DOCUMENT RESUME

24

ED 023 123 CG 003 014

By-Prediger, Dale J.

New Procedures for Scoring Psychological Measurements (Development of Moderated Scoring Keys for Psychological Inventories). Final Report.

Toledo Univ., Ohio.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.

Bureau No-BR-7-E-030

Pub Date Feb 68

Contract-OEC-3-7-070030-2871

Note-48p.

EDRS Price MF-$0.25 HC-$2.50

Descriptors-Answer Keys, Biographies, *College Bound Students, Item Analysis, *Predictive Validity, Research Projects, *Surveys, Testing

The three major project objectives were as follows: (1) development of procedures for determining the optimum number of subgroups (and hence, moderated scoring keys) required for maximizing the predictive effectiveness of an inventory; (2) development of a single scale for reporting the scores obtained from a set of moderated keys; and, (3) determination of the accuracy of moderated scoring key predictions of college attendance as compared with predictions obtained from conventional keying techniques. Basic data consisted of biographical inventory responses and academic aptitude test scores for approximately 20,500 high school boys. Scoring keys were formed for each of 12 ability level subgroups and various combinations of these subgroups. The keying procedures developed as the primary objective of the project work well when applied to actual data. Although a statistically significant difference in favor of the moderated keys was obtained, academic ability was not found to be an effective moderator variable. However, a hit rate of 77% was achieved by biographical data as a predictor of college attendance versus nonattendance. This rate and the equivalent point biserial correlation coefficient of .60 were substantially higher than the corresponding figures for academic aptitude used alone. (AUTHOR)

FINAL REPORT

Project No. 7-E-030
Contract No. OEC 3-7-070030-2871

NEW PROCEDURES FOR SCORING PSYCHOLOGICAL MEASUREMENTS

(Development of Moderated Scoring Keys for Psychological Inventories)

Dale J. Prediger

University of Toledo

Toledo, Ohio

February, 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

CONTENTS

SECTION	PAGE
Summary	1
Problem	2
Related Literature	3
Objectives	6
Procedures	7
Population and Sample	7
Variables	8
Design and Analyses	10
Results	17
Group 1 Analyses	17
Group 2 Analyses	19
Group 3 Analyses	19
Group 4 Analyses	22
Conclusions	30
References	33
Appendix	35

TABLE	PAGE
1. Moderator Variable Score Limits for the 12 Ability Level Groups	12
2. Keys Appropriate to Each of the Twelve Ability Levels	13
3. Key Lengths for Each of the 22 Keys	18
4. Relationship Between Key Length and Key Validity for Each of the 22 Keys	20
5. Determination of Optimum Number of Scoring Keys for Group 3	21
6. Descriptive Data for the Optimum Set of Moderated Scoring Keys for Group 3	23
7. Comparison of Accuracy of 4 Predictors in the Prediction of College Attendance for Group 4	24
8. Joint Hit, Miss Rates for Moderated Keys Versus Other Three Predictors	25

CONTENTS (Cont.)

FIGURE PAGE

1. Phi Trends Across 12 Ability Levels for Item 65, Response 2 . . 27

2. Phi Trends Across 12 Ability Levels for Item 71, Response 2 . . 28

3. Phi Trends Across 12 Ability Levels for Item 96, Response 6 . . 29

Acknowledgments

The author is especially grateful to Darel Eschbach of the University of Toledo Computation Center for his ingenuity and perseverance in trying to get the required data analyses completed locally. Deep gratitude is also owed to Robert Lilley and his staff at the Ohio University Computer Center, where the analyses were finally run.

This investigation utilized the Project TALENT Data Bank, a cooperative effort of the United States Office of Education, the University of Pittsburgh, and the American Institutes for Research. The design and interpretation of the research reported herein, however, are solely the responsibility of the author.

Summary

The primary objective of this study involved exploration and refinement of moderated scoring key techniques for scoring psychological inventories. Work centered in two areas: (1) Development of procedures for determining the optimum number of subgroups (and hence, moderated keys) required for maximizing the predictive effectiveness of an inventory; and, (2) Development of a single scale for reporting the scores obtained from a set of moderated keys. The secondary objective involved determination of the accuracy of moderated scoring key predictions of college attendance as compared with predictions obtained from conventional keying techniques.

