A COMPARATIVE STUDY OF METHODS FOR THE COMBINATION OF PREDICTORS IN PUBLIC PERSONNEL SELECTION

by
Albert P. Maslow

Thesis submitted to the Faculty of the Graduate School
of the University of Maryland in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

1952

F

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the stimulation and advice of the thesis committee, Dr. D. D. Smith, Chairman, Dr. C. N. Cofer, and Dr. R. C. Hackman. Each of them has been generous with his time and his ideas during the development and conduct of this study. I owe a special debt to Dr. Hackman for his insight and suggested solutions to several of the statistical problems posed in this research.

A number of officials of the United States Civil Service Commission have been instrumental in making this study possible. Mr. F. W. Luikart, Chief, Examining and Placement Division, approved and encouraged the formulation of this study, and the use of official records and data. Mr. J. F. Scott, Chief, Test Development Section, and Dr. M. D. Davidoff, Head, Research and Analysis Unit, have provided much administrative and technical support.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF TABLES

LIST OF TABLES (Cont.)

CHAPTER I

INTRODUCTION

A. The general problem of combination of predictors, and typical procedures in use.

Most personnel selection, classification and counselling programs in which objective psychological tests are used present the problem of how best to combine test scores to meet the particular objectives of the program. In personnel selection for a particular occupation, it is generally desired to combine test scores in such a way as to yield a single index which will most accurately predict each applicant's performance in that occupation. Personnel classification, where the goal is to assign applicants to one of several occupations, is a more complicated statistical and administrative procedure, but also requires some method for combining and evaluating a number of test scores with respect to several criteria. Similarly, in vocational guidance and diagnosis, research efforts have been directed toward the combination and representation of test scores in profiles and patterns as a basis for matching the measured characteristics of the individual with specified occupational or academic group norms.

In public personnel work, psychological measurements have been applied more generally for selection and classification than for counselling. Methods for combining test scores as well as for combining tests and other selection devices--such as ratings on training and experience, biographical data, and interviews--are of direct practical as well as theoretical importance since they affect the cost, practicality, and public acceptance of the program.

Three main methods for the selection and combination of predictors are in use. Probably the most general method is that based on multiple correlation analysis. In this method, the correlation of each test with the criterion and with each other test is considered, and the battery of weighted tests derived is that which yields the most accurate prediction of the criterion (in a least squares sense). A second general method is perhaps best known as the multiple cut-off procedure. In this method, "critical scores", or lowest acceptable scores, on each test are defined, and the several tests are then combined in some way to achieve the maximum prediction of the criterion. A third method is the pattern or profile technique, in which persons are identified, individually or as a group, on the basis of the unique configuration of their test scores or categories.

B. Characteristics of multiple correlation methods.

Although multiple correlation techniques are available for use under conditions of non-linear as well as linear regression, they have been developed and used chiefly for the latter case. Ordinarily, the multiple correlation method is used to derive a single distribution of weighted test scores. Selection is made from this distribution in order of total weighted (most predictive) scores. It is important to note the principle of "compensation" in this method. A high score on one variable may compensate for a low score on another; the same total weighted score may be achieved by a variety of patterns of scores on the several predictors. It is also a characteristic of this method that the prediction equation minimizes the error over the whole range of scores. It is not sensitive to particular segments of the range, which may be peculiarly related to criterion performance or which may sharply distinguish successful from unsuccessful employees. When the assumptions of linearity of regression and homoscedasticity are not met, routine use of multiple correlation may be less efficient than some other combining method and may obscure significant non-linear relationships. Richardson, for example, has commented that "the general problem of the combination of measures has been obscured by the indiscriminate adoption of the multiple correlation technique as the 'best' solution, and by the failure to investigate the properties of various weighting systems" (25, pg. 379). A second major limitation of the multiple correlation approach is its computational difficulty, when many variables are employed. Although much research has been directed toward simplifying the procedures, it remains a complex problem both in application and interpretation. Finally, unless carefully applied and cross-validated, a test battery selected by this method exploits sampling errors in the sample used for determining the weights for the several tests.

C. Characteristics of multiple cut-off methods.

The multiple cut-off method, in contrast to multiple correlation, makes no assumptions as to the nature of the regression of criterion scores on test scores. As it has been formally described by Grimsley (11), it is a frankly empirical method for determining the best cut-off scores on each predictor and on various combinations of predictors. Critical scores may be found arithmetically or graphically; in general, they are located at test scores which appear to maximize differences within the criterion group. When these critical scores are determined, each person's scores are coded as above or below the critical scores. This method appears to be simple to apply and adaptable to a variety of situations for which multiple correlation is not appropriate. It permits taking advantage of particular characteristics (breaks, skewness, etc.) of the score distributions. It is also significant to note that each critical score operates as an eliminator: compensation is not permitted.

The concept of multiple cutting-scores has long been applied in public personnel selection programs. Various cut-off procedures, such as "simultaneous" or "successive hurdles," have been devised to reduce the number of applicants reaching successive stages in the testing or scoring process. The principle that the most valid test should be the basis of the first cut-off, and so on, is probably well recognized, although, in large-scale testing programs, the elimination rate and consequent reduction in processing costs no doubt effect the decision as to the order of use. From the standpoint of measurement, the cut-off procedures would seem to make greater demands upon the reliability of each of the tests at the point of cut-off, whether or not the tests were used singly or in combination. Finally, there should be noted one other characteristic of the cut-off method as it has generally been applied. The cut-off point on a particular test has the effect of selecting all persons at or above the critical score. There are no maximum critical scores to define particular segments of the range. In effect the critical score dichotomizes the test score distribution. While this is not a significant limitation in selection programs (particularly in public personnel selection) the method is less flexible for other purposes where several categories of each variable are to be defined.

D. Characteristics of pattern and profile methods.

Pattern and profile methods attempt to capitalize on the predictive significance of the unique relationships among the several test scores. The total of possible individual patterns in several tests, each categorized into a number of score groupings, is entirely too large to handle by any simple means. More important, the accuracy of measurement for most tests does not often justify the use of discrete scores as separate categories. For both these reasons, pattern methods demand some basis for reducing the categories to a manageable number, and some method for manipulating the patterns resulting from combining categories. The graphic device of profiles has been used to represent a pattern of scores for comparison with the profile of a criterion group. Patterns and profiles present similar problems in use and interpretation. These problems chiefly concern: the estimation of the reliability of the pattern and the significance of differences among the several scores; and the measurement of the degree of similarity of profiles or patterns to the criterion group pattern. These methods can conceptually handle a wide variety of data, both quantitative and qualitative, but the problem of interpretation is magnified rather than reduced by use of heterogeneous data.

The relationship between multiple cutting-score methods and pattern methods is very close. The multiple cutting-score method can be considered as a special case of a pattern under the condition that categories of the tests are defined only by their lower limits. The pattern would seem to be more general and flexible in that it permits rejection of persons whose test scores fall _above_ an optimum, as well as below a minimum.

CHAPTER II

REVIEW OF PERTINENT LITERATURE

The following review is not intended to be an exhaustive survey of the great volume of literature about the theory and applications of multiple measurements for the prediction of behavior. Since the study is designed to evaluate several current methods as they may apply to public personnel selection programs, this chapter highlights certain major methods and reviews some typical findings. Particular studies of the multiple cut-off and pattern techniques, which serve as a basis for the formulation and design of this study, are discussed in detail.

A. Methods for the linear combination of measures to predict continuous criteria.

Recently Mosier (24 pg. 764 ff) has reviewed the general theory and sample techniques that have been developed for the linear combination of measures under various conditions of the number of predictors and criterion elements. He points out that although the linear hypothesis may be the simplest it is not necessarily the most accurate; particularly, the hypothesis may break down at the extremes of the range in the area of deviant behavior in which the clinician is especially interested. However, he adds that non-linear relations have seldom been found.

Of particular interest in applications of the multiple correlation technique is the Wherry-Doolittle method for the selection of the most predictive set of variables with respect to a single criterion. While this method is straightforward and permits an estimate of whether each additional variable is contributing error or non-error variance, it is complicated and difficult to carry out and does not lend itself readily to machine or IBM operations. Furthermore, like other multiple correlation methods, it requires cross-validation to guard against capitalizing on chance variation in the sample. Nevertheless, since it represents a typical and frequently used technique, it is a useful standard against which to compare other less well-known methods.

The difficulty in computing multiple regression problems has led to many attempts to develop approximation techniques, including graphic methods and procedures for estimating and using approximate regression coefficients. As Grimsley points out, the finding that rounding off beta weights to single integers makes little difference in the multiple correlation, has raised a question as to the value of full solutions of the regression equation (11). It has also encouraged the search for simpler empirical methods (such as multiple cutting-scores) as substitutes.[1]

The general method of multiple regression has also been applied in recent years to the problem of differential prediction and classification. The purpose is to predict several different criteria, and the size and significance of the differences between sets of predicted criterion scores becomes important for classification and counselling. This problem has been recently explored by Thorndike (30) and Wesman and Bennett (37). Super refers to specific batteries and programs based on this approach (28).

B. Methods for the linear combination of measures to predict categories.

In personnel selection, it is often more practicable to define a categorical criterion, such as pass-fail in a training course, or satisfactory-unsatisfactory in meeting specified work standards, than it is to obtain a

---

[1]The Test Development Section of the United States Civil Service Commission has frequently found it practicable to use integral weights in scoring multiple test batteries, in lieu of fractional regression coefficients.

continuous criterion. A number of methods have been offered to meet the problem of prediction of such categories from continuous measurements. The discriminant function has been developed by R. A. Fisher as a method of weighting variables to maximize the differences between two defined groups. Garrett (10) and Anderson (2) have shown its application to psychological data. More recently, Wherry (38) has demonstrated that the multiple regression weights derived from a solution using biserial or point-biserial criterion correlations are proportional to the weights yielded by more complicated discriminant function analysis. Rulon (26) and Tiedeman (27) have further explored the use of these methods for classification and guidance. In practical application, this method permits the solution of regression weights for a dichotomous criterion and their use in a regression equation by the device of setting up the criterion as a "dummy variate."

Graphic and arithmetic techniques for predicting categories have been devised, at least for the case of a dichotomous criterion, by Guilford and Michael (13) and Betts (4). They are based not on multiple correlation analysis, but upon the determination of that critical test score which produces a prediction of membership in a criterion category at a specified probability level.

These methods, however, deal only with a single predictor. While they might apply to multiple measures, they have not yet been developed to meet the problem of test-selection and combination.

One other approach using a single predictor to predict categories is the multi-serial technique developed by Jaspen (17).

C. Multiple cut-off score techniques to predict a continuous criterion.

Horst (16) has provided a theoretical formulation of the multiple cutting score technique. The monograph by Grimsley (11) is the most detailed available exposition of the particular multiple cutting-score (MCS) method proposed by F. L. Ruch. The procedure calls for the following steps:

1) Define the "critical score levels" on a variable by calculating the mean test score made by specified criterion groups. Since, at any critical score, all persons scoring at or above are considered acceptable, the criterion groups are cumulative. For example, critical scores A, B... F are the mean test scores made by the top 1/8, top 1/4, top 1/2, total group, lowest 1/4 and lowest 1/8 of the criterion group.
2) Determine the percent who would be selected at each of the critical scores on the test (cumulative).
3) Compute the mean <u>criterion</u> score for each selected group (cumulative).
4) Evaluate the discriminating value of a test by graphing or tabling the relation between mean criterion score and percent selected--the greater the decrease in mean score for successive groups, presumably the more valid the test.

5) Evaluate combinations of tests by applying the critical score levels A to F to various combinations, and calculating the mean criterion score and percent selected by the successive A to F levels. For example, in the case of 2 tests, the first group would be those who satisfied the A level on <u>both</u> tests; the second group those who met the B level or above on <u>both</u> tests, etc.

Procedurally, the computations and classification of persons arereadily done from individual data cards. The stability of the "best battery" is checked by applying the critical scores to a second sample and computing the difference in mean criterion score between groups at comparable critical score levels. Comparison is also made with a battery selected by the Wherry-Doolittle method by comparing mean criterion scores of groups at specified critical score levels with the mean criterion score of the same number of cases from a distribution ranged in order of (Wherry-Doolittle) predicted criterion scores.

Grimsley found that the MCS method had a small but insignificant advantage at high selection levels. He concludes that, overall, the MCS method is just as accurate as the Wherry-Doolittle, required only one-third the calculation time, required little knowledge of statistics, and was less affected by shrinkage.

