DOCUMENT RESUME

ED 040 298 | 24 | VT 011 349

AUTHOR Porter, Andrew C.
TITLE A Chi-Square Approach to Discrimination Among Occupations, Using an Interest Inventory.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO TR-24
BUREAU NO BR-5-0216
PUB DATE May 67
CONTRACT OEC-5-10-154
NOTE 46p.

EDRS PRICE EDRS Price MF-$0.25 HC-$2.40
DESCRIPTORS *Occupational Choice, *Occupational Guidance, *Occupational Tests, *Predictive Ability (Testing), Professional Occupations, Prognostic Tests, Skilled Occupations, *Statistical Analysis, Technical Occupations
IDENTIFIERS *Kuder Preference Record Occupational Form D

ABSTRACT

In a rapidly changing technological society, appropriate choice of occupation becomes increasingly important to the individual as well as to the employer. In this study, the effectiveness of a set of weights, established by the chi-square technique, for distinguishing among similar and dissimilar occupations was compared with the weighting scheme for the Kuder Preference Record Occupational Form D. Data were scored by both sets of weights to determine the percentage of males correctly classified according to occupation by each weighting scheme. The findings indicate that the chi-square weights are superior to the Kuder weights for distinguishing among similar occupations: optometrist, pediatrician, veterinarian, physical therapist, and X-ray technician. The total number of subjects in similar occupations was 1,902. An improper scoring key was used to determine the percentages which were based on Kuder weights for dissimilar occupations; therefore the findings for dissimilar occupations were discounted. The dissimilar occupations were clinical psychologist, social case worker, optometrist, forester, and auto mechanic. Double cross-validation studies of the chi-square derived weights indicate that they are general in nature and not just indicative of the sample upon which the present study was based. (CH)

Technical Report No. 24

A CHI-SQUARE APPROACH TO DISCRIMINATION AMONG OCCUPATIONS, USING AN INTEREST INVENTORY.

Andrew C. Porter

Based on a master's thesis under the direction of
Julian C. Stanley, Professor of Educational Psychology

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

May 1967

A grant of time on the Control Data 1604 Digital Computer from the Research Committee of the University of Wisconsin Computing Center and funds from the Laboratory of Experimental Design for the original data used made the study possible.

The preparation of this report was supported by a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

Center No. C–03 / Contract OE 5–10–154

PREFACE

This technical report is based on the master's thesis of Andrew C. Porter. Members of the thesis committee were Julian C. Stanley, Chairman; Chester W. Harris; and Frank B. Baker.

The primary goal of the Wisconsin R & D Center for Cognitive Learning is to extend knowledge about, and to improve educational practices related to, cognitive learning in children and youth. Controlled experimentation is requisite for achieving this objective. The Laboratory of Experimental Design, part of the technical section of the R & D Center, provides valuable assistance to project directors in the design of experiments and also in the analysis of data. Further, the staff of the LED are charged with extending knowledge about experimental design, scaling procedures, data analysis and the like.

This technical report is the fourth in a series describing new developments in the methodological area. In it, a set of weights, established by the chi-square technique, is empirically compared with the Kuder weighting scheme. Data were scored by both sets of weights, and the percentage of males correctly classified according to occupation for each weighting scheme was determined. The results indicate that the chi-square weights discriminate among two similar occupations better than the Kuder weights.

Herbert J. Klausmeier
Co-Director for Research

CONTENTS

LIST OF TABLES

LIST OF APPENDICES

I

INTRODUCTION AND REVIEW OF THE LITERATURE

INTEREST AS A PREDICTOR

Appropriate choice of occupation is an important consideration in a rapidly changing technological society, not only for the individual directly concerned but also for his employer. In the past, ability has been hypothesized as being the one best variable on which to base a choice of occupation. The most superficial check of the problem reveals that even though a person has the ability required by a particular occupation, his interest may be such that the job is totally unsuitable for him. Therefore, it would appear that interest is another important variable to the successful choice of an occupation.

When using a common group as reference, it has been found that the distribution underlying ability for a particular course of study has greater variability than does the distribution of interest because a narrow distribution is more efficient and has less overlap with other distributions than does a wide distribution. If only one variable is to be considered, interest may be a better predictor of success than ability.

In order to use interest as a predictor of success, quantitative measurement becomes important. Although many attempts have been made to establish such an instrument, the following discussion is limited to consideration of the Kuder Preference Record Occupational, Form D (Kuder, 1961). For the remainder of the present paper, the instrument will be referred as the Kuder.

Research has demonstrated that the discriminatory power of a predictive instrument involves both the method of scoring and the instrument itself (Gaier and Lee, 1953). The following study deals with the problem of obtaining a better scoring system for the Kuder than the one presently used.

DESCRIPTION OF INSTRUMENT USED IN THE STUDY

The purpose of the Kuder is to classify males by occupation, using their interests as the criterion. The instrument consists of 100 triadic items for which the testee indicates the activity he most prefers and the activity he least prefers. The directions are easily understood, the vocabulary is controlled throughout the 100 items, and no time limit is set. The testing time usually required ranges from 25 to 30 minutes. The appropriate population is Grades 9 through 16 and adults. Its scoring device uses empirically based keys for specific occupations; over sixty are available at present, with more in the process of being established. Science Research Associates in Chicago publish the test, copyrighted in 1956 by the originator, G. Frederic Kuder, Professor at Duke University, Durham, North Carolina (Buros, 1959). The test packet contains the test booklet, a manual, a research handbook, and a booklet of computational sheets which were especially designed by the author to make possible the establishment of scoring keys other than those offered by the publisher. The first paragraph in the test booklet perhaps gives a better feeling for what the test actually is than could be given by any amount of factual description.

> This blank is used for obtaining a record of your preferences. It is not a test. There are no answers which are right or wrong for everyone. An answer is right if it is true of you. Please follow the directions carefully (Kuder, 1956).

Any further description of the Kuder or the rationale behind the development of the 100 items is not pertinent to the understanding of the procedures investigated here.

The Kuder does not have answers which are necessarily right or wrong for any given occupation. The problem was how to weight each response in a way that the composite provided for maximum discrimination along a continuum of interest. A solution to the weighting is dependent upon the groups which are to be discriminated among, the equipment available, and the desired limits as to the complexity of the weighting procedure. Kuder (1957) stated that theoretically the best method for solving the weighting problem was "namely, the evaluation of all possible combinations of responses to all the items, an astronomical number." Since he considered such a solution impractical, he devised a scoring key involving unit weights. Further explanation of his weighting procedure is in the following section. Many other empirical studies have been made to find a more efficient approach to the weighting problem. Some of these approaches are explained in the review of literature.

REVIEW OF THE LITERATURE

A review of literature indicates that many techniques have been suggested for the various problems of discrimination. K. M. Cowdery (1925) made one of the early attempts to answer the scoring question. He attempted to solve the problem of discrimination between two groups by using dichotomous items. Citing T. L. Kelley (1923) as a reference, Cowdery weights each response

$$b = \frac{\phi}{(1 - \phi^2)\sigma},$$

"where ϕ is the Phi Coefficient and σ is the standard deviation of the frequencies $(a + b)$ and $(b + d)$ cells" (Cowdery, 1925).

	Response 1	Response 2
Group 1	a	b
Group 2	c	d

Kelley demonstrates that such a weight is a good approximation to the best regression weight for a response. The Strong Vocational Interest Blank bases its scoring technique on the above procedure (Stanley, 1964).

Kuder (1963) proposed that response weights be determined by the use of actual differences between proportions of two groups (A and B) marking each particular response. His equation was

$$W_{(A \text{ vs. } B)} = P_A - P_B.$$

Score for subject "j" was determined by the sum of the weights assigned to the responses of the subject. His equation was

$$X_j = \sum_{i=1}^{100} P_{Aji} - \sum_{i=1}^{100} P_{Bji},$$

where the sum was over all the responses of subject "j". Kuder presented arguments for and against such a scoring technique and reported some empirical findings. The percentages of overlapping of the distributions of difference scores for each of the six occupations treated by Kuder were computed by a method described by Tilton and were reported as ranging from 2 to 20 per cent. Tilton's measure (1937) is approximately twice the percentage of incorrect classifications. These findings indicated that such a scoring technique works very well, but consideration must be given to the limitation of using only two groups at a time.

Rao (1948) discussed the problem of classifying groups according to multiple measurements. His method involves dividing the subject space into three regions for a two-group classification. group one, group two, uncertain. It must be known that the subject belongs in either group one or group two in order to use his discriminant function. Possibly, interest may not meet such a restriction. An advantage of Rao's technique is that discrimination is maximized. The likelihood for a particular response pattern for each group is calculated and used to assign all subjects to the group where the likelihood of a similar pattern is maximum. Approximate tests of statistical significance on the separation of the groups on a particular discriminant function are also possible.

A configural approach is another possibility for solving the problem under consideration. Brigham (1932) demonstrated that important information is probably lost if all persons failing an item are assigned a score of zero. His study suggested that it might be possible to determine patterns of response that are typical of a particular group of subjects. Since that time many attempts have been made to develop methods which use pattern-matching techniques (e.g., Cattell, 1949; Cronbach, 1949; Cronbach & Gleser, 1952; DuMas, 1946; Zubin, 1937). Most of the methods are at best of questionable value. Inadequacies which are typical of the studies are failure to correct for unmet assumptions of linearity, equality of reliability, and equality of intercorrelations among items or tests. Even more important, some methods fail to consider the direction of obtained differences, and a few deal only with configuration without consideration of the level at which the configuration operates (e.g., Gaier & Lee, 1953).

DEVELOPMENT OF A KUDER SCORING KEY

Kuder faced the problem of other researchers when he tried to develop a set of weights that could be used on an instrument that would measure interest as a predictor of membership in a particular occupation. His attempts to answer the weighting question through empirical investigation seemed to indicate that the relative effectiveness of different approaches is a complex function of many variables. As a result he hypothesized that a list of those variables includes the number of cases, the composition of the inventory, the content and type of item used, and the extent to which the items can be considered to be uniformly distributed in the domain represented.

The Kuder uses a rather simple method for developing a set of unit weights which could be used for each occupation to differentiate between a particular occupation and men-in-general. The latter group, which represents the norm group, is composed of one thousand male telephone subscribers who were willing to take the Kuder. The norm group represents a stratified sample of 138 cities and towns which were hopefully considered representative of all sizes of population centers and all sections of the United States (Kuder, 1961, p. 15). Kuder gave no set rule as to the minimum number of cases which are necessary for establishing a set of weights but indicates two hundred as a good "rule of thumb." Certainly a larger number of subjects used to find weights for a particular scoring key will result in a more stable key.