Basic data consisted of biographical inventory responses and academic aptitude test scores for approximately 20,500 high school senior boys. An equally weighted combination of two academic aptitude tests served as the moderator variable. Scoring keys were formed for each of twelve ability level subgroups and various combinations of these groups. Decision rules involving comparison of point biserial correlation coefficients for selected keys and key groups were devised in order to identify the optimum number of moderated scoring keys. Criterion group membership probabilities were used to form a common scale on which to report scores. The keying procedures developed as the primary objective of the project appear to work well when applied to actual data. Large samples will generally be required for their use, however.

Academic ability was not found to be an effective moderator variable for the predictor and criterion under study. Although a statistically significant difference in favor of the moderated keys over conventional keys was obtained, the practical implications of this difference were slight.

A hit rate of 77% was achieved by biographical data for predictions of college attendance and nonattendance in a cross-validation sample. This rate and the equivalent point biserial correlation coefficient of .60 were substantially higher than the corresponding figures for academic aptitude used alone. Thus, biographical data would appear to be a powerful predictor of college attendance.

Problem

In the field of measurement, instruments falling in the category called tests are characterized as having questions with one and only one correct answer. Inventories, on the other hand, contain questions which have no one correct answer. At best, the "correct" answer to an inventory item might be thought of as the response most appropriate to the person answering the question. Thus, to the question, "How many brothers and sisters do you have?", one person might respond with "one," another with "six," and another with "none."

A test item is easy to score because there is only one correct answer. However, no such easy scoring procedure is available for the inventory item. Instead, inventory scoring requires the application of specialized techniques for developing scoring keys (e.g., keying techniques). Three general keying techniques are available--logical, homogeneous, and empirical techniques. Logical keying techniques date back to early, unsophisticated attempts at inventory construction and are seldom used by measurement specialists today. Both homogeneous and empirical keying methods have unique merits with the latter generally seen as yielding more valid predictions of specific criteria.

In its simplest form, empirical keying involves contrasting the inventory item responses of two groups (e.g., high school dropouts and persisters), identifying those items to which the groups respond differently, and then scoring those items in such a fashion that members of one group will tend to receive higher total scores than members of the other group. Thus, to the question, "How many brothers and sisters do you have?", responses of "two or three" might be scored +1, indicating a tendency toward persistence; responses of "six or more" might be scored -1; and all other responses may receive no score because they fail to differentiate persisters and dropouts. If a sufficient number of items are found to contain response options that differentiate between persisters and dropouts, the inventory can be used with other individuals for whom predictions of group membership are desired.

Numerous refinements of this simple approach have been available for some time. For instance, Gleser and DuBois (1951) describe a successive approximation method for taking into account both item-criterion correlations and interitem correlations. More recently, Ghiselli (1960) showed that it is possible to differentiate individuals for whom predictions have relatively high versus low accuracy by use of an empirically keyed moderator variable. Finally, Prediger (1966) demonstrated the feasibility of developing separate inventory scoring keys for application to individuals scoring at each of several different levels on a logically related moderator variable. This approach to empirical keying, known as the moderated scoring key approach, appears to represent a new and promising method for improving the validity of many types of inventories.

Surveys of study habits and attitudes, interest inventories, personality inventories, and biographical information blanks are among the general types of inventories representative of those commonly in use in the nation's schools. In addition to the more common uses, these inventories have been applied to the prediction of high school dropout, college attendance, college persistence, and the estimation of delinquency proneness. Yet, for many applications, inventories have been found to contribute a discouragingly small amount of useful information. The development of improved keying techniques would appear to be a basic means of increasing inventory validity.

Within this context, the general purposes of this study were: (a) To investigate procedures for refining the moderated scoring key approach and facilitating its practical application; and (b) to provide information on the generalizability and effectiveness of the approach by comparing the validity of college attendance predictions based on the refined keys with those obtained by conventional procedures.

As a by-product of this investigation, biographical data related to college attendance versus nonattendance was identified for a nationwide sample of high school students. Sufficient information is presented in Appendices A through C of this report for the development of an effective tool for use by school counselors in locating high ability students not likely to go to college.