However, it may be questioned whether the data used by Grimsley provided for a thorough test of the MCS method. The tests he used had relatively restricted ranges of scores and relatively low validities (zero order r's up to .308, R=.403). For example, he reports the following typical means and standard deviations respectively for 3 tests: mean 10.50 s.d. 2.91; mean 10.96 s.d. 3.15; mean 8.64 s.d. 3.01. This restriction in range is important in several respects. In classifying groups whole numbers must be used as critical scores. Therefore, the fewer integral scores in the effective range, the smaller the number of categories that can be isolated, and the larger the frequency within each category. Furthermore, the narrow range implies unreliability of the tests, although no coefficients are reported by Grimsley. It is apparent that he made use of data in which it would be very difficult for any technique to show discriminating ability or stability on cross validation. It is indicative that he found only a small proportion of common cases at high selection levels in comparing the two procedures. Furthermore, the difference in mean criterion score between the highest and lowest groups in his experimental population was less than 3 points, as compared to a standard deviation of 7.8 for the criterion scores.

Several comments about the technique itself may be made. First, as to the means for defining the critical scores, there seems to be no special reason for grouping the population on the criterion and computing critical test scores for those fixed criterion groups. In the following chapter on Methods and Results, a simpler and more direct basis for grouping the population is developed. Second, no rationale is represented as to how many groups are initially justified. Grimsley says only that "it was decided that with 250 cases. . .it should be possible to have six sets of critical scores." (11, pg. 101). Third, the combination of tests by using the same critical score level across all tests would appear to be a limitation of the method, as compared with the possibility of combining tests at varying critical score levels. That is, combination of tests 1, 2, and 3 might theoretically be better if made by using test 1 at level A, test 2 at level C, and test 3 at level B. The method he used, however, examined tests 1, 2, 3, all at level A, then all at level B, and so on. It is recognized that applying a concept of varying levels would immediately increase the number of combinations to be tested empirically, and would tend to negate the simplicity of the MCS method. Grimsley did use one variation which appeared useful, but did not develop a systematic plan for combining tests at varying standards. Finally, it should be noted that the method deals with group prediction and does not yield a practicable means for making individual predictions and evaluating them.

It is clear that the large number of conditions that affect multiple correlation studies also affect the MCS and other procedures for manipulating test scores.

These include the degree and nature of intertest correlations, test reliabilities, number of tests used, nature of the criterion, etc. Few of these conditions have been systematically investigated in comparing MCS and multiple regression methods. In practical application, Thorndike (29) points out four problems:

1) Unless there is clearly a non-linear relationship between predictor and criterion there is no unique basis for selection of a cut-off point.
2) With a large number of predictor variables, the trial and error process with all the possible combinations of cutting points is an overwhelming task.
3) Adjustment of standards to supply and demand is simple with a single composite score, but requires full recomputation with multiple cutting-scores to determine the percentage of applicants selected by a given combination of scores.
4) The multiple cutting-score method is not adaptable to the problem of classification, as distinct from selection, since it gives no quantitative score showing degree of ability for any particular job.

With respect to these points, at least in public personnel selection, it is pertinent to note that, in using composite scores based on the multiple regression method, there is similarly no unique basis for selection of the cut-off point which separates the qualified and unqualified. In both cases, the decision generally is made in terms of supply and demand, and the amount of risk that is tolerable in accepting unlikely applicants, based on the validity of the predictors. The problem of selecting cutting points for trial is soluble, within the limits of the modest number of tests in most actual selection programs, by use of IBM or other sorting devices. These are also essential to any large scale multiple regression analysis and weighting. Similarly, machine procedures should meet the problem of estimation of number of qualified applicants, within the limits of practical test programs. Finally, although the quantitative scores on predictors have not been used in previous studies to rank the applicants, there is no reason why, after cutting scores have been determined, the test scores should <u>not</u> be used. In fact, in public personnel work, ranking of candidates is generally a requirement regardless of the methods of setting the qualifying standards.

In view of the theoretical advantages as well as practical values of such cut-off methods, and despite the difficulties posed by Thorndike, further study of the relative values of such methods as compared to multiple regression seems fully justified.

D. Multiple cut-off score methods to predict a categorical criterion.

The methods reviewed above are related in that they make use of the actual test scores and are concerned with manipulating these scores in maximizing group differences or predicting a criterion variable. A somewhat different approach for combining tests to predict a dichotomized criterion has been developed by Franzen and

Lazarsfeld (9), using frequency comparisons and chi-square analysis. Essentially, the procedure is to find, in terms of chi, the point in the test score distribution which yields the greatest difference between the pass-fail frequencies in the total group as compared to the pass-fail frequencies in the failure group alone. When the cut-off points for the individual predictor variables have been determined, a simple method using partial deltas (derived from the computation for chi-square in a 2 by 2 table) is used for computing all likely combinations of predictors to find the most discriminating set.

In one comparison of the method to multiple-biserial regression analysis, the "multiple-chi" was found superior at several selection levels examined. It is also pointed out that an advantage of the method is to permit analysis of compensatory relations that exist among the tests. This does not mean that the principle of compensation operates with respect to the scores of a particular individual: rather, it means that, cut-off points on the tests having been established, the relations among the tests can be evaluated. For example, the method can demonstrate whether failure on one or more variables is as significant as failure on all of them.

This method would appear to have value for situations where the criterion is a dichotomy, and where distinctions among, or a ranking of, the selected group is not a requirement, and where non-linear relations might exist. However, the problems that would be raised by the need to adjust the cut-off points to yield a required number of selections have not been fully investigated.

E. Pattern and profile methods.

The essence of the methods which may conveniently be grouped as pattern and profile techniques lies in the assumption that the total population under study can be classified on each predictor variable or attribute, and that particular combinations of such classifications will have unique value for the prediction of criterion performance. Toops (33, 34) has discussed in detail the theory underlying these methods. Where each variable can be classified with accuracy (for example, such attributes as sex, occupation, etc.) the major problem that remains is the procedural one of identifying all possible combinations of such categories, so that each combination can be evaluated independently. Toops proposed "addend" coding of variables uniquely to identify a particular combination, or as he calls it, "the ultimate breakdown society to which a given person belongs" (34, pg. 41). This "uśstrith" is, in effect, a cell entry in an l by k table of l variables each categorized in up to k groups. He points out that the homogeneity and uniqueness of a sub-population varies with the validity and independence of the predictor variables. Thus it is desirable for research that there be only a limited number of valid and independent predictors. Also, there should be only a very few categorizations per variable, since each sub-population must have a large enough frequency to yield reliable differential criterion means.

Johnson (18) has developed a similar coding method based on the properties of the binary number system. He also describes a procedure for comparing the obtained frequency of sub-populations with expectancy by a chi-squared criterion. Toops' comments as to number of categories and independence of variables also apply to Johnson's method. Johnson points out one advantage of such coding methods over multiple correlation; they permit the score or code of the sub-populations to show <u>how</u> the score was derived, whereas in multiple correlation this information is not apparent in predicted criterion scores.

Tucker (35) recently has studied the unique pattern technique as compared to multiple regression, in predicting both continuous and dichotomous criteria. His procedure was, in general, to classify a population into sets of patterns varying the number of patterns by changing the number of categories of each variable. (Thus, 3 variables each trichotomized would yield 27 unique patterns.) He found that the multiple regression technique was superior for quantitative, linearly-related variables, but that unique patterns were equally effective for qualitative or a mixture of both kinds of variables. The operating advantage in using only a small number of categories of the predictors did not result in a loss of validity, and showed less shrinkage. In general, a crude grouping with more predictor variables was superior to refined categorizations with fewer predictors.

This study confirms certain theoretical expectations as to the grouping effect in correlational analysis and as to the increase in validity related to adding independent and valid predictors to an existing battery. As Tucker points out, it supports the view that a clinician or counsellor can work more effectively with a specific number of patterns based on a few categories and variables, than with a large and unmanageable number of poorly defined patterns. As the basis for categorization of quantitative variables, however, Tucker used arbitrary groupings or stanine scores. He did not investigate reliability of differences for varying numbers of categories, except in terms of their effect on the shrinkage of multiple correlations in cross-validation.

Patterns and profiles which represent them have been a major tool of clinicians, vocational counsellors, and others concerned with the advisement of individuals and prediction of individual success. Mosier (24, pg. 794) cautions that:

> "Because profiles are simple to construct and are superficially easy to interpret, they constitute one of the most popular methods of summarizing the results of multiple measurement. They. . .enable one to 'picture the total set of test scores and their interrelations at a glance.' Here, more than in any other aspect of test interpretation, do we need to beware of seeming simplicity. By failing to question the reliability of differences between scores, and by relying on the judgment of the interpreter to make the over-all summary, we ignore the possible unreliability and invalidity of the over-all summation which would be instantly revealed if less 'simple' methods were used."

There are many discussions of the use of these methods in vocational guidance. Toops' (33) review of the general problem, Barnette's (3) discussion of occupational aptitude patterns, and Harmon's (15) review of vocational applications are representative. Harmon summarizes his review with the statement that the quantitative study of test patterns has barely begun. Until such tools are available, the counselor will depend on "qualitative clinical insight and judgment" applied to clinical clues available in patterns and profiles. Whether psychometric devices, when available, will, or should, supplant clinical judgment remains a question of faith at this time. Super (28) believes that clinical interpretation is always necessary for proper use of psychometric tools. Mensh (23) has reviewed in detail the variety of approaches being explored for the development of statistical techniques to meet the needs of clinical work rather than group prediction. He notes the shift to "individual-centered" statistics, and reviews the methods of pattern analysis, P-technique of factor analysis, rating, scaling, and multivariate analysis. Some recent examples of studies which deal with patterns may be cited. Cronbach (5) proposed a method for defining intra-individual profile scores (for 3 variables) and plotting them on triangular, homogeneous-coordinate diagrams to determine whether and what patterns existed in the group. DuMas (7) has devised a measure of profile similarity. Meehl (22) has investigated patterns of <u>item</u> responses, in order to exploit predictive possibilities of combinations of response in situations in which each item of a pair separately may have zero validity, but the inter-item correlation is different in the two categories of the criterion.

CHAPTER III

STATEMENT OF THE PROBLEM AND ITS SIGNIFICANCE

A. Objectives of this study.

The general objective of this study was to investigate several methods for the combination of predictors with respect to their effectiveness in predicting continuous and categorical criteria, and their practicality, as compared to multiple correlation analysis. Because of the orientation of the study toward methods which are applicable to selection and placement rather than to guidance and counselling functions, the methods selected for study were the Multiple Cutting-Score, for prediction of a continuous criterion, and the Multiple-Chi, for prediction of a categorical criterion. The choice of these methods for study was based largely on the theoretical and practical advantages they seemed to offer, but also on the finding that, particularly for the Multiple Cutting-Score method, the only detailed study reported (11) was conducted under serious limitations.

A second objective developed during the course of the study. Attempts to simplify and to develop a rationale for the MCS procedure revealed the need for a basis for categorizing continuous variables preliminary to recombination into patterns. The objective then formulated was to investigate methods for determining the optimum categorization of test distributions, and to explore procedures for combining such categories for prediction purposes. The initial design of the experiment did not fully encompass this objective. As a result, certain limitations in the data, particularly with respect to the size of N, affect this phase of the study. However, it was practicable to devise a categorizing method and a pattern analysis procedure as a pilot study for future elaboration.

B. Significance of this study.

The theoretical significance of this research lies in the possibility of establishing methods for the categorization, combination and interpretation of test scores, in a particular kind of prediction situation, which may be more effective than the typical multiple regression method. A method of analysis which does not involve the assumptions of multiple regression, and which permits making use of various patterns of test scores, would appear to have value both for selection and placement or counselling uses. Of practical interest to public and private personnel selection organizations, this study may result in guides for the effective reduction of testing, scoring and test analysis time. It offers the possibility of more effective prediction devices than are yielded by multiple regression procedures for the particular program used in this study.

Although the present analysis has been carried out in the framework of a selection program, the results, particularly as to means for defining test patterns, would apply to the general problem of use of test scores for vocational guidance and clinical purposes.

CHAPTER IV

METHODS AND RESULTS

A. General procedure.

This section presents a brief overview of the design and method of this study. The general plan included, as a first step, the determination and selection of basic data in a prediction situation which met these criteria: 1) multiple predictors which were homogeneous in content, relatively independent, and satisfactorily reliable, 2) substantial validity for the battery as a whole, 3) criteria which could be utilized both as continuous and as categorical variables, 4) a realistic and meaningful prediction situation so that practical problems in application of the procedures and results of the study could be studied, and so that the results might be of direct value in a current program. Next, random sampling was carried out from the total experimental population to select experimental and cross-validation samples, and a second cross-validation sample for study of the stability of techniques other than the multiple regression method.