In the development of a key, Kuder first obtains a set of responses to his 100 items from as many subjects as possible in a particular occupation. A subject indicates his responses to an item by marking his answers on an answer sheet. A model for one item follows:

	Most	Least
Activity 1	1 - ||	|| - 4
Activity 2	2 - ||	|| - 5
Activity 3	3 - ||	|| - 6

The answer positions are numbered as indicated in the model above for purposes of discussion. For each of the six answer positions a count is made of the number of subjects who marked a position, resulting in a total for each position. Each count is divided by the total number of subjects, yielding a percentage for each of the first three answer positions and each of the last three answer positions of each item. The corresponding percentages are found for men-in-general. Zubin's inverse arc sine transformation nomograph (1939) is used to determine a "d" statistic which indicates the difference between the percentages of the norm group and the occupational group for each answer position. A negative "d" statistic for an answer position indicates that the percentage for men-in-general group is greater than the percentage for subjects in the particular occupation being considered. The largest "d" for the first three answer positions and the largest "d" for the last three answer positions are checked for significance at the .05 level. The critical value for "d" at the .05 level is determined by the formula

$$d = \left(\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}\right)1.96,$$

where N_1 equals 1000, the size of the norm group, N_2 equals the size of the occupational group, and 1.96 is the unit normal deviate for a two-tailed test of significance at the .05 level. If "d" is significantly large and positive, the item is keyed so that the answer position with the significant "d" is weighted one; if "d" is significantly large and negative, the item is keyed so that the remaining two answer positions from the one yielding the significant "d" are weighted one. Note that the first three positions are dealt with separately from the last three, since they are the ones indicating the "most liked" activity. After the technique has systematically been applied to each of the one hundred items, the resultant weights are used to score cross-validation groups for both the occupational and men-in-general groups. An iterative procedure is used to determine the "best" set of keyed positions. The criterion for this iteration is to minimize the percentage of overlap between the distributions of the two cross-validation groups (Tilton, 1937). These answer positions are each given a weight of one, and a subject's score on the occupational key is the number of his responses which correspond to the keyed positions.

Well-established theory and practice indicate that the use of fractional weights is not practical because the validity or reliability gained is not worth the extra work involved (Kuder, 1961, p. 3). For example, Kuder limited his scoring key to the use of unit weights principally because of his desire to keep scoring time to a minimum. Such mechanical limitation has greatly been relieved in the past few years by the development of high speed digital computers. The use of fractional weights then becomes a practical consideration.

DEVELOPMENT OF A CHI-SQUARE SCORING KEY

The present study proposes the use of a chi-square technique for dealing with the problem of weighting items which are used for discrimination. The technique was developed for specific use on the Kuder and was used to differentiate among more than two occupational groups simultaneously. Using the Kuder, consider a contingency table whose rows represent the several occupations to be discriminated and whose columns represent the six possible response patterns of a Kuder triadic item. A response pattern was defined as one "most" and one "least" answer position, with the restriction that the "most" and the "least" answer positions cannot be chosen for the same activity. Table 1 indicates the answer positions equivalent to each numbered response pattern.

Table 1

Answer Positions Equivalent to Response Pattern

Response Pattern	Answer Positions
1	1 and 5
2	1 and 6
3	2 and 4
4	2 and 6
5	3 and 4
6	3 and 5

An example of the proposed contingency table is shown below.

Occupational Group	Pattern for Item 1	2	3	4	5	6
Doctor						
Lawyer						
Merchant						
Chief						

To establish the entire set of weights for a set of occupations, 100 contingency tables are necessary, a table for each item in the Kuder.

The responses to the Kuder are obtained for as many subjects as possible. Counts of these subjects' response patterns to an item are taken to determine observed cell frequencies in a contingency table similar to the example above. The marginal totals for rows and the marginal totals for columns are found, and one of the two sets of totals is summed to give the grand total. Using these totals, expected cell frequencies are calculated. Expected cell frequency is defined as the frequency one would expect to find if "the two variables were independent of each other, given the marginal totals of the rows and columns" (Ferguson, 1959). More specific to considerations here, expected cell frequencies are the frequencies one would have expected to observe if in fact the occupations for the table did not differ according to interest as measured by the Kuder. The method for calculating the expected cell frequencies is based on the multiplication theorem of probability which states that the joint occurrence of two or more mutually independent events is the product of their separate probabilities. Therefore, the expected probability for a cell is the product of its row marginal divided by the grand total and its column marginal divided by the grand total. The expected cell frequency is the expected cell probability multiplied by the grand total. Symbolically,

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i},$$

where E equals the expected cell frequency and O equals the observed cell frequency. The individual cell contributions to the overall chi-square statistic are the weights proposed, plus a sign, to be used for differentiating among occupational groups simultaneously. For example, in the proposed contingency table above, a subject who indicates his answer as response pattern one receives a weight on the doctor scale of

$$X = \pm \frac{\left(n_{doctors,1} - \frac{n_{doctors}\, n_1}{N}\right)^2}{\frac{n_{doctors}\, n_1}{N}},$$

where $n_{doctors}$ equals the number of doctors, n_1 equals number choosing response pattern one, and N equals total number of subjects on which the table was based. The sign of the above weight is determined by the sign of its unsquared numerator. A subject's total score is the sum over all weights corresponding to the response patterns that he indicated on the instrument.

The use of chi-square weights is defensible in an intuitive manner on a number of points. Since each weight is in a sense a deviation of an observed frequency from an expected frequency (found by using the assumption of independence of patterns and occupations), the weight indicates a peculiarity of that particular response to a particular occupation. If the response is not peculiar to the occupation, the weight is zero. Retaining the sign of the unsquared numerator gives the peculiarity direction. That is to say, a particular response may be peculiarly like an occupation, or it may be peculiarly unlike an occupation. Since each weight involves a summing over both rows and columns, it necessarily operates within a framework specified by the occupations to be differentiated among. Note that if an occupation were to be added to or deleted from the contingency table, it would probably affect every weight in the table. The above approach to the problem of discrimination should be better than pair-wise contrasts by the similar argument that an analysis of variance is better than all possible combinations of t tests.

PROBLEM TO BE INVESTIGATED

The present study was an empirical investigation of the proposed chi-square weighting technique. The problem was to determine the relative efficiency of the chi-square technique to the technique used by the Kuder. The following statements specify the investigation performed in the present study.

Hypothesis 1. A set of weights established by the chi-square technique discriminates among more than two similar occupations better than the weights used by the Kuder.

Hypothesis 2. A set of weights established by the chi-square technique discriminates among more than two dissimilar occupations at least as well as the weights used by the Kuder.

II
METHODS

Several techniques were required to test the two hypotheses which were stated in Chapter I. Because of the similarity of these hypotheses, the same procedures were used for both; however, a different set of data was investigated for each. A set of chi-square weights was developed using the method described in Chapter I. Sets of weights were developed on each of two independent samples in order to double cross-validate. The weights which were developed on a sample were applied to the same sample and to the opposite sample, and the resulting scores were correlated. The percentages of correctly identified males in the cross-validation groups were used as an indication of the efficiency of the weights. The same sample groups of individuals were scored using the Kuder weights, and the percentages of correctly identified males were again determined. The percentages found by the two methods were compared to indicate the correctness of the previously stated hypotheses.

The present study required the use of a high speed digital computer. The Control Data Corporation 1604 Digital Computer at the University of Wisconsin was used for all programs unless indicated. The University of Wisconsin Research Committee granted fourteen hundred dollars worth of computer time, which was equivalent to eight hours on the 1604 computer, for use in the study. During the process of handling the data, all facilities of the University of Wisconsin Computing Center were used. All of the programs written for the study were written in Control Data's FORTRAN 63, which is equivalent to IBM's FORTRAN 4 (Control Data Corporation, 1964).

In order to test the hypotheses under consideration in the paper, large quantities of data were required. The task would have been impossible had it not been for G. Frederick Kuder, Professor at Duke University and originator of the Kuder, who made the necessary data available. His data were particularly useful because they were the same data used for developing his keys.

The data of the present study were responses of all available males for each of nine occupations. These data for nine occupations were divided into two sets; Set I contained five occupations thought to be homogeneous in interest, and Set II contained five occupations thought to be heterogeneous in interest. One of the occupations was used in both Set I and Set II. Set I was the responses of 406 optometrists, 436 pediatricians, 400 veterinarians, 386 physical therapists, and 274 x-ray technicians. Set II was the responses of 500 clinical psychologists, 452 social case workers, 406 optometrists, 348 foresters, and 298 automobile mechanics. The total number of subjects was 3906.

Kuder's data were on both cards and answer sheets. The data on cards were in four different formats, one of which involved double punching. A single format, requiring two cards per individual, was chosen, and the data not already on cards were manually punched. Three computer programs were written to rearrange Kuder's data and punch it out in the chosen format of the study (See Appendix E, F, and G). The cards which were double punched presented a greater problem. Equipment was not available to read these cards; therefore, an IBM 407 Tabulator was wired to print out the data by scanning certain sections of the card at a time. These printed sheets of data were then manually punched in the chosen format (See Appendix H). In order to make the data easier to use, an IBM 1460 Computer was used to put the data of the 7812 cards on magnetic tape as card images. A program was written to check each card image for particular types of errors, to count the number of card images, and to print out certain card images, thus making the tape well edited (See Appendix I). Each occupation was split into two random halves by means of Program RANDOM, which was written especially for the study. (See Appendix J.) Random halves, resulting from this program, will be used later in the study of cross-validation, which requires two groups. Program RANDOM took the subjects in an occupation and randomly assigned them to positions in an array by use of a random number generator called IRAN (See Appendix K). This generator was obtained from the Computer Science Department of the University of Wisconsin. IRAN generates random numbers, from one up to and including N, where N is the argument required to call the subroutine. The array built by RANDOM was then written on tape, one half at a time, and labeled according to Set I or II, occupation, and random Half A or B.

Programs WTONEA, WTONEB, WTTWOA, and WTTWOB were written to yield the chi-square weights based on each of the four groups of data (Appendix L). All four programs are equivalent except for the part that handles input. The data are read in to build a three-dimensional array of observed frequencies. The dimensions of the array are occupation, response, and item. Within each item, marginal totals are found for each row and each column. A grand sum is found by summing over either one of the sets of marginal totals. The chi-square values for each cell were computed, and the unsquared signs of the numerators were kept. These values were put in their appropriate cells in place of the original frequencies. A sum was made over all cells within each item to yield 100 chi squares, one for each item in the Kuder. The three-dimensional array was both printed and punched for output.