Related Literature

Research on the application of moderator variables to prediction problems has been reviewed by Dunnette (1963) and Ghiselli (1963). The common approach is to develop, by means of empirical keying techniques, a scale that differentiates between individuals for whom inventory or test predictions have greater and lesser degrees of accuracy. The scale is either based on unused items within the predictor itself or on the items in a separate measure. The scale so developed is termed a moderator variable because it is seen as representing a characteristic that affects (moderates) or at least reflects different degrees of relationship between two variables in much the same way that two variables interact with a third in analysis of variance. These moderator variables usually have been found to add little to the accuracy of predictions when combined with the predictor in a multiple regression equation. Hence, their chief function is to "predict predictability."

Another application of moderator variables, which in addition to facilitating the prediction of predictability, promises aid in understanding the reasons for differential predictability, involves the use of logical rather than empirically derived moderator variables. Thus Goodstein and Heilbrun (1962), in a study of college sophomores, found only one significant relationship between the Edwards Personal Preference Schedule and GPA when vocabulary level was controlled by means of partial correlation. However, when academic aptitude was used as a logical moderator variable by dividing both male and female groups into low, medium, and high ability subgroups of equal size, a number of statistically significant partial correlations appeared, especially for the males. In addition, the scales for which correlations were found to be significant varied greatly with ability level.

In a more recent study, Heilbrun (1965), using college persistence over a one-year period as the criterion, paired persisters and dropouts on the basis of ability and sex. He then divided the male and female groups into subgroups of high, medium, and low ability. Hypothesized differences between persisters and dropouts on the Adjective Check List occurred primarily at the high ability level. Significant ability level interaction was found. Hence, in both of the above studies, academic ability was used as a logical moderator variable that added to the understanding of the relationship between personality characteristics and college success and, at the same time, identified subgroups with differing predictability.

It should be noted that Hakel (1966), in a replication of the Goodstein and Heilbrun study, failed to substantiate the generality of the results. Hakel's additional finding that none of the EPPS-achievement correlations were stable across three heterogeneous subgroups of the total sample may be the result of his small total sample size (N=102). The small size of the ability level subgroups could also have a bearing on the negative results of his replication. In any case, Hakel justifiably emphasizes the need for cross-validation and validity generalization studies in the investigation of moderator variable effects.

Among other recent studies involving moderator variables is one by Berdie and Hood (1966) in which the relationship between a number of measures and the college attendance plans of high school students grouped by sex and area of residence was investigated. High school rank, scholastic aptitude test scores, personality variables, and biographical data were among the predictors used in the study. According to Berdie and Hood, the results suggest that predictions of college plans are more accurate when prediction equations are developed for homogeneous groups of students. In addition, there was evidence of differential predictability across the subgroups. It is interesting to note that the moderator variables in this study were categorical rather than continuous.

The usefulness of a number of other moderators in specific situations has been investigated. For example, Korman (1966, 1967) successfully used self-esteem as a moderator of the relationship between various self perceptions and nature of vocational choice. Clark and Campbell (1965) report that ability serves as a moderator variable for the relationship between interests and grades. Brown and Scott (1966) were unsuccessful in their efforts to improve predictions of grades for various groups of college students through the use of study habits, study attitudes, and personality variables as moderators. On the other hand, Barger and Hall (1965), using ability as a moderator instead of a predictor, showed that parents' marital status, family income, father's education, and family size were differentially related to GPA for college females. No statistically significant relationships were found for males at any of the three ability levels under investigation. Parent's marital status was the only biographical variable found to be related to persistence in college. This relationship appeared for both males and females but only for those in the upper third on ability. Thus, ability but not sex served as a moderator variable in this instance. On the basis of results obtained in the above and other studies, it would appear that Ghiselli and Sanders' observation (1967) that moderator variables are situation-specific is very appropriate.