Following the selection of the random samples, multiple regression analysis was made for both continuous and categorical criteria in order to select prediction batteries as bases for comparison with other methods. Multiple cutting-score and multiple-chi techniques for selection of a prediction battery were then applied. Comparison of the several methods was next made in terms of similarity in predictors selected, predictive efficiency, stability in cross-validation, and practicality. Finally, experimental studies of methods for determining the optimum categorization of predictors as a basis for evaluating patterns of predictor variables were carried out.

B. Specific methods and results.

1. Description of populations and data selected for study. (see Table I).

The variables used in this study are the tests in the Designation test battery (predictors) developed and administered by the U. S. Civil Service Commission for the use of Congressmen in making selections to the U. S. Military and Naval Academies. This battery consists of 5 subtests of vocabulary, reading comprehension, spatial relations, surface development, and algebra. Samples of these item types are given in Appendix I. A more detailed description of the tests is shown in Table I.

The continuously distributed criterion is the sum of scaled scores on the Entrance Examination. This examination consists of aptitude tests (several verbal sub-tests and algebraic computations) and achievement tests in English, Mathematics, and U. S. History. Not all the achievement tests are taken by each candidate; certain tests may be waived where educational prerequisites are met. Relatively few candidates compete in the History test. In this study, in order to have a homogeneous group for the criterion and in terms of educational prerequisites, only those candidates taking the aptitude, Mathematics and English tests have been included, from the group applying for West Point.

The categorical criterion is the pass-fail rating on the Entrance Examination. Candidates are rejected for failure on any one of the several parts of the test.

The population from which the study groups have been selected is 2184 candidates in the June 1950 Designation examination. This population is represented in the analysis by a 20% random sample, N = 436, identified as Group T. Of the total population 292 candidates were among those who took the Entrance battery in March 1951. These 292 cases for whom both predictor and criterion data were available are designated as Group C. Group C was subdivided by random sampling into two groups, of 146 cases each, for experimental and cross-validation use. These groups are designated Groups C1 and C2, respectively. Finally, a third random sample, Group C3, was selected from the total Group C, for further study of sampling stability of the multiple cut-off methods.

A series of unpublished studies conducted by the Test Development Section, United States Civil Service Commission, have consistently shown high validity, (R = approximately .7) for prediction of pass-fail on the Entrance examination, high reliability of the component subtests, and moderate or low intertest correlations. These studies have shown relatively less validity for the spatial variables (variables 3 and 4) than for the verbal (variables 1 and 2)or numerical variable (variable 5). The battery is under further study with particular attention being paid to the contribution of the spatial variables to prediction of Entrance Examination success and Academy course grades.

TABLE I

Description of Groups and Variables

A. Groups

Gr. No.	N	Description of Group
T	436	Random sample of 20% of total of 2184 competitors in the predictor battery.
C	292	All competitors for whom the uniform criterion is available.
C1	146	The Experimental Sample: a randomly-selected 50% of Group C used as the predictor group for determining multiple-regression weights and multiple cut-off points.
C2	146	First Cross-Validation Sample - the remaining 50% of Group C after selection of Group C1, used for cross-validation of multiple regression and multiple cut-off methods.
C3	146	Second Cross-Validation Sample - a randomly selected 50% of Group C used for further cross-validation of multiple cut-off methods.

B. Predictors:

Var. No.	Content	No. of Items	Scoring Method
1	Vocabulary	35	No. Right
2	Reading Comprehension	25	No. Right
3	Spatial Relations	25	R - W/4
4	Surface Development	25	R - W/4
5	Algebra	40	No. Right

C. Criteria

Var. No.	Content	Scoring Method
6	Sum of scaled scores on Entrance Tests (West Point Aptitude, plus Math Achievement, plus English Achievement)	Each test based on scaled score with Mean = 500 SD = 50
7	Pass-or-fail on Entrance Tests	Failure on any one or more of 3 Tests below a scaled score of 450

2. Descriptive statistics and significance tests for population and experimental groups.[1]

Table 2 summarizes the descriptive statistics for each variable and group. The significance tests indicate that Group C is not a random sample from the total population as represented by Group T, but is a selected group with higher mean scores and restricted variability in each variable. This is an expected result. The Designation test results are used by Congressmen to screen candidates before they take the Entrance Examination. The rejection of candidates on the basis of low scores on the Designation tests, which are positively correlated with the Entrance Examination, would result in the higher mean and reduced variability shown by Group C as compared to the total population, Group T.

It is apparent that the mean differences among Groups C1, C2, and C3 are <u>not</u> significant for any of the variables. Therefore, it is judged appropriate to apply regression weights and cutting scores derived from the experimental sample directly to the cross-validation samples.

---

[1] Raw scores for all of Group C on all variables are given in Appendix II.

TABLE 2

Descriptive statistics, estimated reliability[1] and tests of significance of mean differences of variables for all groups

Variable	Item Type	No. of Items	Gr. T (N=436) M	Gr. T (N=436) SD	Gr. C (N=292) M	Gr. C (N=292) SD	Gr. C (N=292) est. rel.	Gr. C1 (N=146) M	Gr. C1 (N=146) SD	Gr. C1 (N=146) est. rel.	Gr. C2 (N=146) M	Gr. C2 (N=146) SD	Gr. C2 (N=146) est. rel.	Gr. C3 (N=146) M	Gr. C3 (N=146) SD	Gr. C3 (N=146) est. rel.
1	Vocab.	35			18.33	4.85	.647	18.66	4.92	.659	18.00	4.75	.631	18.37	4.96	.664
2	Reading	25			12.16	4.09	.653	12.[illegible]6	4.04	.643	11.97	4.14	.612	12.28	4.13	.661
1 plus 2		60	28.78[2]	8.11	30.50	7.81	.767									
3	Spatial	25	17.30	5.31	18.51	4.43	.786	18.90	4.34	.787	18.12	4.47	.782	18.45	4.55	.799
4	Surf. Dev.	25	13.11	5.07	14.64	4.65	.749	15.06	4.76	.767	12.23	4.49	.725	14.27	4.63	.744
3 plus 4		50	30.41	9.43	33.15	8.09	.846									
5	Algebra	40	18.57	8.97	21.68	8.67	.890	21.34	8.76	.870	22.02	8.57	.887	22.31	8.89	.898
6 crit. sum					1585.77	226.05	—	1583.68	[illegible].07	—	1587.86	238.35	—	1600.12	235.54	—
7 Pass-Fail on crit.					.746			.705			.726			.733		

Hypotheses as to mean differences

a) that group C is a random sample from Group T — $P < .001$ for all variables compared

b) that groups C1, C2, and C3 are random samples from group C — $P > .05$ for all variables compared

---

[1]Reliability estimated by Kuder-Richardson formula (1, pg. 154):

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma_t^2 - \frac{M_t}{n}(n - M_t)}{\sigma_t^2}$$

[2]Total score for variables 1 plus 2 only were available for the 2184 candidates from which Group T was sampled.

3. Comparison of multiple regression and multiple cutting-scores for prediction of a continuous criterion.

The Wherry-Doolittle test selection method (27) was applied to the data for Group CI. Table 3 shows the matrix of intercorrelations used in this analysis. The intercorrelations of the two verbal variables (variables 1 and 2) and the two spatial variables (variables 3 and 4) are higher than the average. The spatial variables have the lowest validities.

TABLE 3
Intercorrelations of predictor variables and continuous criterion for experimental Group CI (N=146)

Variable	Predictors 2	3	4	5	Criterion 6
1	.527	.233	.064[1]	.342	.433
2		.286	.130[1]	.219	.461
3			.660	.262	.311
4				.300	.262
5					.563

The Beta weights, and multiple correlation for combinations of predictors, are shown in Table 4. Each R, for one or more variables, is significant at the 1% level. The increase in R for the addition to variables 5 and 2,

[1] these correlations are not significant at approximately the 5% level; for df=150, the correlation at this level should be .159 or better (12, pg. 610).

TABLE 4

Multiple correlation of predictor variables with continuous criterion and order of importance as determined by Wherry-Doolittle Test Selection Method; partial regression coefficients and regression equation for Group C1; correlation of predicted and actual criterion scores for Groups C2 and C3 and computed multiple correlation for Group C2.

Predictor Variables	Multiple Correlation[1]	Beta
5	.563	.4295 (5)
5,2	.658	.2798 (2)
5,2,1	.662	.1224 (1)
5,2,1,3	.665	.0553 (3)
5,2,1,3,4	.663	.0524 (4)

Prediction equation based on Group C1

$$X_6 = 5.30\,X_1 + 14.76\,X_2 + 2.72\,X_3 + 2.34\,X_4 + 10.45\,X_5 + 992.75$$

Cross-validation:

Correlation of predicted and actual criterion scores, Group C2 = .774, C3 = .761

Multiple correlation of predictors with criterion, Group C2 = .782

---

[1]All multiple correlations significant at 1% level. The σ_R for R = .563 is .057; for all other R's, the σ_R = .047.

of variable 1, and particularly 3 and 4, is less than the standard error of R. However, all 5 variables were retained for cross-validation, since there was no significant decrease in R by use of the largest number of variables, and it was desired to study all of the variables by the several techniques employed if at all possible.

Table 4 also shows the prediction equation based on Group C1, and cross-validation results, when predicted criterion scores for Groups C2 and C3 were obtained by the equation and correlated with actual criterion scores. It is an unusual finding that, rather than shrinkage, the correlations between predicted and actual criterion scores in Groups C2 and C3 are higher than in Group C1. (This result is confirmed by the direct computation of R for Group C2.) A possible cause for this may lie in the difference in criterion variance in the two samples. Group C1 has a smaller variance than either Groups C2 or C3, (but not significantly smaller on the basis of F-tests).

Selection of a prediction battery by the Multiple Cutting-Score method required as a first step the definition of the cutting or critical scores. The procedure used for defining such scores on the predictor variables is empirical and judgmental. As described by Grimsley (11) a distribution of criterion scores is prepared, and a number of cumulative criterion groups are established; e.g., top 10%, top 20%, etc. The mean test score of each such (cumulative) group is defined as a selection level. In exploratory work with this method, it was found that if the size of such criterion groups was too small, the differences between successive mean test scores would be so small as to have no practical significance. If too large, the relationships between criterion and test at various segments in the range of scores become difficult to distinguish.

Several definitions of critical scores were studied. One procedure was to establish criterion groupings of equal frequencies prior to cumulation. A second procedure was to categorize the criterion distribution on the basis of standard score groupings prior to cumulation. However, this was found ineffective because of the sharp reduction in the number of cases in the extreme categories. The definition finally adopted was based on categories with equal frequencies such that no category would be smaller than 10% of the total N. It was also considered advisable to establish initially a too large, rather than a too small, number of categories. There is a self-correcting effect if too many categories are used (since mean differences between successive categories approach zero) but not if too few are used.

In the course of exploration of this approach, a more direct method than that used by Grimsley was developed for definition of critical scores. This alternative, (identified as the Revised MCS in this study), defines the critical score directly from categorization of the test score distribution (using as close to equal frequencies as practicable) rather than indirectly through the criterion distribution. This technique has a number of advantages:

a) ordinarily, it is simpler to examine and work with test score distributions in test analysis programs than with criterion data, since the test results are generally arrayed and examined for other purposes.

b) the identity of the particular individuals in a criterion grouping is lost as soon as several tests are combined; thus there seems to be little value in using a criterion grouping in lieu of a more readily obtained test grouping.

c) test categories give a direct indication of percent selected at the defined cut-off scores, whereas criterion groupings require a computation of mean test score, then an examination of the percent who would be selected by that test score. The percent of cases in a criterion group and in the test group based on that critical level varies widely as a function of the correlation between criterion and test.

For these reasons, both the Ruch-Grimsley MCS procedure and the Revised MCS were used for selection of test batteries, and the resulting batteries compared, with the intention to use the most practicable and effective method in the subsequent comparisons with the multiple regression method. Table 5 summarizes the defined critical scores and percent selected for each variable, for both the MCS and Revised MCS methods. For preparation of the tabled data, test and criterion scores were entered on IBM cards for each person. Categorization was made by card-sorting; calculations of mean criterion scores were made directly from the data cards.[1]

The judgment as to selective value is guided by the size of differences in mean criterion score between the successive critical score levels, and by the occurrence of no difference or actual reversals in mean criterion scores between successive levels.[2]

---

[1]The computations are also readily adaptable to IBM tabulating operations; in this case, a listing, in criterion score order, showing cumulative criterion score, and all test scores, and listings for each variable, in test score order, showing cumulative criterion scores, are desirable.

[2]The MCS as described by Grimsley includes the plotting of the data in Table 5, as an aid in selecting tests. Graphs were drawn, but were not found particularly useful or essential in judging the relative value of tests.

TABLE 5

Determination of critical test scores ($\bar{X}$) mean criterion score of selected group ($\bar{Y}$) and percent selected, for each variable, by Multiple Cutting Score and Revised Multiple Cutting Score methods for Group CI, N = 146.