Program SCORE was written to use the weights from the above programs as input, plus each individual's set of responses (See Appendix M). Program SCORE was used for four runs to yield the total scores for each individual on each occupation. In each of the four runs, the individuals were scored on the weights derived from the individuals in the opposite half. Since the halves were independent of each other, the resultant scores were not due to any "overfitting" of the data.

The criterion of highest raw score was used as the basis for deciding in which occupation an individual should be placed. In order to obtain a measure of the efficiency of the weights developed, a count was made of the number of people in a given profession who actually scored the highest on that profession. These counts were made for each of the four groups of data.

The procedure of double cross-validation (Mosier, 1951) was used to determine the stability of the scores resulting from the chi-square technique. For example, the total scores for Set I, Half A, using the weights derived on Set I, Half A, were correlated with the total scores of Set I, Half A, using the weights derived on Set I, Half B. The intercorrelations of occupations within a particular set and half were computed; thus total scores of Set I, Half A, using weights derived on Set I, Half B, were correlated with themselves. These correlations show the similarities of the occupational keys within a set. The standard library statistical program (Unified Means, Standard Deviations, and Correlation Program) was used to find the correlations (Schacht, 1964).

In order to compare the chi-square weights with the Kuder weights, it was necessary to have the data scored using the Kuder keys, which Dr. Kuder supplied along with his original data. A program titled SCORE K was written to use Kuder Scoring Keys to score each individual on all nine occupations considered in the study (Appendix N). The Kuder key weights answer positions rather than response patterns. Therefore, each Kuder key had to be coded so that the six possible response patterns received a weight determined by the number of keyed answer positions included in that pattern. The possible weights were 0, 1, or 2 points for any one item.

Consider the item

|| ||
|| ||
|| || ,

with positions 1, 2, and 6 keyed, using the Kuder technique. The coding for the above item would be that response patterns one through six receive weights of 1, 2, 1, 2, 0, 0, respectively. The nine Kuder keys were coded in this manner, and a card deck of the weights was manually punched. The input for SCORE K was the card deck of weights and individuals' responses which were on tape. The output was scores, both punched and printed, for each of the 3906 individuals.

Since the total number of points possible was not equal across occupational keys, direct comparisons of the raw scores resulting from Program SCORE K were not possible. Program COMPAR was written to take each raw score on a particular occupation and divide it by the total number of points possible on the Kuder key for that occupation (Appendix O). Each individual's score on each occupation was then a percentage of the total possible on that occupation. COMPAR then determined which percentage was the highest for the set of five occupations in which a particular individual's occupation fell. A count was then made of the number of people in a given profession who actually were picked by COMPAR as belonging in that occupation. These decisions were again made on a "hit or miss" basis. In other words, an individual either was or was not picked correctly, using his actual occupation as the criterion.

The percentages of correctly identified individuals, using chi-square weights, were then compared with the percentages of correctly identified individuals, using the Kuder weights. These comparisons were made by using a procedure for determining the significance of the difference between two correlated proportions (McNemar, 1949), which is a special test for comparing proportions when both are based on the same sample of individuals (Ferguson, 1959). The technique allows for the correlation between the paired observations.

III
RESULTS

In testing the stated hypotheses, the resulting data were tabulated and arranged in meaningful form. Only results which were directly necessary for the conclusions made in the paper were reported. Much valuable intermediate data, not reported in the paper, has been kept in printed form and will be made available by the writer to interested researchers for three years. Probably within these data are the answers to many questions not considered here. Available are frequencies necessary for calculating chi squares, individual chi-square weights, over-all chi-square statistics, and total scores for each individual scored on each occupation in his set, using weights derived on Half A and weights derived on Half B.

Four correlation matrices were computed in the cross-validation study. Within Set I the total scores for males in Half A, scored on the weights derived on Half A, were correlated with the total scores of males in Half A, scored on the weights derived on Half B (Table 2). Within Set I the total scores for males in Half B, scored on the weights derived on Half A, were correlated with the total scores of males in Half B, scored on the weights derived on Half B (Table 3). Within Set II the total scores for males in Half A, scored on the weights derived on Half A, were correlated with the total scores of males in Half A, scored on the weights derived on Half B (Table 4). Within Set II the total scores for males in Half B, scored on the weights derived on Half A, were correlated with the total scores of males in Half B, scored on the weights derived on Half B (Table 5). The entries in these tables of interest to cross-validation are in the main diagonals; that is, the correlation of an occupation with itself when scored on two different keys. The correlations range from .9528 for veterinarians to .8057 for physical therapists for Set I, Half A, and from .9511 for veterinarians to .7906 for physical therapists for Set I, Half B. The correlations for Set II range from .9865 for clinical psychologists to .9088 for optometrists for Half A, and from .9867 for clinical psychologists to .9039 for optometrists for Half B. These values probably have meaning to only two decimal places of accuracy. The high correlations indicate that the degree of "overfit" of any one set of weights to the sample on which they were derived is quite small. This means that the total scores obtained by using any one of the four sets of weights in this study were general in nature, not just applicable to a particular sample. An-

Table 2

Correlations of Total Scores for Males in Set I, Half A,
Scored on Weights Derived on Set I, Half A, with
Total Scores for Males in Set I, Half A,
Scored on Weights Derived on Set I, Half B

A on A	A on B				
	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	.9007	-.0826	-.5056	-.1374	-.1696
Pediatrician	-.1036	.9157	-.2651	-.1558	-.4838
Veterinarian	-.4653	-.3392	.9528	-.4155	-.0651
Physical Therapist	.0081	.0737	-.5762	.8057	.2129
X-ray Technician	-.2542	-.5647	.0141	.3014	.8650

Note.—The entries in these tables of interest to cross-validation are in the main diagonals; that is, the correlation of an occupation with itself when scored on two different keys.

Table 3

Correlations of Total Scores for Males in Set I, Half B, Scored on Weights Derived on Set I, Half A, with Total Scores for Males in Set I, Half B, Scored on Weights Derived on Set I, Half B

B on A	B on B				
	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	.9037	-.0736	-.5403	-.0932	-.1720
Pediatrician	-.0656	.9263	-.2927	-.1600	-.4651
Veterinarian	-.4862	-.3535	.9511	-.4050	-.0639
Physical Therapist	.0232	.0420	-.5572	.7906	.2651
X-ray Technician	-.2624	-.5612	.0387	.2853	.8610

Note.—The entries in these tables of interest to cross-validation are in the main diagonals; that is, the correlation of an occupation with itself when scored on two different keys.

Table 4

Correlations of Total Scores for Males in Set II, Half A, Scored on Weights Derived on Set II, Half A, with Total Scores for Males in Set II, Half A, Scored on Weights Derived on Set II, Half B

	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	.9865	.6997	-.0603	-.7811	-.8994
Social Case Worker	.6774	.9752	-.1718	-.8190	-.7299
Optometrist	-.2360	-.3457	.9088	.0293	.0645
Forester	-.7567	-.8483	-.1341	.9853	.7395
Auto Mechanic	-.8996	-.7346	-.1367	.7284	.9873

Table 5

Correlations of Total Scores for Males in Set II, Half B, Scored on Weights Derived on Set II, Half A, with Total Scores for Males in Set II, Half B, Scored on Weights Derived on Set II, Half B

	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	.9867	.6996	-.0202	-.7700	-.9038
Social Case Worker	.6854	.9751	-.1040	-.8252	-.7461
Optometrist	-.2321	-.3066	.9039	-.0106	.0458
Forester	-.7590	-.8475	-.1724	.9837	.7480
Auto Mechanic	-.9038	-.7400	-.1822	.7289	.9894

Table 6

Percentages of Males in a Particular Occupation in Set I, Half A, Classified into Each of the Five Occupations in Set I, Using the Chi-Square Weights Derived on Set I, Half B

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physicial Therapist	X-ray Technician
Optometrist	66.50	15.76	7.88	7.39	2.46
Pediatrician	6.88	64.22	18.81	6.88	3.21
Veterinarian	5.50	5.00	86.50	3.00	0.00
Physical Therapist	17.10	22.28	10.36	41.97	8.29
X-ray Technician	13.87	7.30	19.71	21.90	37.23

Table 7

Percentage of Males in a Particular Occupation in Set I, Half B, Classified into Each of the Five Occupations in Set I, Using the Chi-Square Weights Derived on Set I, Half A

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	54.19	13.79	17.24	6.40	8.37
Pediatrician	3.67	64.22	20.64	5.50	6.00
Veterinarian	3.00	6.00	89.50	1.00	0.50
Physical Therapist	11.40	15.54	18.65	36.79	17.62
X-ray Technician	9.49	8.76	21.17	10.95	49.64

other method for measuring the shrinkage in predictive effectiveness of weights when applied to the sample on which they were derived versus some other sample was to compare the main diagonal entries of Table 6 with Appendix A, Table 7 with Appendix B, Table 8 with Appendix C, and Table 9 with Appendix D. The Appendix values are the percentages correctly identified, using weights which take advantage of the "overfit" of the data. These tables show percentages when these weights were applied to the cross-validation group. The keys, indicated as stable by the correlations, were generally the keys that had less shrinkage when applied to the cross-validation group. The decrease in the percentages of correctly identified males ranged from 27.00 to -5.42 for Set I and from 16.38 to 2.02 for Set II.

In order to study the inter-correlations of the occupational keys developed in this study, the total scores, resulting from scoring one of the halves using the weights derived on the opposite half, were correlated with each other (Appendixes P, Q, R, and S). The off-diagonal elements were for the most part negative, which indicated that these keys were good for discriminating among occupations. A few exceptions were physical therapist and X-ray technician in Set I, clinical psychologist and social case worker in Set II, and forester and auto mechanic in Set II. Averages for inter-occupational correlations were -.2161 for Set I, Half A; -.2245 for Set I, Half B; -.2356 for Set II, Half A; and -.2337 for Set II, Half B.

Using the criterion of highest score, the percentages of correctly identified males were found for each occupation in each of the two sets and each of the two halves, using the weights established on the opposite half. The findings are reported for chi-square weights in Tables 6, 7, 8, and 9, and for Kuder weights in Tables 10, 11, 12, and 13. The main diagonals of these tables reported the percentages of males in a given occupation that were correctly chosen as being in that occupation. The off-diagonal elements indicated the percentage of that occupation which were misclassified as one of the four other possible occupations.