Hobert and Dunnette (1967) continued the reformulation of the prediction problem begun by Dunnette (1963). Using what was called the quadrant-analysis technique, they divided the bivariate, predictor-criterion frequency distribution at the predictor median. The predictor, in this case, consisted of a composite score on a comprehensive battery of biographical, personality, ability, and managerial judgment measures. The same biographical data and personality variables were then used to differentiate subjects above and below the median on a criterion consisting of a comprehensive measure of managerial effectiveness. Separate moderator scales were developed for the "high" predictor group (above predictor median) and the "low" predictor group. These moderator scales were then used to identify and eliminate overpredicted individuals (lower right quadrant) and underpredicted individuals (upper left quadrant) from a cross-validation sample. Of 220 cases in the sample, 25% were eliminated as unpredictable. The point biserial correlation between the predictor and criterion in the cross-validation sample was .73. The comparable figure for the total cross-validation sample was .65. Thus a substantial increase in validity was accomplished at the expense of some reduction in sample size due to the elimination of the least predictable cases. One cannot help wondering, however, whether comparable results could have been achieved by simply eliminating the middle 25% of the cases on the predictor variable. Hobert and Dunnette think not but no data were provided.

There is another way in which to look at the Hobert and Dunnette data. Suppose the predictor variable is viewed as the moderator variable, in this case dichotomized at the median. Suppose, then, criterion scores above the criterion median are considered to be "successes" and those below as "failures." One then has success-failure groups with "high" predictor (moderator variable in this case) scores and similar groups with "low" moderator variable scores. In the process of developing what were called moderator tests, Hobert and Dunnette were really forming separate scoring keys for the high-low groups. The procedure is very similar to that reported by Prediger (1966). However, in the study by Prediger, application of the results of the procedures was different as will be discussed below.

Earlier work by Prediger (1965) on the relationship of biographical data (biodata) to persistence in college led him to hypothesize that the relationship might actually be moderated by student characteristics such as academic aptitude and prior achievement. Using two-year persistence data for 1469 University of Missouri entering freshmen males and a biographical inventory constructed specifically for the study, Prediger (1966) demonstrated that academic ability is, in fact, an effective moderator variable in the prediction of persistence from biodata. Separate scoring keys were developed for the biographical inventory at each of three ability levels. For two of the ability groups the contribution of biodata to the differentiation of persisters and dropouts in a cross-validation sample was found to be greater than it was for the total group. For the third ability level the results were mixed. Overall differentiation of persisters and dropouts was greater when separate ability level keys (e.g., moderated keys) were constructed than when one overall key was used. Hence, in this study, a technique for increasing the overall predictive validity of an inventory by the development of moderated scoring keys was demonstrated. This approach also allows for the identification of subgroups that have greater or lesser degrees of predictability, which is as far as the previous approaches developed by Ghiselli and others have gone.

Sophisticated techniques for the identification of subgroups have been reviewed by Dunnette (1963) and most recently demonstrated by Schoenfeldt (1966). Rock, Barone, and Linn (1967) describe a computer program that groups individuals on up to five moderator variables in such a fashion that the relationship between ten predictors and the criteria is optimized for the total sample. However, for the purpose of investigating ways of refining moderated scoring key techniques, the less cumbersome procedures for identifying ability level subgroups discussed in the section on design would appear to be adequate.

Before experimentation with moderated scoring keys can take place, a relationship must be present between items in the inventory for which keys are to be developed and the event or characteristic to be predicted. In this study, biodata was used as a predictor of college entrance for a nationwide sample of students in the Project TALENT study (Flanagan, Davis, Dailey, Shaycroft, Orr, Goldberg, and Neyman, 1964). The criterion of college entrance is one that has been frequently shown to be related to various aspects of biodata. Such relationships have already been reported for a few Project TALENT biodata items (Flanagan and Cooley, 1966).

Objectives

The primary objective of this project involved the exploration and refinement of the techniques for developing moderated keys first reported by Prediger (1966). Investigation centered in two areas:

(a) Development of procedures for determining the optimum number of subgroups (and hence, moderated scoring keys) required for maximizing the predictive effectiveness of an inventory.

(b) Development of a single scale for reporting the scores obtained from a set of moderated scoring keys, thus facilitating the practical application of moderated keys to prediction problems.

The development of a single scale is especially important since moderated keys are usually of different length and have different means and standard deviations. Conversion to standard scores represents no solution because the predictive implications of comparable standard scores on different moderated scoring key scales are likely to differ greatly.