MCS Method: Critical test scores ($\bar{X}$) defined as the mean test score for the specified criterion group.

Selection level & Criterion Group[1]	A top 10%			B top 20%			C top 30%			D top 40%			E top 50%		
Variable	X	Y	%	X	Y	%	X	Y	%	X	Y	%	X	Y	%
1 Vocabulary	24	1736	16	22	1710	26	21	1692	33	20	1689	39	20	--	--
2 Reading	15	1703	29	15	--	--	14	1673	40	14	--	--	14	--	--
3 Spatial	21	1653	43	[illegible]	1628	55	20	--	--	20	--	--	20	--	--
4 Surf. Dev.	18	1601	32	16	1619	54	16	--	--	16	--	--	16	--	--
5 Algebra	32	1816	16	28	1738	27	27	--	--	26	1721	32	25	1708	34

Selection level & Criterion Group[1]	Mn avg. test sc			F bot 50%			G bot 40%			H bot 30%			I bot 20%			J bot 10%		
Variable	X	Y	%	X	Y	%	X	Y	%	X	Y	%	X	Y	%	X	Y	%
1 Vocabulary	19	1656	49	17	1639	65	17	--	--	17	--	--	16	1627	72	14	1606	87
2 Reading	12	1636	60	11	1620	73	11	--	--	10	1612	80	9	1612	82	9	--	--
3 Spatial	19	1626	61	18	1612	67	17	1602	75	18	--	--	17	--	--	16	1604	81
4 Surf. Dev.	15	1625	58	14	1619	66	13	1613	73	14	--	--	12	1616	79	12	--	--
5 Algebra	21	1672	49	18	1650	62	17	1637	68	16	1638	70	15	1624	76	13	1617	84

Revised MCS Method: Critical test scores X defined as the test score reached or exceeded by the specified criterion group.

Selection level & test group[2]	A top 10%			B top 20%			C top 30%			D top 40%			E top 50%		
Variable	X	Y	%	X	Y	%	X	Y	%	X	Y	%	X	Y	%
1 Vocabulary	26	1753	10	23	1710	21	21	1692	33	20	1689	39	19	1656	49
2 Reading	18	1796	10	16	1718	21	15	1703	29	14	1673	40	13	1649	52
3 Spatial	24	1722	15	22	1679	29	21	1653	43	20	1628	55	19	1626	61
4 Surf. Dev.	21	1672	14	19	1646	23	18	1601	32	17	1628	46	16	1619	54
5 Algebra	35	1844	10	31	1786	20	26	1721	32	23	1698	41	21	1672	49

Selection level & test group[2]	F top 60%			G top 70%			H top 80%			I top 90%			J top 100%		
Variable	X	Y	%	X	Y	%	X	Y	%	X	Y	%	X	Y	%
1 Vocabulary	18	1644	58	17	1639	65	15	1618	80	13	1598	91	9	1584	100
2 Reading	12	1636	60	11	1620	73	10	1612	80	7	1604	89	2	[illegible]	100
3 Spatial	18	1612	67	17	1602	75	15	1604	86	12	1592	93	4	1584	100
4 Surf. Dev.	14	1619	66	13	1613	73	11	1609	82	8	1596	91	4	1584	100
5 Algebra	18	1650	62	16	1638	70	14	1619	81	10	1607	91	4	1584	100

[1] Following Grimsley's procedure, below the test mean, critical scores are defined by cumulation from the tail end of the distribution.

[2] Because of the fewer discrete scores in the tests, as compared to the criterion, in order to avoid splitting cases with a given test score, the groups only approximate increments of 10%. The percent selected show the actual percent.

From the MCS data in Table 5, variable 5 shows the greatest differences among the successive groups in mean criterion score, (from 1816 at the A level, to 1617 at the J level), and no reversals. Therefore, it was considered that variable 5 is most selective, followed by variables 1 and 2, then 3 and 4. The Revised MCS approach also selects variable 5 first, then variables 2 and 1, then 3 and 4. For the size of the categories used, there is little distinction to be made between these methods.

The combination of tests for selection of the most predictive battery, for both the MCS and Revised MCS method, was made by taking variable 5 as the base, adding to it a second variable, and computing the mean criterion score and percent selected at the various selection levels. Since variables 3 and 4 appeared relatively ineffective, they were added to the batteries after combinations of 5, 1, and 2 were studied. To the best 2 tests were next added other variables, and so on, until the maximum predictive arrangement was found. The results of this operation are shown in Table 6.

TABLE 6

Mean criterion scores ($\bar{Y}$) and percent selected at various critical score levels for specified test combinations.

MCS Method:

	Selection level																					
	A		B		C		D		E		Mn		F		G		H		I		J	
Variables	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%
5	1816	16	1738	27	1738	27	1721	32	1708	34	1672	49	1650	62	1637	68	1638	70	1624	76	1617	84
5,1	1879	5	1829	10	1814	11	1814	11	1776	16	1696	30	1682	45	1682	45	1663	52	1647	60	1631	76
5,2	1931	7	1870	10	1811	13	1809	16	1689	33	1664	49	1650	60	1639	64	1639	64	1635	71	--	--
5,1,2	2032	3	1952	6	1881	7	1871	10	1752	21	1684	37	1661	44	1665	46	1649	53	1635	67	--	--
5,1,2,3	2057	3	2040	3	2002	5	2002	5	1724	12	1700	27	1667	36	1668	36	1651	42	1636	60	--	--
5,1,2,4	2108	3	2040	3	2040	3	2002	5	1799	11	1712	25	1676	34	1684	35	1674	40	1658	57	--	--
5,1,2,3,4	2088	2	1961	5	1913	6	1922	6	1922	6	1752	15	1688	27	1666	33	1666	36	1663	42	--	--

Revised MCS Method:

	Selection level																			
	A		B		C		D		E		F		G		H		I		J	
Variables	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%	$\bar{Y}$	%
5	1844	10	1786	20	1721	32	1698	41	1672	49	1650	62	1638	70	1619	81	1607	91	1584	100
5,2	2002	2	1850	7	1851	11	1791	18	1717	28	1679	43	1647	56	1638	67	1618	84	1584	100
5,1	2025	2	1860	8	1810	13	1763	20	1696	30	1684	41	1663	52	1639	67	1614	85	1584	100
5,2,1	2021	1	1953	4	1919	8	1871	10	1727	19	1701	30	1661	44	1648	58	1617	82	1584	100
5,2,1,3	2113	1	2075	1	2002	5	2002	5	1767	14	1722	23	1667	36	1649	54	1623	77	1584	100
5,2,1,4	--	--	1952	2	1952	2	1924	6	1787	10	1723	24	1668	37	1664	51	1630	75	1584	100
5,2,1,3,4	--	--	2075	1	2088	2	2002	5	1823	9	1735	21	1671	33	1661	50	1629	74	1584	100

Evaluation of the various combinations was made by inspection of the mean criterion scores and percents selected. A comparison of the two methods for given combinations of variables shows no appreciable differences between them. For example, comparison of the MCS battery 5,1,2 with the Revised MCS battery 5,2,1 shows no large differences, and at a number of points the mean criterion scores and percents selected are identical. Because of this similarity only the Revised MCS data were used as a basis for selection of a final battery for comparison with multiple regression method. The most selective battery was judged to be one which included variables 5,2,1,3 at the A to D levels, and only 5 and 2 at the E to J levels (see Table 6). This battery would select 5% of the group with a mean criterion score of 2002 (D level). For selection of larger percents, tests 5 and 2 appear to yield the highest mean criterion scores, in general. Except for extreme groups in which the percent selected is so small as to be not only unreliable, but of little practical significance, this battery appeared to be most effective over the entire range. It is recognized, however, that the distinctions among the various combinations become small and probably unreliable as the percent selected increases. The cutting scores for this battery are summarized in Table 7.

The Revised MCS battery and the Wherry-Doolittle battery, as described above, were cross-validated and them compared. The cutting scores as shown in Table 7 were applied to Groups C2 and C3, and the percent selected and mean criterion score computed. The multiple regression equation from the Wherry-Doolittle analysis for Group C1 was used to estimate criterion scores for Groups C2 and C3. Then the C2 and C3 groups were arranged in order of predicted criterion scores. From this array, it was possible to compute the actual mean criterion scores for various percents selected, corresponding to the percents selected by the Revised MCS method.

TABLE 7

Cutting-scores for the most predictive battery, selected by the Revised MCS method

	Selection Level									
Variable	A	B	C	D	E	F	G	H	I	J
5	35	31	26	23	21	18	16	14	10	4
2	18	16	15	14	13	12	11	10	7	2
1	26	23	21	20	--	--	--	--	--	--
3	24	22	21	20	--	--	--	--	--	--

The two batteries were compared in the following ways:

a) similarity in tests selected. There is close agreement as to the order of values of the tests. Variables 5 and 2 are of most value, then 1 and 3. Variable 4 was omitted in the Revised MCS (and might have been eliminated from the Wherry-Doolittle also).

b) predictive value of various selection ratios. The data in Table 8 show little difference between the two methods where the percent selected is relatively large. Differences at high selection ratios are consistently in favor of the Wherry-Doolittle method. This holds true for both Groups C2 and C3, but the differences, while consistent in direction, are not very large when compared to the standard deviation of the criterion distribution.

c) stability on cross-validation. The cross-validation stability of the Wherry-Doolittle method has been described earlier in Table 4, in which no shrinkage was found. For the Revised MCS method, Table 8 shows that there are relatively small differences among the three groups with respect to the percent selected and mean criterion scores, except at the highest, and, therefore, least reliable selection levels where the frequencies are very small.

d) practicality of the methods. The Revised MCS method as developed in this study was found considerably easier to apply than the Wherry-Doolittle. The possible savings in any particular program will depend upon the extent to which machine methods are adapted to the computation of correlational data and the application of regression weights, and the extent to

TABLE 8

Percent selected by the Revised Multiple Cutting-Score battery and mean criterion scores ($\bar{Y}$) in Groups C1, C2, and C3; and mean criterion scores for comparable percents selected in Groups C2 and C3 by use of Wherry-Doolittle (WD) multiple regression equation derived from Group C1.

Selection level	Group C1 % sel.	Group C1 $\bar{Y}$	Group C2 % sel.	Group C2 $\bar{Y}$ Rev.MCS	Group C2 $\bar{Y}$ WD	Group C3 % sel.	Group C3 $\bar{Y}$ Rev.MCS	Group C3 $\bar{Y}$ WD
A	1	2113	0	--	--	0	--	--
B	1	2129	1	2203	2203	1	2118	2118
C	4	1997	1	2099	2129	6	1885	1955
D	5	2002	4	1895	2023	8	1881	1933
E	28	1717	27	1786	1810	30	1760	1797
F	43	1679	38	1742	1770	41	1725	1753
G	56	1647	48	1692	1741	51	1685	1735
H	67	1638	61	1672	1701	63	1674	1714
I	84	1618	86	1621	1632	86	1636	1651
J	100	1584	100	1588	1588	100	1600	1600

which time-saving methods, such as use of approximate betas, can be applied. In this study, the time for developing the Revised MCS battery and applying it to the two cross-validation samples was estimated at approximately one-half the time required for the Wherry-Doolittle analysis and cross-validation.

However, the reduction in processing time does not imply that a lesser degree of technical competence in the use and interpretation of the MCS methods is required. As has been pointed out, there are a number of stages in the process for which no adequate criteria or guide lines have yet been developed. A high level of knowledge about the behavior of test scores, and considerable information about the uses to which the data are to be put are considered essential to the proper use of these methods.

4. Comparison of multiple-chi and multiple-R-biserial for prediction of a dichotomous criterion.

In order to evaluate the multiple-chi technique for the selection of combinations of tests and cut-off scores for prediction of a dichotomous criterion, the procedure outlined by Franzen and Lazarsfeld (9) was used with variables 1, 2, and 5. The criterion, described as variable 7 in Tables 1 and 2, was pass-fail on the Entrance test battery. In the experimental group, C-1, the number passing was 103, the number failing, 43.

Each variable was studied at successive cut-off scores, (approximately .5σ intervals) to determine the score which showed the greatest differentiation, in terms of chi[1], between the distribution of the 43 failures and the 146 total competitors. Those cut-off points on each variable were selected which showed the highest chi-values and which did not reject more competitors than the actual rejection rate. For Group C1, 43 out of 146 were rejected by the criterion; thus no cut-off which rejected more than 43 persons was used for later combinations.

---

[1]Chi instead of Chi-square was used, since, with df=1, it may be interpreted as a standard score.

Table 9 summarizes the results of this analysis. All the chi's but two are significant at the .01 level. For combination studies, cut-off scores of 10 and 13 for variable 1, 5 and 7 for variable 2, and 12 and 14 for variable 5 were selected.