Table 8

Percentages of Males in a Particular Occupation in Set II, Half A, Classified into Each of the Five Occupations in Set II, Using the Chi-Square Weights Derived on Set II, Half B

Actual Occupation	Test Indicated Occupation: Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	86.40	7.20	1.60	.40	4.40
Social Case Worker	32.74	52.65	3.54	3.98	7.08
Optometrist	22.66	3.94	39.41	3.94	30.05
Forester	8.05	1.15	1.72	43.68	45.40
Auto Mechanic	0.00	.67	1.34	2.68	95.30

Table 9

Percentages of Males in a Particular Occupation in Set II, Half B, Classified into Each of the Five Occupations in Set II, Using the Chi-Square Weights Derived on Set II, Half A

Actual Occupation	Test Indicated Occupation: Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	88.40	4.00	.80	2.00	4.80
Social Case Worker	29.20	59.73	1.33	2.21	7.52
Optometrist	21.18	8.37	42.36	2.46	25.62
Forester	8.05	2.87	1.72	44.83	42.53
Auto Mechanic	1.34	0.00	.67	2.01	95.97

Table 10

Percentages of Males in a Particular Occupation in Set I, Half A, Classified into Each of the Five Occupations in Set I, Using the Kuder Weights

Actual Occupation	Test Indicated Occupation: Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	41.87	9.36	31.03	10.84	6.90
Pediatrician	9.63	46.79	31.65	8.26	3.67
Veterinarian	3.00	5.50	86.00	3.50	2.00
Physical Therapist	11.40	11.40	23.83	40.41	12.95
X-ray Technician	5.84	5.84	39.42	10.95	37.96

Table 11

Percentages of Males in a Particular Occupation in Set I, Half B, Classified into Each of the Five Occupations in Set I, Using the Kuder Weights

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	49.75	7.39	28.57	8.37	5.91
Pediatrician	11.01	47.25	30.28	7.80	3.67
Veterinarian	3.00	5.00	87.50	3.50	1.00
Physical Therapist	9.33	8.81	36.27	32.64	12.95
X-ray Technician	7.30	8.76	36.50	8.03	39.42

Table 12

Percentages of Males in a Particular Occupation in Set II, Half A, Classified into Each of the Five Occupations in Set II, Using the Kuder Weights

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	12.40	44.80	40.80	2.00	0.00
Social Case Worker	.44	74.34	19.47	3.54	2.21
Optometrist	0.00	5.91	75.86	8.37	9.85
Forester	0.00	2.30	9.20	81.61	6.90
Auto Mechanic	0.00	3.36	2.68	8.72	85.23

Table 13

Percentages of Males in a Particular Occupation in Set II, Half B, Classified into Each of the Five Occupations in Set II, Using the Kuder Weights

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	7.60	48.00	37.20	6.80	.40
Social Case Worker	0.00	75.22	20.80	2.65	1.33
Optometrist	0.00	7.88	73.40	9.36	9.36
Forester	0.00	5.17	4.60	83.33	6.90
Auto Mechanic	0.00	.67	5.37	7.38	86.58

These tables clearly indicated the occupations which were most difficult to identify. These tables also indicated which occupations caused the difficulties in discrimination.

As was expected from the preceding results, within a set the percentage of classifications by occupation, although not identical, was quite similar for Half A, scored on B weights, and Half B, scored on A weights. Testing for significant differences between the matched proportions of these pairs of tables seemed unimportant. A comparison of the corresponding pairs of tables reporting Kuder weights would have no meaning because these tables were the results of applying a standard set of weights to two random halves in each of two sets of data. The apparent differences could only be attributed to differences in random samples. However, in a crude sense, the comparison would check Program RANDOM.

The percentage of correctly identified males in Set I (homogeneous group) using chi-square weights ranged from 86.50 to 37.23 for Half A and 89.50 to 36.70 for Half B. The percentage of correctly identified males in Set II (heterogeneous group) using chi-square weights ranged from 95.30 to 39.41 for Half A and from 95.97 to 42.36 for Half B. As was expected from the preceding findings in the correlation matrices, clinical psychologist was a major distractor for picking forester. Somewhat surprising to note, however, was that the reverse of these findings was not true. For example, social case worker was not a major distractor for picking clinical psychologist. Other major distractors were clinical psychologist and auto mechanic for picking optometrist. In Set I distractors were not nearly as pronounced. The percentage of misclassifications were more evenly distributed over the remaining four occupations.

By using the data in Tables 6 through 13, comparisons of the discriminating power of the two techniques under consideration were possible. Kuder classification percentages were subtracted from chi-square classification percentages, and the resulting differences in percentages were reported (see Tables 14, 15, 16, and 17). The entries of major interest were in the main diagonal of each table. Positive entries indicated that the chi-square weights provided better discrimination, and negative entries indicated that the Kuder weights provided better discrimination.

Testing the statistical significance of the difference of two proportions based on the same sample of males then became necessary. The data may be represented

Frequencies

		Half B		
		Miss	Hit	
Half A	Hit	A	B	A + B
	Miss	C	D	C + D
		A+C	B+D	N

Proportions

		Half B		
		Miss	Hit	
Half A	Hit	a	b	p_1
	Miss	c	d	q_1
		q_2	p_2	1.00

(Ferguson, 1959, p. 149). For the problems considered here

Table 14

Kuder Classification Percentages
Subtracted from Chi-Square Classification Percentages
on Set I, Half A

Actual Occupation	Test Indicated Occupation: Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	24.63	6.40	-23.15	-3.45	-4.44
Pediatrician	-2.75	17.43	-12.84	-1.38	- .46
Veterinarian	2.50	- .50	.50	- .50	-2.00
Physical Therapist	5.70	10.88	-13.47	1.97	-4.66
X-ray Technician	8.03	-1.46	-19.71	10.95	- .73

Table 15

Kuder Classification Percentages
Subtracted from Chi-Square Classification Percentages
on Set I, Half B

Actual Occupation	Test Indicated Occupation: Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	4.44	6.40	-11.33	-1.97	2.46
Pediatrician	-7.34	16.97	- 9.64	-2.30	2.33
Veterinarian	0.00	1.00	2.00	-2.50	- .50
Physical Therapist	2.07	6.73	-17.62	4.15	4.67
X-ray Technician	2.19	0.00	-15.33	2.92	10.22

Table 16

Kuder Classification Percentages
Subtracted from Chi-Square Classification Percentages
on Set II, Half A

Actual Occupation	Test Indicated Occupation: Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	74.00	-37.60	-39.20	- 1.60	4.40
Social Case Worker	32.30	-21.69	-15.93	.44	4.87
Optometrist	22.66	- 1.97	-36.45	- 4.43	20.20
Forester	8.05	- 1.15	- 7.48	-37.93	38.50
Auto Mechanic	0.00	- 2.69	- 1.34	- 6.04	10.07

Table 17

Kuder Classification Percentages
Subtracted from Chi-Square Classification Percentages
on Set II, Half B

Actual Occupation	Test Indicated Occupation: Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	80.80	-44.00	-36.40	- 4.80	4.40
Social Case Worker	29.20	-15.49	-19.47	- .44	6.19
Optometrist	21.18	.49	-31.04	- 6.90	16.26
Forester	8.05	- 2.30	- 2.88	-38.50	35.63
Auto Mechanic	1.34	- .67	- 4.70	- 5.37	9.39

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{a+b}{N}}}$$ is a unit normal deviate,

where $\sqrt{\frac{a+b}{N}}$ is the standard error which takes into account the correlation between the paired observations. The computation of the above statistic for a large N requires a great deal of effort. A conservative estimate of this standard error was much easier to compute and just as appropriate, since large differences are the only ones of interest. (See Appendix T for the rationale of the standard error used in the following tests of significance.)

For Set I, Half A, the chi-square weights classified optometrist and pediatrician significantly better than Kuder weights at the .05 level. For Set I, Half B, the chi-square weights classified pediatrician significantly better than Kuder weights at the .05 level. Other comparisons were nonsignificant at the .05 level. For Set II, Half A, the chi-square weights classified clinical psychologist and auto mechanic significantly better than did the Kuder weights at the .05 level. The Kuder weights, however, classified social case worker, optometrist, and forester significantly better than did the chi-square weights at the .05 level.

The preceding tables give percentages calculated on each separate half. In order to present the overall picture and gain stability, the percentages for Half A were averaged with the percentages for Half B. The entries in Table 6 were pair-wise averaged with the percentages in Table 7, and the results reported in Table 18; the entries of Table 8 were pair-wise averaged with the percentages of Table 9 and the results reported in Table 19; the entries of Table 10 were pair-wise averaged with the entries of Table 11 and the results reported in Table 20; the entries in Table 12 were pair-wise averaged with the entries in Table 13 and the results reported in Table 21. Comparisons of the discriminating power of the two techniques under consideration were then possible, using the resulting average percentages in Tables 18 through 21. Average Kuder classification percentages were subtracted from average chi-square classification percentages, and the resulting differences in percentages were reported (Tables 22, 23). Again the entries of major interest were in the main diagonal of each table. Positive entries indicated that the chi-square weights provided better discrimination, and negative entries indicated that the Kuder weights provided better discrimination.

A look at the data showed one quite implausible result. The percentage of correctly identified male clinical psychologists was less than chance when the Kuder scoring keys were used. Since these data appeared questionable, another key was immediately ordered from Science Research Associates for comparison with the key used by the present study. As suspected, the key used in the study (It had been obtained directly from Dr. Kuder.) was faulty, but the new key arrived too late to use with the relevant data and make all of the necessary corrections. Therefore, the results reported in Tables 12, 13, 16, 17, 21, 23 are incorrect for the first row and first column. Possibly the fact that a key other than clinical psychologist was used may have affected all of the results in those tables but only because some occupation other than clinical psychologist was the distractor for picking the remaining four occupations. The magnitude of such an effect could not be determined.

