *B* score." This procedure led to assignments as follows:

To Job *A*: 81 cases, or 50.0%
To Job *B*: 48 cases, or 29.6%
To Job *C*: 33 cases, or 20.4% .

This is certainly as close correspondence to the desired percentages as one could hope to obtain.

Using the same set of scores, another problem was worked out in which the required percentages were 10 for *A*, 20 for *B*, and 70 for *C*. With three trials and about the same amount of time as before, a set of rules was reached which gave the following results:

To Job *A*: 17 cases, or 10.5%
To Job *B*: 32 cases, or 19.6%
To Job *C*: 113 cases, or 69.8% .

In this case, the rules were: "Assign to *A* if *A* score is 10 points above B and C scores; assign to B if B score is 4 points above C score."

With a greater number of variables, or with a very large N, this cut-and-fit procedure would become more laborious; but even so, the time involved would probably be a small item in the total planning of a classification program. Again, Brogden* has considered a similar problem and worked out a similar iterative procedure for the situation in which a portion of the job applicants may be rejected and not assigned to any of the jobs.

The Method of Predicted Yield would be feasible only in a large-scale program which was going to continue on a basis of stable requirements for a considerable period of time. It also requires provision for absorbing temporary unbalance between quotas and yield. Within those limitations, it seems to offer maximal opportunity for effective classification.

By the very nature of things, classification of personnel can never have the neatness and elegance that are possible in an enterprise of pure selection. Classification will relatively rarely be called for in its pure form. Even when it is, the variety of points at which professional judgment must enter will increase as some function of the number of jobs among which the available personnel must be divided. This paper has tried to suggest, however, some directions in which we may move to make more effective our design and selection of a test battery and our use of the resulting test scores for differential assignment of personnel.

*Brogden, Hubert E. An approach to the problem of differential prediction. *Psychometrika*, 1946, 11, 139-154.

Manuscript received 10/10/49
Revised manuscript received 1/3/50