TABLE 9

Rejection rates in the failure group (N_F=43) and total group CI (N=146) at specified cut-off scores for each variable, and chi values for comparison of failure group to total group.

Variable	Cut-off Score	Number Rejected: In failure group	Number Rejected: In total group	Chi
1	10 and below	5	8	1.7
	13 " "	12	19	2.9
	14 " "	16	30	2.7
	15 " "	20	41	2.7
2	5 and below	8	10	3.1
	6 and "	10	16	2.6
	7 " "	15	21	3.8
	8 " "	17	26	3.7
	10 " "	23	40	3.8
	11 " "	27	59	3.0
5	8 and below	5	7	2.1
	10 " "	11	17	2.9
	12 " "	13	24	2.8
	14 " "	18	35	3.3
	16 " "	24	27	3.1

Table 10 summarizes the analysis of the combinations of cutting points. (For economy in computing, not every possible combination was examined; certain possibilities, such as variables 1 and 2 alone, were judged from Table 9 not to be effective.)

TABLE 10

Rejection rates in failure group (N=43) and total group CI (N=146) at specified cut-off score combinations, and chi values for comparison of pass-fail classification made by multiple cut-offs with pass-fail on criterion.

Variables	Respective Cut-off Scores	Number Rejected: In failing Group	Number Rejected: In total Group	Chi
1,2,5	10,5,14	23	42	4.3
1,5	10,14	21	39	3.9
2,5	5,14	21	39	3.9
1,2,5	13,7,12	26	45	5.0
1,5	13,12	20	35	4.1
2,5	7,12	22	37	4.6
1,2,5	10,7,14	25	46	4.5
2,5	7,14	24	44	4.3

Regression analysis was made by computing the multiple correlation of variables 1, 2, and 5 with the pass-fail criterion, using biserial correlations for the validity coefficients. For the prediction equation, a dummy variate was set up (10) by coding all pass cases as 1, all fail cases as 0. Accordingly, the mean and standard deviation of this variable for Group CI was .705, s.d. .429. The correlations and prediction equation are given in Table 11.

TABLE 11

Biserial correlations, multiple-R-biserial and regression equation for prediction of pass-fail criterion, based on Group C1 (N=146).

Variable	r_{bis}	
1	.440	$R_{bis} = .628$
2	.501	Prediction equation:
5	.462	$X_7 = .0120X_1 + .0377X_2 + .0165X_5 - .3370$

For cross-validation, the cut-off scores and the prediction equation were applied to Groups C2 and C3. The predicted and actual pass-fail scores were recorded in fourfold tables and tetrachoric correlations computed. These results are summarized in Table 12.

TABLE 12

Relationship between criterion pass-fail categories and pass-fail categories predicted by multiple-chi and multiple-R-biserial methods, for Groups C1, C2 and C3.

Multiple-chi

Group C1

	Criterion fail	Criterion pass	
above cut-offs	17	84	101
below cut-offs	26	19	45
	43	103	146

$r_{tet} = .650$

Group C2

	Criterion fail	Criterion pass	
above cut-offs	13	84	97
below cut-offs	27	22	49
	40	106	146

$r_{tet} = .683$

Group C3

	Criterion fail	Criterion pass	
above cut-offs	16	89	105
below cut-offs	23	18	41
	39	107	146

$r_{tet} = .637$

Multiple-R-biserial

Group C1

	Criterion fail	Criterion pass	
above cut-offs	8	67	75
pred. fail	35	36	71
	43	103	146

$r_{tet} = .670$

Group C2

	Criterion fail	Criterion pass	
pred. pass	3	69	72
pred. fail	37	37	74
	40	106	146

$r_{tet} = .808$

Group C3

	Criterion fail	Criterion pass	
pred. pass	7	75	82
pred. fail	32	32	64
	39	107	146

$r_{tet} = .700$

These results show the multiple-chi method to be more accurate with respect to the total number of failures predicted than the R-biserial, which would reject many more than the actual rate. However, it is less accurate in identifying the actual criterion failures. For example, for Group C1, the multiple-chi method properly identifies 84 of the 103 pass cases, and 26 of the 43 failures, whereas the multiple-R-biserial identifies 67 of the 103 pass cases, and 35 of the 43 failures. Results for Groups C2 and C3 are similar.

From the point of view of selection, it might be more desirable to predict the maximum number of failures, even at the cost of rejecting a sizeable number of pass cases, than to misclassify failures. In this light, the R-biserial is the preferred method for this data.

In terms of cross-validation stability, there is less variation among the three samples for the multiple-chi method, than for the multiple R-biserial.

With regard to ease in application, the multiple-chi procedure was found to be quite simple to compute. It cannot be mechanically used, however. One must consider such factors as the desirable rejection rate and must select likely combinations for trial, in order to reduce labor. The multiple-R-biserial involves as much work as any typical multiple regression problem.

5. Methods for categorizing test score distributions.

In this section are discussed three methods which were applied to the problem of how best to group test scores preliminary to pattern analysis. As pointed out in the discussion of the Multiple Cutting-Score methods, procedures which require grouping together all cases at or above a particular cutting-point in a distribution are not easily adapted to combining test categories when the standards vary for the several tests. When it is desired to combine categories or segments of distributions without regard to their relative order or rank in their original distributions, a more flexible procedure is needed than the Multiple Cutting-Score methods provide.

The study was not originally designed to investigate this problem. However, one of the methods, based on a single-classification analysis of variance, appears, even on the basis of the data used, as a practicable and theoretically sound approach.

a) Comparison of frequency distributions of criterion and test defined by Multiple Cutting-Score (MCS) categories.

In the MCS method, before cumulation of the criterion distribution, the mean test scores made by specified (equal) criterion groups are determined. When these test score cut-offs are applied to categorize the test distribution, a new frequency distribution is generated. It was desired to find the relationship between the test score distribution based on the defined cut-off scores[1] and the rectangular criterion distribution. Presumably, the more similar the criterion distribution and the test distribution associated with it through the defined cutting scores, the more valid the test.

As a test of this relation, the chi-square test of the homogeneity of variances (18) was applied. For comparisons based on 10 criterion groups, each containing 10% of the bases, all of the test score distributions were found to be significantly different in variance from the criterion distributions associated with them ($p < .001$). The criterion groupings, and, consequently, the test groupings, were then made successively coarser and chi-square computed. However, the null hypothesis as to homogeneity of variance was accepted ($p > .05$) only for variable 5 and only with as few as 3 categories.

There are a number of problems in connection with this technique which sharply limit its usefulness here. First, when criterion categories are combined, it becomes necessary to recompute the median or mean test score for the

---

[1]The median, rather than the mean, test score was used to define the cut-off score, because of the small N and the wide variability in the test score arrays for each criterion group.

new larger group, which will be used as the test cut-off score. Second, when categories are relatively refined, reversals occur; that is, the computed median test score for one criterion group may be larger than the median test score for the next higher criterion group. Using such reversed cut-offs results in test categories with zero frequency. Third, for the bottom criterion category, its median test score selects only those persons in the category who scored at or above that median. For those below that median, there is no corresponding frequency in the criterion distribution. These irregularities may be handled in the chi-square computation by combining categories, but the rationale for so doing is not fully satisfactory.

For these reasons, as well as the general computational difficulty, this approach was judged not to be worth further study or cross-validation.

b) Application of the Guttman scale-analysis

The procedure described by Guttman (14) for scale analysis was applied to the data for Group CI for variable 5, the most valid test. It was desired to determine whether a cutting point or points on the test could be located which could sort out test categories related to criterion categories. Therefore, the group was arrayed in test score order on the vertical axis of a tabulation form, and 10 criterion categories defined as the horizontal axis. Each individual was then recorded in the column representing his criterion score. The procedure then was to attempt to adjust and recombine criterion categories so as to find an arrangement in which a parallelogram appeared, and in which Guttman's criterion, that no column have more error than non-error for the cutting point selected, was satisfied. The criterion groups had to be condensed to two broad categories before a sizeable distinction between categories appeared. However, even with this grouping, the criterion of 80% reproducibility was not met.

It is likely that the criteria developed by Guttman and utilized by him and others for the scaling of qualitative data are too rigid for direct transfer to the present problem. The tabulation method does provide a useful summary form from which inferences may be drawn (as from a correlation diagram). However, there is a basic problem presented by the <u>order</u> implicit in the criterion categories, which is not found in scaling discrete items or qualitative varietes.

Since this method did not prove fruitful for the most valid variable, no further attempt was made to apply it to other variables in this study.

c) Categorization and pattern analysis of test scores by an analysis of variance procedure.

This approach to the problem of defining meaningful categories of the test variables is based on the concept that if test score cut-off points can be located which will separate the population into groups which are significantly different in terms of mean criterion score, then those test score cut-off points will provide a defensible and stable set of categories as a basis for pattern analysis. For this purpose, the F-technique was considered applicable to the present data, in a single classification analysis of variance design (19,21). The F-technique is a general test for the significance of group differences in mean scores, irrespective of any logical order among the groups.

There are two major assumptions of this method. The first, a normal distribution of the measurement in the population sampled, is satisfied by the distribution of the continuous criterion for these data.[1] The second, that there is equal variability among the groups, is perhaps more difficult to satisfy. The application and interpretation of tests of significance of mean differences among the groups and between groups is complicated because of the shape of the criterion score distributions within the categories. Since the criterion categories are defined by test score limits, and the test is positively correlated with the criterion, the distribution of criterion scores within a category would not be expected to be the same as the distribution that would result from random sampling.

---

[1]A X^2 test showed the distribution of criterion scores not to depart significantly from normality ($p > .30$).

However, despite this complication, as well as the relatively small N for precise tests of significance, it was considered very desirable to explore the application of the analysis of variance to the present problem. As McNemar points out, there is "some evidence that moderate departure from normality and moderate lack of homogeneity regarding variances do not seriously disrupt the applicability of the technique" (2), pg. 249). The particular way in which the method was carried out is described below.

Group CI was arrayed in test score order, divided into 16 categories, with frequencies of 9, and means and variance estimates for criterion scores computed. An F-test, t-ratios and eta were then computed. The group was then divided into successively larger categories, and F and eta computed for each categorization.[2] Then, t-tests for mean differences between adjoining categories were made, up to the point at which most or all were less than 1.0.

Table 13 summarizes these computations. From this table the optimum categorization was selected for each variable, for pattern analysis. This point was chosen at the categorization at which there appeared to be a sharp

---

[2]This procedure was also carried out with categories based on standard deviation groups with similar results; the equal-n groups are considered preferable, to avoid very small n's in extreme categories, and to simplify calculations.

change in the rate of decrease in F, as well as generally significant t's between categories. Using these criteria, for variables 1, 2, and 5, the optimum groupings were 3, 3 and 4 respectively.

TABLE 13

Analysis of variance of criterion scores for a varying number of categories of the predictor variables. Group C1, N=144.[1]

Var.	Range	No. Cat.	Freq.	F[2]	eta	t for adjoining categories
1	09-31	2	72	18.4	.359	4.3
		3	48	15.0	.413	2.7, 2.7
		4	36	9.5	.411	1.9, 1.0, 2.4
		6	24	7.1	.452	1.8, 1.1, 1.0, 1.3, 1.4
		8	18	5.1	.454	1, 1, 1, 1, 1, 1.8, 1.4
		12	12	6.9	.603	
		16	9	2.4	.467	
2	02-23	2	72	13.1	.304	3.8
		3	48	15.0	.419	1.6, 4.0
		4	36	8.9	.398	1.4, 1.1, 2.4
		6	24	6.5	.437	1.5, 1, 1, 2.3, 1
		8	18	5.4	.465	2.2, 1, 1, 1, 1, 1.0, 2.0
		12	12	3.5	.474	
		16	9	2.7	.492	
5	04-39	2	72	28.1	.406	5.3
		3	48	20.2	.472	2.3, 4.0
		4	36	15.9	.584	2.1, 1.1, 3.5
		6	24	12.8	.562	2.2, 1, 1, 1, 3.7
		8	18	12.1	.620	2.7, 1, 1, 1, 1, 1.3, 4.2
		12	12	6.6	.595	
		16	9	5.6	.631	

[1] For simplicity in computing, the original N of 146 was reduced to 144, by dropping 2 randomly selected cases.

[2] $p < .01$ for all F's

Two methods have been explored for evaluating combinations of test score categories as defined above.