Table 18

Averages of Percentages of Males in a Particular Occupation In Set I, Half A, and Set I, Half B, Classified into Each of the Five Occupations in Set I Using Chi-Square Weights

Actual Occupation	Test Indicated Occupation: Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	60.35	14.78	12.56	6.89	5.42
Pediatrician	5.28	64.22	19.73	6.29	4.60
Veterinarian	4.25	5.50	88.00	2.00	0.25
Physical Therapist	14.25	18.91	14.50	39.38	12.95
X-ray Technician	11.68	8.03	20.44	16.43	43.43

Table 19

Averages of Percentages of Males in a Particular Occupation in Set II, Half A, and Set II, Half B, Classified into Each of the Five Occupations in Set II Using Chi-Square Weights

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	87.40	5.60	1.20	1.20	4.60
Social Case Worker	30.97	56.19	2.43	3.09	7.30
Optometrist	21.92	6.15	40.89	3.20	27.83
Forester	8.05	2.01	1.72	44.25	43.96
Auto Mechanic	.67	.33	1.00	2.35	95.63

Table 20

Averages of Percentages of Males in a Particular Occupation in Set I, Half A, and Set I, Half B, Classified into Each of the Five Occupations in Set I Using the Kuder Weights

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	45.81	8.37	29.80	9.61	6.40
Pediatrician	10.32	47.02	30.97	8.03	3.67
Veterinarian	3.00	5.25	86.75	3.50	1.50
Physical Therapist	10.37	10.10	30.05	36.52	12.95
X-ray Technician	6.57	7.30	37.96	9.49	38.69

Table 21

Averages of Percentages of Males in a Particular Occupation in Set II, Half A, and Set II, Half B, Classified into Each of the Five Occupations in Set II Using the Kuder Weights

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	10.00	46.40	39.00	4.40	.20
Social Case Worker	.22	74.78	20.13	3.09	1.77
Optometrist	0.00	6.89	74.63	8.87	9.60
Forester	0.00	3.73	6.90	82.47	6.90
Auto Mechanic	0.00	2.01	4.02	8.05	85.90

Table 22

Average Kuder Classification Percentages Subtracted from Average Chi-Square Classification Percentages on Set I

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	14.54	6.41	-17.24	-2.72	- .98
Pediatrician	-5.04	17.20	-11.24	-1.74	.93
Veterinarian	1.25	.25	1.25	-1.50	-1.25
Physical Therapist	3.88	8.81	-15.55	2.86	0.00
X-ray Technician	5.11	.73	-17.52	6.94	4.74

Table 23

Average Kuder Classification Percentages Subtracted from Average Chi-Square Classification Percentages on Set II

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	77.40	-40.80	-37.80	- 3.20	4.40
Social Case Worker	30.75	-18.59	-17.70	0.00	5.53
Optometrist	21.92	- .74	-33.74	- 5.67	18.23
Forester	8.05	- 1.72	- 5.18	-38.22	37.06
Auto Mechanic	.67	- 1.68	- 3.02	- 5.70	9.73

IV

DISCUSSION

Results reported in Chapter III indicated the appropriateness of the chi-square technique as opposed to the Kuder technique for discriminating among males according to occupation. The comparisons of the two weighting techniques contained a bias which favored the Kuder technique. These biases resulted because the data scored with Kuder weights were the data Kuder used for deriving his weights. The weights contained all the idiosyncracies of the data to which they were applied and, therefore, were able to classify a greater percentage of males correctly for an occupation than would have been possible on an independent sample.

On the other hand, when the chi-square weights were used to identify males, the sample was independent of the one on which the weights were based. The independence of samples made the results generalizable, but perhaps more parallel percentages for comparison with the Kuder percentages would have been the ones where the weights were applied to the same half from which they were derived.

Scores were computed by using Kuder weights, but Kuder's method of interpretation could not be used because the study involved more than two groups, and Kuder considered only two at a time. Raw scores could not be compared over occupations because the distribution of scores for one occupation might be quite different from the distribution of scores for another. Evidence for these differences was shown in part by the fact that the total numbers of points possible were not equal over professions. Because of the nature of the scoring, some occupational keys provided more ways to get points than others. For example, on a particular item for one occupation a score of two might be obtained in either of the two ways and for another occupation by only one way. For either key the total possible on that item is two. A summary of the number of points possible per key and the ways of getting such a score can be found in Appendix U. Probably a better way to handle the differential effect of various keys on comparisons of raw scores would have been to transform them to standard scores. Such transformations would take care of differences in levels of various occupational distributions and also differences in their variances. Such a technique involves a tremendous amount of computation and, even when considered in the light of high speed digital computers, would still have been a large task. Instead, raw scores were transformed into percentages. Such a procedure took into account the differences in levels of occupational distributions but did not touch upon the problem of different variances. By studying Appendix V, an estimate of the importance of variability for the reported sets of occupations is possible.

Another point is that the criterion of either right or wrong is probably not the most meaningful one for interest in the real world sense. A person may have interests which are just as complementary to one occupation as they are to another. With this in mind the criterion of highest score becomes a poor one. Possibly a minimum score should be set, below which a person would not be advised to enter the occupation. This, however, is a negative approach, and a better approach might be to consider all high scores as possible occupations. After all, the important thing is to eliminate the occupations which are completely unsuitable. Picking the one best occupation would seem of lesser importance, as well as being a difficult task.

If at the start of the study complete freedom of choice had been possible in selecting occupations, a factor analytic approach would have insured the nature of the two sets of data. However, the choices made by intuitive reasoning were supported in part by two factor analytic studies (Schutz & Baker, 1962; King & Norrell, 1964). These studies found that clinical psychologist and forester were in separate factors. Further support was given by Schutz and Baker (1962), who placed pediatrician, veterinarian, and x-ray technician in a common factor. Unfortunately, the other occupations considered in the study were not a part of either factor analysis.

A point relevant to the use of results is that inter-occupational correlation matrices might be a good check on factor-analysis results.

For example, if a factor analysis of occupations yielded the results that two occupations had high loading on a common factor, their inter-occupational correlation should also be high and positive. If one occupation received a high factor loading on one factor, and another occupation received a high factor loading on an orthogonal factor, their inter-occupational correlation should be zero.

Items could be deleted from the instrument by using each item's chi square as the criterion. For any fixed set of occupations, chi-square statistics could be computed for each item. Items which do not have a chi-square statistic significant at or above a specified level probably do not contribute much to the discriminating power of the total scores based on the chi-square weights. Appendix W indicates the number of items which had chi squares significant at the .05 and .01 levels. The items that were found to be not significant for one of the halves were not necessarily the non-significant items for the other half; however, there was considerable overlap (See Appendix X). A more appropriate approach would have been to use both halves in establishing the chi squares for keeping items. After choosing the discriminating items, the procedure described by the study could be used. However, Cowdery (1925) substantiates the merit of keeping all items when weighted to their degree of significance. His reasons are that mere numbers of items might add to the reliability as well as to the validity of the final score of an instrument by considering all available information. In rebuttal, some situations require testing time be kept to an absolute minimum, and an instrument which still provides an adequate level of discrimination would be quite valuable.

A possible use of the chi-square technique might be made by such a school as engineering. The admission officers of an engineering school wish to determine the particular area for which an entering freshman is best suited in order to avoid loss of time, both for the individual and the school. Interest might be one of the important criteria for making this choice. Data could be gathered for each of the possible classifications of engineers, and chi-square weights could be established on the data. One definite advantage of the chi-square technique is that the entire procedure has been programmed. Hence, only a few minutes would be necessary for this important decision. A program is also available for scoring large numbers of people in a very few minutes. This decrease in both time and labor may in some instances be an important factor.

V
SUMMARY AND CONCLUSIONS

Two main hypotheses were tested: (1) A set of weights, established by the chi-square technique, discriminates among more than two similar occupations better than the weights used by the Kuder; and (2) a set of weights, established by the chi-square technique, discriminates among more than two dissimilar occupations as well as, or better than, the weights used by the Kuder. The results of the investigation supported hypothesis one but seemed to repudiate hypothesis two.

The methods for investigating the two hypotheses were identical; however, the data were different. Empirical weights were derived for each of two sets of occupations by the use of a chi-square technique. These weights were applied to cross-validation groups, and the percentages of correctly classified males were determined for each of the occupations. The percentages of all occupations in both Set I and Set II were better than chance. The double cross-validation studies of the derived chi-square weights indicated that the weights were general in nature and not just indicative of the sample upon which the present study was based.

The same data which were scored by chi-square weights were also scored by Kuder weights, and the percentages of correctly classified males were again determined. The percentages based on the Kuder weights were subtracted from the percentages based on the chi-square weights, which allowed for a comparison of the two scoring techniques. As explained in Chapter III of the present study, the percentages for Set II were incorrect because an improper scoring key was provided to use for clinical psychologist. These comparisons indicated that the chi-square weights were superior to the Kuder weights for discriminating among homogeneous occupations. The results of comparisons for heterogeneous occupations were discounted because an improper scoring key was used to determine the percentages which were based on Kuder weights.

The present study supported hypothesis one, but evidence was inconclusive for hypothesis two. The value of self-reported interest as a variable for classifying males, according to occupation pursued, was established. With the exception of clinical psychologist, both techniques were capable of picking a percentage significantly better than chance for every occupation incorporated into the investigation.

APPENDIX A

Percentages of Males in a Particular Occupation in Set I Half A
Classified into Each of the Five Occupations in Set I
Using the Chi-Square Weights Derived on Set I Half A

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	61.08	13.30	10.34	9.36	5.91
Pediatrician	.92	74.78	16.06	3.67	4.59
Veterinarian	1.50	4.50	90.50	2.00	1.50
Physical Therapist	6.22	11.40	13.47	60.62	8.29
X-ray Technician	4.38	10.22	13.14	8.03	64.23

APPENDIX B

Percentages of Males in a Particular Occupation in Set I Half B
Classified into Each of the Five Occupations in Set I
Using the Chi-Square Weights Derived on Set I Half B

	Test Indicated Occupation				
Actual Occupation	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	79.80	7.88	6.90	3.45	1.97
Pediatrician	3.67	71.56	16.06	6.88	1.83
Veterinarian	4.50	2.50	89.00	0.00	0.50
Physicial Therapist	10.36	12.43	11.92	56.00	8.29
X-ray Technician	10.22	9.49	16.06	7.30	56.93

APPENDIX C

Percentages of Males in a Particular Occupation in Set II Half A
Classified into Each of the Five Occupations in Set II
Using the Chi-Square Weights Derived on Set II Half A

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	90.80	6.00	.40	.40	2.40
Social Case Worker	18.58	69.03	2.21	2.65	7.52
Optometrist	21.18	4.43	46.80	3.94	23.65
Forester	5.75	1.15	1.72	55.17	36.21
Auto Mechanic	0.00	.67	.67	1.34	97.32

APPENDIX D

Percentages of Males in a Particular Occupation in Set II Half B
Classified into Each of the Five Occupations in Set II
Using the Chi-Square Weights Derived on Set II Half B

	Test Indicated Occupation				
Actual Occupation	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	93.20	1.60	.40	.80	4.00
Social Case Worker	27.43	62.39	1.33	1.60	6.40
Optometrist	15.27	4.43	55.67	0.00	24.63
Forester	8.62	1.15	.57	47.70	41.95
Auto Mechanic	.67	0.00	.67	1.34	97.32