1) In the first, the test scores for variables 1, 2 and 5 were coded into their respective categories. For example, the code for a person scoring in the top categories of variables 1 and 2, and the bottom category of variable 5, would be 114. Then for each unique combination of categories, the frequency, mean criterion score and number of failures were computed.[1] The significance of a particular combination may be inferred by relating it to expected values under the hypothesis of no relationship between unique combinations and criteria. There are a possible 36 patterns. Each should be represented by approximately 4 persons, including 1 failure, whose mean criterion score is the group mean, 1584. Precise tests of significance are not justified with these data, because of the small n's. However, it is illustrative of the method to present the data and to discuss the kinds of inferences that can be drawn.

Table 14 lists these data for Groups C1 and C2. The combination codes are set up in the order of variables 5, 1 and 2, and listed so that the category code changes first for the <u>least</u> valid test. This type of table may be used as follows:

---

[1] Johnson (18) describes a similar method with the addition of a special coding system which was not necessary here, since the position of the digit identifies the variable.

a) by ranking the combinations in order of mean criterion score, that set of combinations can be selected which yields the desired number of persons with the maximum criterion mean.

b) by examining codes only on two of the three digits, the effectiveness of combinations of 2 variables, ignoring the third, may be studied.

c) by combining categories for the same variable, the effectiveness of broader groupings may be studied.

d) particular patterns may be found which include most failures; this would permit use of the patterns especially for identifying failures, and might yield some insight as to the specific psychological causes for criterion failure. For example, to evaluate the effect of cutting off all persons in category 4 on variable 5, all codes from 411 to 433 would be grouped; for Group C1, this would reject 18 of the 43 failures and 17 others.

TABLE 14

Combinations of test categories, mean criterion score ($\bar{Y}$), total number of cases in category (n) and number of failures (n_f), for Groups C1 and C2, N=146 in each group.

Group C1				Group C2			
Combinations	$\bar{Y}$	n	n_f	Combinations	$\bar{Y}$	n	n_f
111	1952	8	0	111	1963	9	0
112	1594	2	0	112	1693	3	0
113	1705	5	0	113	1667	8	1
121	1730	5	1	121	1877	6	0
122	1703	3	1	122	1796	3	0
123	1651	5	2	123	1699	3	0
131	1907	1	0	131	1809	1	0
132	1658	5	0	132	1820	1	0
133	1604	4	1	133	1704	7	0
211	1747	4	0	211	1770	7	0
212	1553	8	2	212	1685	1	0
213	2025	1	0	213	1585	3	1
221	1600	3	2	221	1624	3	0
222	1558	5	0	222	1680	2	0
223	1568	4	1	223	1538	10	4
231	1825	1	0	231	1678	4	1
232	1502	3	0	232	1469	5	1
233	1527	4	2	233	1456	9	6
311	1698	7	0	311	1692	3	0
312	1708	4	0	312	1711	1	0
313	1421	3	1	313	1688	2	0
321	1519	8	2	321	1606	3	0
322	1699	2	0	322	1627	2	0
323	1433	3	2	323	1486	5	1
331	—	0	-	331	1731	1	0
332	1439	5	3	332	1470	2	1
333	1445	8	5	333	1337	8	5
411	1687	2	1	411	1682	3	0
412	1470	2	1	412	1801	1	0
413	1574	2	1	413	1514	1	1
421	1390	2	1	421	1701	2	0
422	1500	3	1	422	1354	4	2
423	1619	4	0	423	1450	6	3
431	1542	1	0	431	1610	1	0
432	1548	3	1	432	1334	2	2
433	1347	16	12	433	1262	14	11

e) the stability of the criterion means and failure rates may be compared between two samples.

Perhaps the most significant feature of this technique, as compared to multiple regression analysis, is that the coded combination score permits some study of the relationships among the test variables and the criterion; that is, it is possible to see the pattern of a person's score, or, in effect, the components of his score, whereas in multiple correlation, it is almost impossible to determine from the total weighted score exactly how the individual achieved it. Although this value has been recognized for the pattern method, when it is coupled with a sound procedure for defining test categories, there appears to emerge a very meaningful and useful analytical program.

2) The second method which has been found practicable[1] is the preparation of a list in the format used for Guttman scale analysis. Such a list is arranged in criterion score order and shows a tally for each person in the test column which identifies his category score. When these columns are grouped so that the top categories in each variable are together, then the second, and so on, if there is a positive relationship between

---

[1] IBM procedures are readily adapted to this type of listing. (20)

the variables and criterion, the tallies should approximate a parallelogram. The headings and entries for such a listing for a 10% sample of Group CI appear in Table 15.

TABLE 15

Sample scale analysis tabulation format, for categorized data, Group CI.

Criterion score \ Variable / Category	5 / 1	11 / 1	2 / 1	25 / 2	1 / 2	21 / 2	5 / 3	21 / 3	2 / 3	5 / 4
2119	1	1	1							
1907		1			1			1		
1825			1	1				1		
1767				1	1				1	
1692						1		1		1
1649		1	1	1						
1613	1					1		1		
1581						1	1	1		
1558				1	1	1				
1532					1				1	1
1475				1		1		1		
1426					1		1		1	
1381								1	1	1
1312				1				1	1	
1201								1	1	1

This type of data may then be further manipulated in accordance with the scale analysis method, or for any other desired purpose.

C. Discussion of results and suggestions for further research.

The results obtained in this study show quite definitely that the empirical methods for combining predictors, specifically the Multiple-Cutting-Score and Multiple-Chi techniques, are practicable for use in selection programs. Their demands, in terms of time and skill level of personnel, are more modest than are made by the more typical regression analysis. IBM and other machine methods may be used for this procedure. However, for the data used here, neither of these methods appeared as favorable as regression analysis in terms of predictive efficiency. In particular, the findings previously reported by Grimsley (11) that the Multiple Cutting-Score seemed superior at high selection ratios was not substantiated here. Probably the difference is due to the fact that in the present study, the data satisfied the assumptions for use of regression analysis, and the variables were both highly reliable and valid. The two methods, Multiple Cutting-Score and multiple regression, were similar in efficiency only for selecting a large proportion of the population. Thus, where the rejection rate is low, the Multiple Cutting-Score method, and particularly the revised technique developed in this study, offers substantial economies in computation.

The empirical nature of the Multiple Cutting-Score and Multiple-Chi methods suggests considerable shrinkage, or at least large differences, in cross-validation. The stability of the cutting-score combinations proved to be one unexpected result. It is recognized, however, that the similarity in the samples used here is perhaps greater than is often found in an operating test program.

The limitation in both these methods results from the need to retain the order of the categories, so that in selecting at a given cut-off score, all persons in categories above that point are considered selected. While this is perhaps essential in public personnel selection program, it definitely restricts the applicability of the methods. Therefore, the second general result of this study, the rationale and exposition of a means for categorizing and combining test scores, takes on added importance. The generality of the method of analysis of variance for test categorization makes it useful both in selection work and in counselling. This technique, coupled with graphic and tabulating devices for analysis of patterns, provides a practical and flexible tool, applicable to other kinds of selection devices besides test scores.

Further research with the categorization and pattern analysis system described here would be highly desirable and very likely profitable, using sufficiently large N's to identify stable patterns. Analysis of the patterns could then be made along several lines. For example, one possibility is to develop means for differentiating occupational or other groups by determining those patterns significantly and differentially associated with those groups. Another challenging line of attack is the use of patterns for developing hypotheses as to the nature, degree and compensatory relations among measurements, with respect to criterion performance.

Further studies of the several cutting score methods should be done under conditions other than those of this study. Since the advantage of these procedures for linearly-related data has not been demonstrated, perhaps their major usefulness does lie in dealing with non-linear relations.

CHAPTER V

SUMMARY AND CONCLUSIONS

This study was designed to investigate the effectiveness of several current methods for the combination of test scores, as compared to multiple regression techniques. For this purpose, a prediction situation was selected in which there were reliable and valid predictors, and both continuous and categorical criteria for a homogeneous test population. The data were the 5 test scores of 292 competitors on the United States Civil Service Commission's battery used by Congressmen for designation of candidates to the United States Military Academy. The criteria were total score on the entrance examination and pass-fail on the entrance examination. The total group of 292 cases was sub-divided into 3 random samples of 146 each. The first sample was used for development of the procedures, the second and third samples for cross-validation.

For prediction of the continuous criterion, a simplified form of the multiple cutting-score method proposed by Ruch was compared to multiple regression analysis. For prediction of the dichotomous criterion, the multiple-chi technique was compared with multiple-biserial analysis. Because of limitations in both these techniques as to the ways in which test categories can be combined, a more general solution to the problem of definition of categories as a basis for pattern analysis was explored.

The following conclusions are offered:

1. A revised multiple cutting-score technique and multiple regression analysis applied to the same data show close agreement in the tests selected and in the order of value of the tests. The multiple regression method is a superior prediction method at high selection ratios. Both methods are relatively stable on cross-validation. The multiple cutting-score method is simple to apply and adaptable to machine computations. Therefore, it is a practicable method for use when the selection ratio is low and economy in computation is an important factor.

2. For prediction of a dichotomous criterion, the multiple-chi method is more stable on cross-validation than the multiple-biserial and is more accurate with respect to the prediction of failure rate. However, the multiple-biserial is superior in identifying actual failures. The multiple-chi technique is relatively simple to compute and adaptable to machine operations.

3. The analysis of variance technique offers a systematic method for determining the optimum number of categories of the criterion distribution which can be defined by test score cut-offs. Procedures for using such categories in combinations, for the purpose of analysis of the unique predictive value of the various patterns, are presented. This

method offers a practicable and rational approach to a basic problem underlying the techniques for combining and evaluating test score patterns.

SELECTED BIBLIOGRAPHY

1. Adkins, Dorothy, et al. Construction and analysis of achievement tests. Washington, U. S. Govt. Printing Office, 1947, 292 pp.

2. Anderson, T. W., Classification by multivariate analysis. Psychometrika, Vol. 16, No. 1, March 1951, pp. 31-49.

3. Barnette, W. L. Jr. Occupational aptitude pattern research. Occupations, Vol. 29, No. 1, 1950, pp. 5-12.

4. Betts, Gilbert L. Test calibration for categorical classification. Educ. and Psychol. Msmt., Vol. 9, No. 3, 1947, pp. 269-80.

5. Cronbach, Lee J. "Pattern tabulation": a statistical method for analysis of limited patterns of scores, with particular reference to the Rorschach test. Educ. and Psychol. Msmt., Vol. 9, No. 2, 1947, pp. 149-172.

6. Cronbach, Lee J. Statistical methods for multi-score tests. J. Clin. Psychol., Vol. 6, 1950, pp. 21-25.

7. Du Mas, Frank M. On the interpretation of personality profiles. J. Clin. Psychol., Vol. 3, 1947, pp. 57-65.

8. Educational Testing Service. United States Military Academy Entrance Examinations, (In) Annual Report, 1950-51, Princeton, New Jersey.

9. Franzen, Raymond. A method for selecting combinations of tests and determining their best cut-off points to yield a dichotomy most like a categorical criterion. Civil Aeronautics Administration, Research Division, Report No. 12, 1943.

10. Garrett, Henry E. The Discriminant function and its use in psychology. Psychometrika, Vol. 8, No. 2, June 1943, pp. 65-79.

11. Grimsley, Glen. A comparative study of the Wherry Doolittle and a multiple cutting-score method. Psychol. Monog., No. 297, 1949.

12. Guilford, J. P. Fundamental statistics in psychology and education. McGraw Hill, N. Y., 1950.

13. Guilford, J. P. and Michael, W. B. The prediction of categories from measurements. Sheridan Supply Co., Beverly Hills, Calif., 1949.

14. Guttman, Louis. The Cornell technique for scale and intensity analysis. Educ. and Psychol. Msmt., Vol. 7, No. 2, 1947, pp. 247-80.

15. Harmon, Lindsay R. Test patterns in the vocational clinic. Educ. and Psychol. Msmt., Vol. 7, 1947, pp. 207-20.

16. Horst, Paul. An analytical formulation of the multiple cutting score technique. (In) Horst, P. et al. The prediction of personal adjustment, Social Science Research Council, Bulletin 48, New York, 1941.

17. Jaspen, N. Serial correlation. Psychometrika, Vol. 11, 1946, pp. 23-30.

18. Johnson, H. M. Multiple contingency versus multiple correlation, an old time-saving way of handling multiple contingency. Am. J. Psychol., Vol. 57, 1944, pp. 49-62.

19. Johnson, Palmer O. Statistical methods in research. Prentice-Hall, New York, 1949.

20. Kahn, L. A. and Bodine, A. J. Guttman scale analysis by means of IBM equipment. Educ. and Psychol. Msmt., Vol. 11, No. 2, 1951, pp. 298-314.

21. McNemar, Quinn. Psychological statistics. Wiley and Sons, New York, 1949.

22. Meehl, Paul E. Configural scoring. J. Consult. Psychol., Vol. 14, 1950, pp. 165-171.

23. Mensh, Ivan N. Statistical techniques in present-day psychodiagnostics. Psychol. Bull., Vol. 47, 1950, pp. 475-92.

24. Mosier, Charles I. Batteries and profiles. (In) Lindquist, E. F. (ed), Educational Measurement, American Council on Education, Washington, D. C., 1951.

25. Richardson, Marion W. The combination of measures. (In) Horst, P. et al. The prediction of personal adjustment, Social Science Research Council, Bulletin 48, New York, 1941.

26. Rulon, P. J. Distinctions between discriminant and regression analysis and a geometric interpretation of the discriminant function. Harv. Educ. Rev., Vol. 21, 1951, pp. 80-90.

27. Stead, William H., et al. Occupational counselling techniques. American Book Co., New York, 1940.

28. Super, Donald. Appraising vocational fitness. Harpers, New York, 1949.

29. Thorndike, Robert L. Personnel selection test and measurement techniques. Wiley, New York, 1949.

30. Thorndike, Robert L. The problem of classification of personnel. Psychometrika, Vol. 15, 1950, pp. 215-37.

31. Thorndike, Robert L. Tests as research instruments. Rev. of Educ. Res., Vol. 21, No. 5, 1951, pp. 450-62.

32. Tiedeman, D. V. The utility of the discriminant function in psychological and guidance investigations. Harv. Educ. Rev., Vol. 21, 1951, pp. 71-80.

33. Toops, Herbert A. Philosophy and practice of personnel selection. Educ. and Psychol. Msmt., Vol. 5, 1945, pp. 95-124.

34. Toops, Herbert A. The use of addends in experimental control, social census and managerial research. Psychol. Bull., Vol. 45, No. 1, 1948.

35. Tucker, Joseph. Relative predictive efficiency of multiple regression and unique pattern techniques. Teachers College, Columbia University, 1950.

36. United States Military Academy, Catalogue. 1951-2, U. S. Govt. Printing Office, Washington, D. C., 1951.

37. Wesman, A. G. and Bennett, G. K. Problems of differential prediction. Educ. and Psychol. Msmt., Vol. 11, No. 2, 1951, pp. 265-272.

38. Wherry, Robert J. Multiple biserial and multiple point biserial correlations. Psychometrika, Vol. 12, 1947, pp. 189-95.

39. __________. Symposium: the need and means of cross-validation. Educ. and Psychol. Msmt., Vol. 11, No. 1, 1951, pp. 4-28.

APPENDIX I

SAMPLE QUESTIONS FOR DESIGNATION EXAMINATION

UNITED STATES CIVIL SERVICE COMMISSION

SAMPLE QUESTIONS FOR DESIGNATION EXAMINATION

The purpose of these questions is to familiarize applicants with the types of questions which will be asked in the Designation Examination. Read these directions carefully; be sure that you understand exactly how the questions in the test are to be answered and how the separate answer sheet is to be used.

In the actual test you will mark your answers on a separate answer sheet similar to the Sample Answer Sheets on this sheet. To answer the sample questions, indicate your answers on the Sample Answer Sheets. After each number on the answer sheet are five pairs of dotted lines labeled A, B, C, D, and E. Read each question carefully; decide which one of the suggested answers is best; then on the separate answer sheet blacken the space between the dotted lines under the letter corresponding to your answer. (Make a solid black mark.) If you make a mistake, completely erase the black mark; do not merely cross it out. Mark only one answer to each question; double answers are counted as incorrect. When you finish the questions, compare your answers with those given in the Correct Answers to Sample Questions.

SAMPLE QUESTIONS

For vocabulary questions choose the suggested word that means most nearly the same as the underlined word means in the illustrative sentence.

1. It seems feasible to start naval maneuvers now.
   FEASIBLE means most nearly
   A) urgent
   B) justifiable
   C) practicable
   D) beneficial
   E) praiseworthy

2. Surveillance of enemy aliens is customary in time of war.
   SURVEILLANCE means most nearly
   A) close supervision
   B) subversive activity
   C) constant protection
   D) unwarranted suspicion
   E) continued confinement

3. The product of $(3m - n)$ and $(3m + n)$ is
   A) $9m^2 - n^2$
   B) $m^2 - 9n^2$
   C) $9m^2 - 6mn^2$
   D) $6m^2 - n^2$
   E) none of these

4. The value of y that satisfies the equation $\sqrt{4y - 3} + 2 = y - 10$ is
   A) -3
   B) -7
   C) 10
   D) 12
   E) none of these

5. The value of y that satisfies the simultaneous equations, $14x - 5y = 31$, $4x + 8y = 56$, is
   A) 1/3
   B) 4
   C) 5
   D) 7 4/5
   E) none of these

6. The roots of the equation $3x^2 - 9x + 6 = 0$ are
   A) irrational and equal
   B) rational and equal
   C) imaginary and unequal
   D) irrational and unequal
   E) rational and unequal

7. Enough iron ore containing 15% pure iron is to be mixed with x tons of iron ore containing 8% pure iron to obtain a mixture of 100 tons containing 10% pure iron. An equation that can be used to find x is
   A) $.15(100-x) + .08x = .10(100)$
   B) $.15x + .08(100 - x) = .10(100)$
   C) $.08x + .15(100 - x) = .10(100 + x)$
   D) $.08x + .10(100 - x) = .15(100)$
   E) $.08(100 - x) + .10x = .15(100)$

Sample Answer Sheet

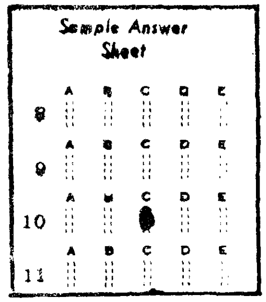
	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					

OVER

In each of the next two questions read the quotation, select the one statement that is best supported by the quotation, and then mark the answer space that has the same letter as this statement.

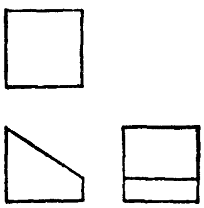
8. (Reading) "The English language is peculiarly rich in synonyms and there is scarcely a language spoken among men that has not some representative in English speech. The spirit of the Anglo-Saxon race has subjugated these various elements to one idiom, making not a patchwork, but a composite language."
Select the alternative that is best supported by the quotation. The English language
A) has few idiomatic expressions
B) is difficult to translate
C) is used universally
D) is composed chiefly of foreign phrases
E) has absorbed words from other languages

9. (Reading) "More patents have been issued for inventions relating to transportation than for those in any other line of human activity. These inventions have resulted in a great financial saving to the people and have made possible a civilization that could not have existed without them."
Select the alternative that is best supported by the quotation. Transportation
A) would be impossible without inventions
B) is an important factor in civilization
C) is still to be much improved
D) is more important than any other activity
E) is carried on through the Patent Office

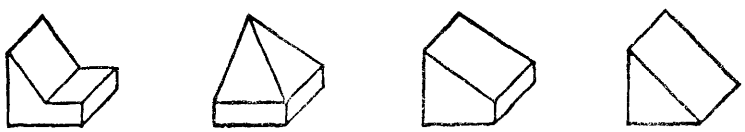


In questions like No. 10 you are to select the one of the drawings of objects, A, B, C, or D, that would have the TOP, FRONT, AND RIGHT views shown in the drawing at the left.

10. TOP



FRONT RIGHT

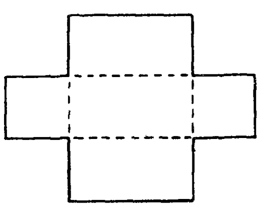


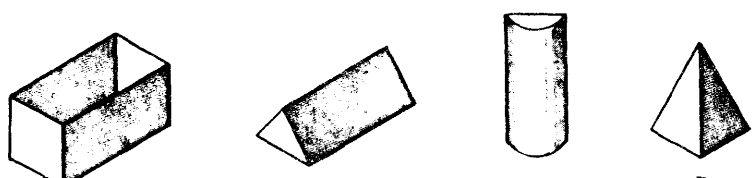
A B C D

In question No. 10, object C looks like the view marked "TOP" when looked at from directly above, and like the views marked "FRONT" and "RIGHT" when looked at from the front and right side respectively. Therefore, the space under C has been blackened for question No. 10 on the Sample Answer Sheet.

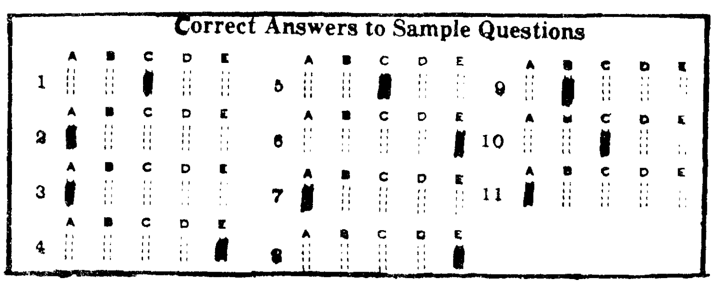
In questions like No. 11 you are to select the one of the drawings of objects, A, B, C, or D, that could be made from the flat piece drawn at the left, if this flat piece were folded on the dotted lines shown in the drawing or rolled.

11.





A B C D



APPENDIX II

List of raw test scores and criterion scores
for total Group C, indicating membership
in random samples C1, C2, C3

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1,C2,C3.
5004	23	14	17	14	19	1636	P	1
5220	23	20	24	13	19	1779	P	1
5380	17	19	20	18	19	1567	P	1
6960	18	15	20	14	18	1640	P	1
6120	30	11	21	20	32	1410	P	1
0144	18	11	20	18	21	1582	P	1
2832	30	14	21	14	21	1419	P	1
0704	28	20	19	14	36	1929	P	1
5500	25	12	21	20	25	1394	F	1
1108	21	16	22	19	22	1688	P	1
6516	19	13	21	15	31	1789	P	1
1476	25	16	19	19	31	1712	P	1
2776	19	18	24	22	33	1597	P	1
6072	25	18	21	17	31	1968	P	1
7684	18	17	25	17	29	1746	P	1
3244	30	23	25	20	38	2113	P	1
3320	26	15	24	21	34	2119	P	1
3532	14	3	11	6	14	1145	F	2
2260	14	7	21	15	12	1107	F	2
6224	16	7	14	18	13	1153	F	2
6664	16	8	16	4	3	1207	F	2
0412	12	10	11	5	19	1194	F	2
4848	20	12	17	13	9	1187	F	2
3800	9	4	15	12	4	1236	F	2
3036	16	13	21	18	24	1295	F	2
6768	17	5	11	16	27	1331	F	2
3344	11	27	14	23	26	1326	F	2
3524	18	9	22	19	21	1347	F	2
4824	12	13	14	12	11	1344	F	2
2588	17	11	20	13	24	1361	F	2
0784	20	8	15	16	6	1349	F	2
7856	14	9	16	20	22	1407	F	2
4480	12	7	22	17	18	1427	P	2
5580	20	9	20	7	25	1451	F	2
2244	18	12	20	12	14	1438	P	2
7568	12	12	20	15	25	1433	P	2
2356	23	10	9	6	22	1460	F	2
4160	14	10	24	21	15	1480	P	2
2228	8	10	17	16	24	1478	F	2
5428	12	10	21	14	12	1493	P	2
3776	15	13	20	15	21	1487	P	2
8700	19	7	10	4	14	1505	P	2
5228	19	11	20	13	26	1516	P	2

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
2948	15	12	24	20	22	1524	P	2
7796	18	14	18	16	26	1526	P	2
7432	18	15	18	12	19	1533	P	2
3972	18	11	12	8	22	1553	P	2
5520	18	15	22	16	13	1583	P	2
0604	22	11	13	16	30	1572	P	2
5452	17	12	16	9	19	1559	P	2
3388	25	15	16	13	25	1594	P	2
8792	18	9	16	16	9	1603	P	2
8112	13	14	19	17	15	1602	P	2
5264	13	12	25	18	27	1608	P	2
3756	16	17	20	10	12	1610	P	2
7348	23	11	21	12	29	1614	P	2
2240	16	10	20	12	31	1641	P	2
4980	19	15	24	14	18	1636	P	2
6100	14	11	22	21	36	1645	P	2
1276	24	16	22	19	12	1672	P	2
6392	16	17	21	24	28	1667	P	2
2400	13	7	23	11	29	1680	P	2
2108	16	9	11	10	36	1684	P	2
5888	24	13	15	7	25	1685	P	2
1936	19	10	22	19	27	1693	P	2
5572	18	12	23	18	18	1694	P	2
0880	21	12	19	17	19	1711	P	2
8714	17	14	13	4	29	1719	P	2
1884	17	8	20	18	8	1729	P	2
7056	13	11	24	19	29	1725	P	2
9008	21	13	24	17	29	1776	P	2
6732	15	13	21	21	29	1820	P	2
2292	26	15	19	14	21	1821	P	2
8736	20	12	20	16	29	1823	P	2
4408	17	14	22	20	36	1847	P	2
3780	24	10	21	17	16	1851	P	2
7860	18	10	18	7	26	1866	P	2
5360	24	17	14	18	21	1882	P	2
7020	28	22	19	15	16	1921	P	2
7888	19	18	17	17	31	1930	P	2
1136	25	19	4	21	39	1964	P	2
8760	27	17	22	16	30	2021	P	2
2992	24	19	22	21	36	2054	P	2
2712	12	6	7	0	5	1022	F	2-3
3288	14	8	23	12	13	1116	F	2-3
8827	18	5	14	10	4	1146	F	2-3