APPENDIX E

```
      PROGRAM SAM
      TYPE INTEGER DATA4, DATA5
      TYPE INTEGER DATA1, DATA2, DATA3
      DIMENSION DATA1(80), DATA2(80), DATA3(80)
      DIMENSION DATA4(80), DATA5(5)
      DO 100 I=1,406
      READ 1, (DATA1(J),J=1,80)
      READ 1, (DATA2(J), J=1,80)
    1 FORMAT (80A1)
      DO 8 J=1,5
      DATA3(J)=0
    8 DATA4(J)=0
      DATA3(6)=4
      DATA4(6)=3
      DO 9 J=1,52
      K=J+6
    9 DATA3(K)=DATA1(J)
      DO 10 J=1,12
      K=J+58
   10 DATA3(K)=DATA2(J)
      DO 11 J=13,25
      K=J-6
   11 DATA4(K)=DATA2(J)
      DO 16 J=27,49
      K=J-7
   16 DATA4(K)=DATA2(J)
      DO 17 J=43,54
   17 DATA4(J)=0
      DATA5(1)=9
      DATA5(2)=9
      DO 13 J=3,5
      K=J+52
   13 DATA5(J) = DATA1(K)
      PRINT 14, (DATA3(J),J=1,70) , (DATA5(K),K=1,5)
      PRINT 15, (DATA4(J),J=1,54) , (DATA5(K),K=1,5)
      PUNCH 14, (DATA3(J),J=1,70) , (DATA5(K),K=1,5)
      PUNCH 15, (DATA4(J),J=1,54) , (DATA5(K),K=1,5)
   14 FORMAT(6I1,64A1,5X,2I1,3A1)
   15 FORMAT(6I1,36A1,12I1,21X,2I1,3A1)
  100 CONTINUE
      END
      END
```

APPENDIX F

```
      PROGRAM SAM
      TYPE INTEGER DATA1, DATA2, DATA3, DATA4
      TYPE INTEGER DATA5, DATA6, DATA7
      DIMENSION DATA1(80), DATA2(80), DATA3(80)
      DIMENSION DATA4(80), DATA5(80), DATA6(80), DATA7(5)
```

```
      DO 100 I=1,650
      READ 1, (DATA1(J),J=1,80)
      READ 1, (DATA2(J), J=1,80)
      READ1, (DATA3(J),J=1,80)
      READ1, (DATA4(J),J=1,80)
    1 FORMAT (80A1)
      DO 8 J=1,5
      DATA5(J)=0
    8 DATA6(J)=0
      DATA5(6)=4
      DATA6(6)=3
      DO 9 J=1,26
      K=J+6
    9 DATA5(K)=DATA1(J)
      DO 10 J=1,26
      K=J+32
   10 DATA5(K)=DATA2(J)
      DO 11 J=1,12
      K=J+58
   11 DATA5(K)=DATA3(J)
      DO 12 J=13,25
      K=J-6
   12 DATA6(K)=DATA3(J)
      DO 13 J=1,23
      K=J+19
   13 DATA6(K)=DATA4(J)
      DO 14 J=43,54
   14 DATA6(J)=0
      DO 15 J=1,2
      K=J+61
   15 DATA7(J)=DATA1(K)
      DO 18 J=3,5
      K=J+52
   18 DATA7(J)=DATA1(K)
      PRINT 16, (DATA5(J),J=1,70) ,J=1,70) , (DATA7(K),K=1,5)
      PRINT 17, (DATA6(J),J=1,54) , (DATA7(K),K=1,5)
      PUNCH 16, (DATA5(J),J=1,70) , (DATA7(K),K=1,5)
      PUNCH 17, (DATA6(J),J=1,54) , (DATA7(K),K=1,5)
   16 FORMAT (6I1,64A1,5X,5A1)
   17 FORMAT(6I1,36A1,12I1,21X,5A1)
  100 CONTINUE
      END
      END
```

APPENDIX G

```
      PROGRAM SAM
      TYPE INTEGER DATA, DATA1, DATA2
      TYPE INTEGER DATA3, DATA4, MDATA
      DIMENSION DATA(80), DATA1(80), DATA2(80), MDATA(80)
      DIMENSION DATA3(3), DATA4(3)
      DO 100 I=1,1140
      READ 1, (DATA(J),J=1,80)
      READ 1, (MDATA(J),J=1,80)
    1 FORMAT (80A1)
      DO 8 J=1,5
```

```
      DATA1(J)=0
    8 DATA2(J)=0
      DATA1(6)=4
      DATA2(6)=3
      DO 9 J=31,80
      K=J-24
    9 DATA1(K)=DATA(J)
      DO 10 J=31,44
      K=J+26
   10 DATA1(K)=MDATA(J)
      DO 11 J=45,80
      K=J-38
   11 DATA2(K)=MDATA(J)
      DO 12 J=43,54
   12 DATA2(J)=0
      DO 14 J=1,3
      K=J+23
      DATA3(J)=DATA(K)
      L=J+1
   14 DATA4(J)=DATA(L)
      PRINT 13, (DATA1(J),J=1,70), (DATA3(K),K=1,3)
      PRINT 15, (DATA2(J),J=1,54), (DATA4(K),K=1,3)
      PUNCH 13, (DATA1(J),J=1,70), (DATA3(K),K=1,3 )
      PUNCH 15, (DATA2(J),J=1,54), (DATA4(K),K=1,3)
   13 FORMAT (6I1,64A1,7X,3A1)
   15 FORMAT (6I1,36A1,12I1,23X,3A1)
  100 CONTINUE
      END
      END
```

APPENDIX H

```
      PROGRAM SAM
      TYPE INTEGER DATA1, DATA2, DATA3
      TYPE INTEGER DATA4, DATA5
      DIMENSION DATA1(80), DATA2(80), DATA3(80), DATA4(80)
      DIMENSION DATA5(5)
      DO 100 I=1,1033
      READ 1, (DATA1(J),J=1,74)
      READ 1, (DATA2(J),J=1,74)
    1 FORMAT (6X,74A1)
      DO 31 J=1,5
      DATA3(J)=0
   31 DATA4(J)=0
      DATA3(6)=4
      DATA4(6)=3
      DO 2 J=1,64
      K=J+6
      IF (DATA1(J)-1H+) 3,4,3
    3 IF (DATA1(J)-1H-) 5,6,5
    5 IF(DATA1(J)-1H0) 7,8,7
    7 IF (DATA1(J)-1H1) 9,10,9
    9 IF (DATA1(J)-1H2) 11,12,11
   11 IF (DATA1(J)-1H3) 99,77,99
   77 DATA3(K)=6
      GO TO 2
```

```
12 DATA3(K)=5
   GO TO 2
10 DATA3(K)=4
   GO TO 2
 8 DATA3(K)=3
   GO TO 2
 6 DATA3(K)=2
   GO TO 2
 4 DATA3(K)=1
   GO TO 2
99 IF (DATA1(J)-1H4) 14, 15, 14
14 IF (DATA1(J)-1H5) 16, 17, 16
16 IF (DATA1(J)-1H6) 18, 19, 18
18 IF (DATA1(J)-1H7) 20, 21, 20
20 IF (DATA1(J)-1H8) 22, 23, 22
22 DATA3(K)=6
   GO TO 2
23 DATA3(K)=5
   GO TO 2
21 DATA3(K)=4
   GO TO 2
19 DATA3(K)=3
   GO TO 2
17 DATA3(K)=2
   GO TO 2
15 DATA3(K)=1
 2 CONTINUE
   DO 13 J=1, 36
   K=J+6
   IF (DATA2(J)-1H+) 33, 34, 33
33 IF (DATA2(J)-1H-) 35, 36, 35
35 IF (DATA2(J)-1H0) 37, 38, 37
37 IF (DATA2(J)-1H1) 39, 40, 39
39 IF (DATA2(J)-1H2) 41, 42, 41
41 IF (DATA2(J)-1H3) 200, 78, 200
78 DATA4(K)=6
   GO TO 13
42 DATA4(K)=5
   GO TO 13
40 DATA4(K)=4
   GO TO 13
38 DATA4(K)=3
   GO TO 13
36 DATA4(K)=2
   GO TO 13
34 DATA4(K)=1
   GO TO 13
200 IF (DATA2(J)-1H4) 44, 45, 44
44 IF (DATA2(J)-1H5) 46, 47, 46
46 IF (DATA2(J)-1H6) 48, 49, 48
48 IF (DATA2(J)-1H7) 50, 51, 50
50 IF (DATA2(J)-1H8) 52, 53, 52
52 DATA4(K)=6
   GO TO 13
53 DATA4(K)=5
   GO TO 13
51 DATA4(K)=4
   GO TO 13
```

```
 49 DATA4(K)=3
    GO TO 13
 47 DATA4(K)=2
    GO TO 13
 45 DATA4(K)=1
 13 CONTINUE
    DO 30 J=43, 54
 30 DATA4(J)=0
    DO 24 J=1, 5
    K=J+69
 24 DATA5(J)=DATA1(K)
    PRINT 25,  (DATA3(J), J=1, 70) , (DATA5(K), K=1, 5)
    PRINT 26,  (DATA4(J), J=1, 54) , (DATA5(K), K=1, 5)
    PUNCH 25,  (DATA3(J), J=1, 70) , (DATA5(K), K=1, 5)
    PUNCH 26,  (DATA4(J), J=1, 54) , (DATA5(K), K=1, 5)
 25 FORMAT (70I1, 5X, 5A1)
 26 FORMAT (54I1, 21X, 5A1)
100 CONTINUE
    END
    END
```

APPENDIX I

```
    PROGRAM TAPE
    DIMENSION ID(80), IB(80)
    K=0
 15 READ INPUT TAPE 33, 4, (ID(I), I=1, 80)
    IF (ID(6)-1H4) 70, 12, 70
 70 IF (ID(6)-1HT) 69, 12, 69
 12 READ INPUT TAPE 33, 4, (IB(I), I=1, 80)
    IF (IB(6)-1H3) 71, 14, 71
 71 IF(IB(6)-1HT) 69, 14, 69
 14 K=K+1
    IF (K-5) 19, 19, 20
 19 PRINT 11,  (ID(I), I=1, 80)
    PRINT 11,  (IB(I), I=1, 80)
 20 IF (K-2007) 25, 16, 25
 16 PRINT 11,  (ID(I), I=1, 80)
    PRINT 11,  (IB(I), I=1, 80)
    GO TO 15
 25 IF (K-3908) 15, 26, 26
 26 PRINT 11,  (ID(I), I=1, 80)
    PRINT 11,  (IB(I), I=1, 80)
    GO TO 30
 69 PRINT 18, K
    IF (K-3909) 15, 30, 30
 18 FORMAT (12H  ERROR  K= , I5)
 11 FORMAT (2X, 80A1)
  4 FORMAT (80A1)
 30 CONTINUE
    END
    END
```