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
0676	19	18	24	17	27	1681	P	2-3
6736	20	19	21	17	24	1681	P	2-3
1528	19	16	20	15	29	1689	P	2-3
1164	21	8	22	17	34	1692	P	2-3
4544	29	16	17	15	22	1716	P	2-3
8820	18	11	17	20	30	1730	P	2-3
1944	10	10	17	17	17	1729	P	2-3
6160	16	17	18	15	20	1731	P	2-3
5960	17	15	12	10	32	1741	P	2-3
0176	14	11	11	12	31	1754	P	2-3
5092	31	21	18	11	11	1770	P	2-3
1240	17	9	17	13	22	1774	P	2-3
8740	22	9	21	14	28	1774	P	2-3
7880	23	13	20	15	29	1776	P	2-3
5104	21	11	14	11	34	1784	P	2-3
5272	12	10	20	16	31	1800	P	2-3
1752	24	14	20	18	9	1801	P	2-3
2716	11	17	18	16	35	1809	P	2-3
4208	18	7	25	21	18	1809	P	2-3
2324	20	17	22	14	13	1819	P	2-3
2732	17	12	15	12	24	1834	P	2-3
0504	19	10	24	18	33	1834	P	2-3
6404	21	17	17	22	35	1835	P	2-3
2516	16	15	13	15	27	1838	P	2-3
0732	21	16	17	19	26	1864	P	2-3
4556	16	16	23	16	26	1868	P	2-3
5712	19	17	24	18	36	1887	P	2-3
2296	26	21	24	19	22	1920	P	2-3
2572	27	15	20	14	30	1888	P	2-3
4416	18	15	25	17	29	1925	P	2-3
2796	21	10	19	16	31	1950	P	2-3
4072	22	16	21	13	37	1995	P	2-3
2628	23	17	19	16	36	2034	P	2-3
6124	19	20	18	18	29	2090	P	2-3

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
2860	32	21	23	18	40	2203	P	2 -3
8984	17	11	18	17	18	1148	F	2 -3
7380	12	11	18	10	16	1171	F	2 -3
3440	9	4	8	8	18	1154	F	2 -3
3660	16	6	10	13	11	1211	F	2 -3
6252	14	9	21	11	11	1235	F	2 -3
1016	15	10	10	5	11	1217	F	2 -3
1768	16	8	11	11	16	1244	F	2 -3
3168	10	8	23	+12	18	1295	F	2 -3
6848	13	12	13	9	12	1324	F	2 -3
5288	13	14	15	10	20	1337	F	2 -3
0364	15	16	22	22	22	1339	F	2 -3
8420	15	3	8	8	8	1347	F	2 -3
6372	8	10	14	12	26	1374	F	2 -3
9116	20	10	22	19	9	1366	F	2 -3
2596	17	13	20	17	10	1395	F	2 -3
1808	18	13	21	17	13	1394	F	2 -3
3560	12	11	21	16	24	1413	F	2 -3
1848	18	8	17	15	20	1416	P	2 -3
6400	16	5	11	16	24	1445	F	2 -3
1612	16	10	15	4	23	1461	P	2 -3
1412	22	10	20	13	29	1480	P	2 -3
8096	17	11	20	17	21	1483	P	2 -3
0292	19	11	18	13	18	1500	P	2 -3
7290	19	21	13	7	25	1511	P	2 -3
3528	23	6	13	13	11	1514	F	2 -3
3620	22	8	17	10	22	1522	P	2 -3
3548	27	19	21	16	15	1525	P	2 -3
7732	23	11	21	20	15	1524	P	2 -3
6636	14	10	11	4	12	1532	P	2 -3
7092	21	12	17	17	32	1528	P	2 -3
0564	18	8	20	9	32	1533	P	2 -3
7116	19	8	20	16	20	1557	P	2 -3
0264	4	5	10	14	27	1588	P	2 -3
7350	21	17	25	10	26	1591	P	2 -3
2788	22	18	17	19	7	1605	P	2 -3
8132	15	9	21	13	27	1608	P	2 -3
6532	21	7	25	17	32	1611	P	2 -3
8412	21	9	21	12	34	1633	P	2 -3
0652	27	18	17	12	20	1630	P	2 -3
0322	10	8	19	16	9	1642	P	2 -3
4220	17	15	22	15	19	1649	P	2 -3
4204	21	16	21	8	30	1672	P	2 -3

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
2532	9	5	4	4	9	1510	P	1-3
1800	12	15	25	17	25	1825	P	1-3
6112	10	12	17	11	32	1613	P	1-3
2916	11	12	21	17	11	1472	P	1-3
6984	16	10	5	4	9	1381	P	1-3
5080	20	16	10	7	12	1484	P	1-3
6600	23	6	25	22	38	1895	P	1-3
5172	23	8	18	6	31	1733	P	1-3
3088	19	8	21	25	29	1625	P	1-3
4976	21	12	22	18	12	1581	P	1-3
8856	21	9	13	18	30	1584	P	1-3
5248	13	10	24	20	13	1616	P	1-3
7276	21	17	17	9	19	1767	P	1-3
3172	17	11	15	9	10	1532	P	1-3
3152	25	14	10	9	29	1601	P	1-3
0644	22	10	13	12	24	2025	P	1-3
4152	24	14	17	7	37	1587	P	1-3
6428	15	12	20	10	21	1475	P	1-3
1204	19	15	20	9	15	1235	F	1-3
3616	16	10	21	12	12	1309	F	1-3
0696	13	11	18	14	21	1312	F	1-3
2444	23	16	12	8	13	1430	F	1-3
3228	19	16	20	6	23	1202	F	1-3
2812	13	5	10	6	10	1201	F	1-3
2120	12	10	22	11	7	1238	F	1-3
7468	11	2	12	8	7	1014	F	1-3
7388	10	10	22	25	15	1543	F	1-3
0396	15	7	17	11	9	1362	F	1-3
4532	10	5	10	5	9	1304	F	1-3
2560	22	15	16	15	37	1778	P	1-3
2100	15	14	22	16	37	1845	P	1-3
1792	16	13	20	16	18	1411	F	1-3
7152	27	14	16	13	17	1576	P	1-3
1572	14	12	22	13	20	1582	P	1-3
7976	18	14	20	20	14	1613	P	1-3
7192	20	10	21	17	24	1727	P	1-3
5100	16	11	25	21	39	1749	P	1-3
3748	26	19	18	17	15	1800	P	1-3
4256	19	9	24	16	32	1889	P	1-3
8428	16	16	15	17	36	1907	P	1-3
8548	21	14	22	18	20	1907	P	1-3
6964	29	21	19	14	24	1864	P	1-3
3824	19	13	21	17	16	1718	P	1-3

List of raw test scores and criterion scores for total Group indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
7520	18	15	20	21	20	1547	P	1-3
7960	18	15	17	13	16	1522	P	1-3
7148	22	11	21	18	20	1491	P	1-3
0268	18	11	24	17	17	1426	F	1-3
3696	17	13	17	16	29	1422	F	1-3
8832	21	12	22	14	25	1459	F	1-3
5884	20	11	19	17	25	1575	P	1-3
1076	18	7	23	21	30	1638	F	1-3
6920	20	13	19	15	24	1655	F	1-3
6988	20	19	22	21	18	1664	P	1-3
6816	19	16	25	17	21	1685	F	1-3
4612	25	18	21	15	20	1707	P	1-3
6092	21	14	21	14	20	1714	P	1-3
8396	18	13	21	21	35	1898	P	1-3
5120	20	15	22	17	34	2032	P	1-3
4516	25	18	21	17	37	1966	P	1-3
3588	31	17	22	21	38	2032	P	1-3
3028	22	16	21	17	26	1786	P	1-3
6704	24	13	22	15	28	1690	P	1-3
0820	26	13	24	16	24	1590	P	1-3
7728	16	14	21	14	30	1574	P	1-3
8652	16	14	20	21	21	1408	F	1-3
2460	23	13	22	16	24	1410	F	1-3
8980	18	13	20	17	23	1558	P	1
7772	18	14	22	17	22	1504	P	1
1008	17	17	13	12	6	1296	F	1
3476	14	10	18	17	7	1283	F	1
5900	13	6	15	11	9	1370	F	1
8540	15	11	24	16	9	1439	F	1
6720	14	5	19	18	5	1026	F	1
4392	10	6	11	12	4	1824	P	1
4508	15	13	21	17	6	1481	P	1
5308	11	7	18	13	14	1192	F	1
0664	10	3	12	12	19	1249	F	1
6304	9	13	22	13	13	1692	F	1
1716	20	6	24	17	13	1575	P	1
2376	10	6	14	20	20	1239	F	1
6996	16	4	15	16	35	1707	P	1
3436	19	5	21	18	26	1386	F	1
7316	14	4	20	20	29	1384	F	1
6884	15	6	17	20	18	1496	P	1
0136	21	5	17	4	14	1532	F	1
3328	27	9	18	7	20	1330	F	1

List of raw test scores and criterion scores for total Group
Indicating membership in random samples C1, C2, C3

Ident. No.	Raw Score on Predictor Variable 1	2	3	4	5	Total Score on Criterion Variable	Pass or Fail on Criterion Variable	Membership in Groups C1, C2, C3.
4012	18	13	15	7	12	1561	P	1
0660	22	16	16	5	26	1649	P	1
0328	19	8	16	14	31	1389	F	1
3876	22	13	10	5	10	1358	F	1
8908	22	8	25	20	35	1903	P	1
5668	15	7	10	8	14	1475	F	1
7328	13	10	20	22	23	1557	P	1
1436	23	11	19	16	10	1616	P	1
7356	19	14	18	13	11	1325	F	1
8532	17	7	17	8	15	1340	F	1
2324	20	12	9	10	21	1553	P	1
6500	14	12	13	15	35	1632	P	1
0256	13	8	20	21	15	1461	F	1
6956	29	21	25	19	12	1943	P	1
7332	17	11	20	10	14	1791	P	1
7272	21	14	14	9	21	1532	P	1
8276	19	16	14	3	31	1439	F	1
4296	17	15	16	11	19	1651	P	1
0756	15	13	9	12	18	1581	P	1
5808	27	14	15	12	26	1927	P	1
7260	14	11	19	19	30	1575	F	1
0512	14	11	21	18	15	1420	P	1
8344	14	10	18	12	28	1878	P	1
5896	14	17	19	18	14	1542	P	1
1796	14	14	24	22	32	1626	P	1
8120	15	11	16	18	18	1658	P	1
2576	15	11	22	18	17	1497	F	1
6860	17	11	22	16	14	1578	P	1
8216	19	11	16	13	17	1533	P	1
6992	17	16	16	17	33	1838	P	1
0628	15	12	25	25	22	1567	P	1
3304	17	16	20	12	23	1824	P	1
5596	19	13	16	16	26	1520	P	1
3404	14	14	17	16	15	1347	F	1
1760	16	13	15	14	18	1273	F	1
7184	20	12	25	19	15	1680	P	1
7128	16	10	22	18	26	1360	F	1
5864	17	11	20	13	38	1713	P	1
4784	28	19	22	14	16	1710	P	1
6896	21	11	18	20	17	1443	P	1
6728	26	18	25	21	17	1538	P	1
5060	23	15	25	21	17	1584	P	1
6220	18	15	21	18	17	1322	F	1

Albert P. Maslow

3525 East Capitol Street, Washington, D. C.

Ph. D, 1952

Date of birth: November 12, 1916

Place of birth: Cleveland, Ohio

Secondary Education: Glenville High School, Cleveland, Ohio

Collegiate Institutions attended	Dates	Degree	Date of Degree
Western Reserve University	1934-8	A.B.	1938
		A.M.	1938
University of Minnesota	1938-9	--	--

Positions held, including present or prospective occupations

Psychologist, U. S. Civil Service Commission, 1939-43, 1946-51

Psychologist, Army of the United States, 1943-46

Asst. Chief, Test Development Section, United States Civil Service Commission, 1951--