APPENDIX J

```
      PROGRAM RANDOM
      DIMENSION B(500,20), A(20), NA(10), C(20,20)
      READ 2, (NA(I),I=1,10)
    2 FORMAT (10I3)
      READ 3, ((C(L, M), M=1, 10), L=1, 20)
      I=1
      L=1
   11 WRITE OUTPUT TAPE 43, 3, (C(L, M), M=1, 10)
      PRINT 70, (C(L, M), M=1, 10)
      L=L+1
      N=NA(I)
      DO 5 J=1, N
    5 B(J, 1)=-5.0
      KOUNT=0
    4 READ INPUT TAPE 33, 3, (A(J), J=1, 20)
   73 K=IRAN(N)
      IF(B(K, 1)) 6, 73, 73
    6 DO 9 J=1, 20
    9 B(K, J)=A(J)
      KOUNT=KOUNT+1
      IF(KOUNT-N) 4, 20, 20
   20 NR=N/2
      NB=NR+1
      WRITE OUTPUT TAPE 43, 3, ((B(K, J), J=1, 20), K=1, NR)
      WRITE OUTPUT TAPE 43, 3, (C(L, M), M=1, 10)
      PRINT 70, (C(L, M), M=1, 10)
      L=L+1
      WRITE OUTPUT TAPE 43, 3, ((B(K, J), J=1, 20), K=NB, N)
      I=I+1
      IF (I-10) 11, 11, 25
   25 CONTINUE
    3 FORMAT (10A8)
   70 FORMAT (2X, 10A8)
      END
      END
```

APPENDIX K

```
        IDENT      IRAN
        ENTRY      IRAN
IRAN    SLJ        **
        SIL     1  EXIT-1
        LIU     1  IRAN
        SIU     1  FETCH
        INI     1  1
        SIU     1  EXIT
FETCH   LIU     1  **
        LDA        MAGIC
        MUF        RAND
        ENA
        LRS        1
        STQ        RAND
        LDA     1
        MUF        RAND
        INA        1
```

```
          ENI         1 **
EXIT      SLJ           **
MAGIC     DEC           1220703125
RAND      OCT           77777777
          END
```

APPENDIX L

```
      PROGRAM WTTWOB
      DIMENSION ANR(6,7,100), IDATA(100), NA(5)
      DO 60 K=1,100
      DO 60 J=1,7
      DO 60 I=1,6
   60 ANR(I,J,K)=0.0
C     PRESET ARRAY
C     I=PROFESSION, J=RESPONSE, K=ITEM
      DO 352 I=1,3814
  352 READ INPUT TAPE 33,30,A
      I=0
      READ 31, (NA(M),M=1,5)
   31 FORMAT (5I3)
   99 IF (I-5) 70,71,71
   70 I=I+1
      NR=NA(I)
      NM =NR*2+2
      DO 37 M=1,NM
   37 READ INPUT TAPE 33,30,A
   30 FORMAT (A8)
      DO 32 M=1,NR
      READ INPUT TAPE 33,1, (IDATA(K),K=1,100)
    1 FORMAT (6X,64I1,10X/6X,36I1)
      DO 2 K=1,100
      J=IDATA(K)
    2 ANR (I,J,K)=ANR(I,J,K)+1.0
   32 CONTINUE
      GO TO 99
C     FOUND OBSERVED VALUES
   71 DO 5 K=1,100
      DO 5 J=1,6
      DO 5 I=1,5
    5 ANR(6,J,K)=ANR(6,J,K)+ANR(I,J,K)
      PRINT 339, (IDATA(K),K=1,100)
  339 FORMAT (1X,100I1)
      DO 6 K=1,100
      DO 6 I=1,5
      DO 6 J=1,6
    6 ANR(I,7,K)=ANR(I,7,K)+ANR(I,J,K)
C     ANR=MATRIX OF OBSERVED VALUES, ROW TOTALS, COLUMN TOTALS
      DO 7 K=1,100
      DO 7 I=1,5
    7 ANR(6,7,K)=ANR(6,7,K)+ANR(I,7,K)
      DO 55 K=1,100
      PRINT 56,K
   56 FORMAT(6H ITEM ,I3)
      DO 55 I=1,6
      PRINT 52, (ANR(I,J,K),J=1,7)
```

```
   52 FORMAT(1H ,7F8.4,2X)
   55 CONTINUE
C     COMPUTED TOTAL NUMBER RESPONDING
      DO 10 K=1,100
      DO 10 J=1,6
      DO 10 I=1,5
      SIGN=1.0
      EXP=(ANR(6,J,K)*ANR(I,7,K))/ANR(6,7,K)
      IF(ANR(I,J,K)-EXP) 8,9,9
    8 SIGN=-1.0
    9 ANR(I,J,K)=(ANR(I,J,K)-EXP)**2/EXP
   10 ANR(I,J,K)=ANR(I,J,K)*SIGN
C     COMPUTED CHI SQUARE WEIGHTS WITH SIGN
C     PUT THEM BACK IN ANR
      DO 78 K=1,100
      DO 78 J=1,7
   78 ANR (6,J,K)=0.0
      DO 79 K=1,100
      DO 79 I=1,5
   79 ANR(I,7,K)=0.0
      DO 13 K=1,100
      DO 13 J=1,6
      DO 13 I=1,5
   13 ANR(6,J,K)=ANR(6,J,K)+ABSF(ANR(I,J,K))
C     COLUMN SUM OF CHI SQUARES
      DO 14 K=1,100
      DO 14 I=1,5
      DO 14 J=1,6
   14 ANR(I,7,K)=ANR(I,7,K)+ABSF(ANR(I,J,K))
C     ROW SUM OF CHI SQUARES
      DO 15 K=1,100
      DO 15 I=1,5
   15 ANR(6,7,K)=ANR(6,7,K)+ANR(I,7,K)
C     100 CHI SQUARES
      DO 17 K=1,100
      PRINT 201 ,K
  201 FORMAT(1H0,6H ITEM ,I3)
      DO 16 I=1,6
      PRINT 53,I,(ANR(I,J,K),J=1,7)
   53 FORMAT(12H PROFESSION ,I1,2X,7F9.4)
      PUNCH 301,(ANR(I,J,K),J=1,7)
  301 FORMAT(7F9.4)
   16 CONTINUE
   17 CONTINUE
      END
      END
```

APPENDIX M

```
      PROGRAM SCORE
C   TS FOR SET TWO B ON A THEN B ON B
      DIMENSION ANRA(6, 7, 100), ANRB(6, 7, 100), IDATA(100)
      DIMENSION TSA(5), TSB(5), NA(5)
      DO 2 K=1, 100
      DO 2 I=1, 6
    2 READ 72,  (ANRA(I, J, K), J=1, 7)
      DO 3 K=1, 100
      DO 3 I=1, 6
    3 READ 72,  (ANRB(I, J, K), J=1, 7)
   72 FORMAT (7F9. 4)
C  WEIGHTS ARE IN
      DO 352 I=1, 3814
  352 READ INPUT TAPE 33, 30, A
      PRINT 8
    8 FORMAT(12H PROFESSION  , 3X, 3HONE, 6X, 3HTWO, 5X, 5HTHREE, 4X, 4HFOUR,
     15X, 4HFIVE, 6X, 3HONE, 6X, 3HTWO, 5X, 5HTHREE, 4X, 4HFOUR, 5X, 4HFIVE)
      L=0
      READ 31, (NA(M), M=1, 5)
   31 FORMAT (5I3)
   99 IF (L-5)  70, 71, 71
   70 L=L+1
      NR=NA(L)
      NM=NR*2+2
      PRINT 301, L
  301 FORMAT (/13H  PROFESSION  , I1)
      KOUNT=0
      DO 37 M=1, NM
   37 READ INPUT TAPE 33, 30, A
   30 FORMAT (A8)
      DO 32  M=1, NR
      DO 11 J=1, 5
      TSA(J)=0. 0
   11 TSB(J)=0. 0
      READ INPUT TAPE 33, 1, (IDATA(K), K=1, 100)
    1 FORMAT (6X, 64I1, 10X/6X, 36I1)
      DO 6 I=1, 5
      DO 6 K=1, 100
      J=IDATA(K)
      TSA(I)=TSA(I)+ANRA(I, J, K)
    6 TSB(I)=TSB(I)+ANRB(I, J, K)
      KOUNT = KOUNT+1
      PRINT 7, KOUNT, (TSA(I), I=1, 5), (TSB(I), I=1, 5)
    7 FORMAT (2X, I5, 5X, 10F9. 2)
      PUNCH 9, (TSA(I), I=1, 5), (TSB(I), I=1, 5)
    9 FORMAT (10F8. 2)
   32 CONTINUE
      GO TO 99
   71 CONTINUE
      END
      END
```

APPENDIX N

```
      PROGRAM SCOREK
C  TS FOR ALL INDIVIDUALS ON KUDER WTS
      DIMENSION IWTS(10, 6, 100), IDATA(100), NA(10), A(10)
      DIMENSION ITS(10)
      READ 31,  (NA(M), M=1, 10)
   31 FORMAT (10I3)
      READ 32,  (((IWTS(I, J, K), J=1, 6), K=1, 100), I=1, 10)
   32 FORMAT (60I1)
C    WEIGHTS ARE IN
      L=0
   99 IF (L-10) 70, 71, 71
   70 READ INPUT TAPE 33, 30, (A(N), N=1, 10)
   30 FORMAT (10A8)
      PRINT 69,  (A(N), N=1, 10)
   69 FORMAT (2X,  10A8)
      L=L+1
      KOUNT=0
      NR=NA(L)
      DO 32  M=1, NR
      DO 11  J=1, 10
   11 ITS(J)=0
      READ INPUT TAPE 33, 1, (IDATA(K), K=1, 100)
    1 FORMAT (6X, 64I1, 10X/6X, 36I1)
      DO 6 I=1, 10
      DO 6 K=1, 100
      J=IDATA(K)
    6 ITS(I)=ITS(I)+IWTS(I, J, K)
      KOUNT=KOUNT+1
      PRINT 7,  KOUNT,  (ITS(I), I=1, 10)
    7 FORMAT (2X, I4, 10I6)
      PUNCH 9,  (ITS(I), I=1, 10)
    9 FORMAT (10I8)
   32 CONTINUE
      READ INPUT TAPE 33, 30, (A(N), N=1, 10)
      PRINT 69,  (A(N), N=1, 10)
      KOUNT=0
      DO 34  M=1, NR
      DO 12  J=1, 10
   12 ITS(J)=0
      READ INPUT TAPE 33, 1, (IDATA(K), K=1, 100)
      DO 18 I=1, 10
      DO 18 K=1, 100
      J=IDATA(K)
   18 ITS(I)=ITS(I)+IWTS(I, J, K)
      KOUNT=KOUNT+1
      PRINT 7,  KOUNT,  (ITS(I), I=1, 10)
      PUNCH 9,  (ITS(I), I=1, 10)
   34 CONTINUE
      GO TO 99
   71 CONTINUE
      END
      END
```

APPENDIX O

```
      PROGRAM COMPAR
      DIMENSION IA(10), IB(10), IRS(10)
      READ 10, (IA(N), N=1, 10)
   10 FORMAT (10I2)
      DO 1 N=1, 1902
      READ 4, (IB(I), I=1, 10)
    4 FORMAT (10I8)
      DO 2 I=1, 5
    2 IRS(I)=0
      DO 3 I=1, 5
    3 IRS(I)=((IB(I)*100)/IA(I))
      I=1
      J=I+1
   30 IF (IRS(I)-IRS(J)) 60, 70, 70
   60 I=J
      IF (J-5) 61, 71, 100
   61 J=J+1
      GO TO 30
   70 IF(J-5) 62, 71, 100
   62 J=J+1
      GO TO 30
   71 NO=I
      PRINT 7, N, (IRS(I), I=1, 5), NO
    7 FORMAT (2X, I4, 5I6, 5X, I2)
    1 CONTINUE
      DO 15 N=1, 2004
      READ 4, (IB(I), I=1, 10)
      DO 40 I=1, 5
   40 IRS(I)=0
      DO 41 I=6, 10
   41 IRS(I)=((IB(I)*100)/IA(I))
      I=6
      J=I+1
   42 IF(IRS(I)-IRS(J)) 80, 90, 90
   80 I=J
      IF(J-10) 81, 91, 100
   81 J=J+1
      GO TO 42
   90 IF(J-10) 82, 91, 100
   82 J=J+1
      GO TO 42
   91 NO=I-5
      PRINT 7, N, (IRS(I), I=6, 10), NO
   15 CONTINUE
  100 CONTINUE
      END
      END
```

APPENDIX P

Correlations of Total Scores for Males in Set I Half A
Scored on Weights Derived on Set I Half B with
Total Scores for Males in Set I Half A
Scored on Weights Derived on Set I Half B

A on B	A on B				
	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	1.0000	-.0408	-.5348	-.1823	-.2053
Pediatrician		1.0000	-.3421	-.1172	-.5486
Veterinarian			1.0000	-.4273	-.0489
Physicial Therapist				1.0000	.2862
X-ray Technician					1.0000

APPENDIX Q

Correlations of Total Scores for Males in Set I Half B
Scored on Weights Derived on Set I Half A with
Total Scores for Males in Set I Half B
Scored on Weights Derived on Set I Half A

B on A	B on A				
	Optometrist	Pediatrician	Veterinarian	Physical Therapist	X-ray Technician
Optometrist	1.0000	-.1337	-.5179	.0600	-.2298
Pediatrician		1.0000	-.3243	-.0361	-.5210
Veterinarian			1.0000	-.5724	-.0001
Physical Therapist				1.0000	.1504
X-ray Technician					1.0000

APPENDIX R

Correlations of Total Scores for Males in Set II Half A
Scored on Weights Derived on Set II Half B with
Total Scores for Males in Set II Half A
Scored on Weights Derived on Set II Half B

A on B	A on B				
	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	1.0000	.6796	-.0700	-.7586	-.9060
Social Case Worker		1.0000	-.1392	-.8429	-.7460
Optometrist			1.0000	-.1680	-.1470
Forester				1.0000	.7421
Auto Mechanic					1.0000

APPENDIX S

Correlations of Total Scores for Males in Set II Half B
Scored on Weights Derived on Set II Half A with
Total Scores for Males on Set II Half B
Scored on Weights Derived on Set II Half A

B on A	B on A				
	Clinical Psychologist	Social Case Worker	Optometrist	Forester	Auto Mechanic
Clinical Psychologist	1.0000	.7078	-.2251	-.7923	-.9108
Social Case Worker		1.0000	-.3359	-.8530	-.7460
Optometrist			1.0000	.0321	.0377
Forester				1.0000	.7484
Auto Mechanic					1.0000

APPENDIX T

A Note on a Conservative Error of Estimate for Use in Testing Significance of Difference
Between Proportions Calculated on the Same Individual

The following is offered as a logical argument and not as a mathematical proof. The variance of the differences of two correlated variables is estimated by

(1) $$S^2_{(x-y)} = S^2_x + S^2_y - 2\ \mathrm{cov}(x, y)$$

If x and y are measures of a characteristic of a common individual, the correlation between x and y for a set of observations must range between 1 and 0. Therefore, equation one has a maximum value when

(2) $$S^2_{(x-y)} = S^2_x + S^2_y,$$

where

$$-2\ \mathrm{cov}(x, y) = 0.$$

The correlation between two variables is defined as

(3) $$r_{xy} = \frac{\mathrm{cov}(x, y)}{S_x S_y}$$

from which it follows logically that when r_{xy} equals zero, cov(x, y) equals zero. $S^2_{(x-y)}$ is maximum when r_{xy} equals zero. The conclusion is that a test requiring $S_{(x-y)}$ as its error term is conservative if $S_{(x-y)}$ is calculated under the assumption of independence of x and y.

APPENDIX U

A Summary of Kuder Occupational Keys

Occupational Key	No. of Questions Used by Key	Maximum Score Possible on Key
Optometrist	52	83
Pediatrician	47	79
Veterinarian	51	80
Physical Therapist	44	67
X-ray Technician	54	82
Clinical Psychologist	51	97
Social Case Worker	51	88
Forester	44	78
Automobile Mechanic	51	80

APPENDIX V

Analysis of the Ways the Items Used Contribute to the Total Score

Actual Occupation	No. of ways to get points	Percentage of items used	
		Points possible 1	Points possible 2
	1	28.85	3.85
Optometrist	2	11.54	51.92
	4		3.85
	1	12.77	0.00
Pediatrician	2	19.15	61.70
	4		6.38
	1	13.73	1.96
Veterinarian	2	29.41	49.02
	4		5.88
	1	27.27	4.55
Physical Therapist	2	20.45	45.45
	4		2.27
	1	22.22	5.56
X-ray Technician	2	25.93	38.89
	4		7.41
	1	9.80	43.14
Clinical Psychologist	2	0.00	43.14
	4		3.92
	1	19.61	3.92
Social Case Worker	2	7.84	62.75
	4		5.88
	1	22.73	25.00
Forester	2	0.00	43.18
	4		9.09
	1	23.53	9.80
Automobile Mechanic	2	19.61	39.22
	4		7.84

APPENDIX W

Number of Items on the Kuder Whose
Contingency Tables Had Significant Chi Squares

		Number of Items	
		.05 level	.01 level
Set I	Half A	85	69
	Half B	83	69
Set II	Half A	96	95
	Half B	99	97

APPENDIX X

List of Items with a Nonsignificant Chi Square

.05 Level*				.01 Level**					
Set I		Set II		Set I				Set II	
A	B	A	B	A		B		A	B
15	5	26	52	14	58	5	52	26	52
25	24	52		15	59	11	54	52	71
26	25	71		21	65	13	58	70	99
35	26	99		24	66	15	59	71	
42	30			25	68	24	66	99	
44	35			26	69	25	68		
46	41			28	70	26	69		
50	42			32	74	28	71		
59	43			35	75	30	74		
68	50			40	79	31	75		
69	52			42	80	32	79		
74	66			44	84	35	80		
80	68			45	88	41	83		
84	69			46	94	42	94		
99	71			50	99	43	99		
	74			56		50			
	79								

*8 of the nonsignificant items were common to each half at the .05 level.

**20 of the nonsignificant items were common to each half at the .01 level.

REFERENCES

Brigham, C. C. A study of error. New York: College Entrance Examination Board, 1932.

Buros, O. K. (Ed.) The fifth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1959.

Cattell, R. B. r_p and other coefficients of pattern similarity. Psychometrika, 1949, 14, 279-298.

Control Data Corporation. FORTRAN 63/reference manual. Minneapolis: 1964.

Cowdery, K. M. An evaluation of the expressed attitudes of members of three professions. Unpub. doctoral dissertation, Stanford Univer., 1925.

Cronbach, L. J. Statistical methods applied to Rorschach scores. Psychol. Bull., 1949, 46, 393-429.

Cronbach, L. J. & Gleser, Goldine C. Similarity between persons and related problems of profile analysis. Bur. Res. & Service, Univ. of Ill., 1952 (mimeo).

DuMas, F. M. A quick method of analyzing the similarity of profiles. J. Clin. Psychol., 1946, 2, 80-83.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1949.

Gaier, E. L. & Lee, Mar[illegible] C. Pattern analysis: the configural approach to predictive measurement. Psychol. Bull., 1953, 50, 140-148.

Kelley, T. L. Statistical method. New York: Macmillan, 1923.

King, P. & Norrell, Gwen. A factorial study of the Kuder Preference Record Occupational, Form D. Educ. Psychol. Measmt, 1964, 24, 57-63.

Kuder, G. F. Kuder Preference Record Occupational, Form D, Test Booklet. Chicago: S.R.A., 1956.

Kuder, G. F. A comparative study of some methods of developing occupational keys. Educ. Psychol. Measmt, 1957, 17, 105-114.

Kuder, G. F. Kuder Preference Record Occupational, Form D Manual. Chicago: S.R.A., 1961.

Kuder, G. F. A rationale for evaluating interests. Educ. Psychol. Measmt, 1963, 23, 3-10.

McNemar, Q. Psychological statistics. New York: John Wiley, 1949.

Mosier, C. I. Problems and designs of cross-validation. Educ. Psychol. Measmt, 1951, 11, 5-11.

Rao, C. R. Utilization of multiple measurements in problems of biological classification. J. roy. statist. Soc., 1948, 10, 159-193.

Schacht, R. M. Unified means, standard deviations and correlation program (1604 FORTRAN symbolic version), unpublished, Co-Op ID, G 2 - WISC STATI, 1964.

Schutz, R. E. & Baker, R. L. A factor analysis of the Kuder Preference Record Occupational, Form D. Educ. Psychol. Measmt, 1962, 22, 97-104.

Stanley, J. C. Measurement in today's schools. (3rd ed.) Englewood Cliffs: Prentice-Hall, 1964.

Tilton, J. W. The measurement of overlapping. J. educ. Psychol., 1937, 28, 656-662.

Zubin, J. The determination of response patterns in personality adjustment inventories. J. educ. Psychol., 1937, 28, 401-413.

Zubin, J. Nomographs for determining the significance of the differences between the frequencies of events in two contrasted series or groups. J. Amer. Statis. Assoc., 1939, 34, 539-548.