The secondary objective of the project involved investigation of the generalizability of moderated scoring key techniques. In order to achieve this objective and at the same time evaluate the effectiveness of procedures developed through pursuit of the primary objective, the following general and subsidiary hypotheses concerning the application of moderated scoring keys to the prediction of college attendance were made:

(a) General hypothesis--The contribution of biodata to the prediction of college attendance is greater upon cross-validation, for scoring keys developed within ability level subgroups (e.g., moderated keys) than for a key developed on the total group (e.g., the general key).

(b) Subsidiary hypotheses--Moderated scoring key predictions of college attendance-nonattendance for high school students are more accurate than predictions based on:

(S1) the normal empirical keying procedures used in forming the general key.

(S2) the above, plus academic aptitude data in an optimally weighted combination.

(S3) academic aptitude data alone.

In order for the general hypothesis to be supported, it is necessary to confirm only subsidiary hypothesis S1. However, S2 and S3 are directly relevant to the practical application of moderated keys based on biodata since these keys must produce more accurate predictions than could be obtained through other readily available means. In other words, there is no substitute for incremental validity.

Procedures

Population and Sample

All data used in this study were obtained from the Project TALENT Data Bank. The subjects constitute a subgroup of the stratified random sample of the U. S. high school seniors who participated in the data collection phase of Project TALENT in the spring of 1960. Stratification was based on factors such as type of school and geographical area so that the sample would be representative of all high school seniors in the U. S. The nature of the stratification variables and the representativeness of the obtained sample are discussed in a publication entitled The Project TALENT Data Bank (1965).

Of the total senior sample, only males who responded to the Project TALENT follow-up questionnaire mailed in the summer of 1961 and for whom data were available on the Project TALENT Data Bank tape files in July, 1966, were included in this study. (Females were not included because the relationship between their biodata responses and the criterion is likely to differ from that of males, thus necessitating separate analyses.) Approximately 30% of the seniors in the 1960 sample failed to respond to the follow-up questionnaire (Flanagan and Cooley, 1966). Since important differences between respondents and nonrespondents were found to exist, the subjects included in this study cannot be taken as representative of the U. S. population of high school seniors. Screening for missing data, which is described in the section on design and analyses, further reduced the representativeness of the sample.

It should be emphasized, however, that for the purposes of this study, representativeness of the population of U. S. high school seniors is not essential since generalization of this population is not intended. Instead, the relative effectiveness of the various keying techniques within the available sample is compared. The sample used in this study should represent reasonably well a hypothetical population of high school senior males who choose to respond to the type of follow-up procedure used by Project TALENT. While it may be useful for readers to think in these terms, it is not essential to interpretation of results. Of the 21,534 cases supplied by the Project TALENT Data Bank, 20,367 were available for the study after screening was completed.

Variables

A wide variety of data was collected on Project TALENT students during a two-day period in the spring of 1960. Of the approximately 100 separate test and inventory scores which were available, the following were selected for possible use in a composite moderator variable: Vocabulary (172); English Grammar Total (230); Reading Comprehension (250); Abstract Reasoning (290); Arithmetic Computation (410); and Mathematics (320). The number in parentheses after each variable precisely identifies the variable according to the Project TALENT code. The variable titles identify them as typical measures of academic aptitude and achievement. Each is further described in a manual entitled The Project TALENT Data Bank (1965).

Two of the above six variables, Reading Comprehension and Mathematics, were selected for use in the moderator variable. Selection was based on three factors: (a) point biserial correlation of the variables with college entry, otherwise undefined, as revealed in preliminary data developed by Project TALENT (Flanagan, et al., 1964); (b) apparent adequacy of score range, floor, and ceiling of the measures as indicated by Project TALENT normative data (Flanagan, et al., 1964); and (c) the number of scores available for each possible combination of two variables as revealed during preliminary editing of the magnetic tapes containing the sample data. The correlation between Reading Comprehension scores and entry into college between high school graduation and the one-year follow-up was reported as .42 while that for Mathematics was estimated to be .48 on the basis of this variable's similarity to others for which data was reported. These were the highest correlations reported for the six variables. In fact, the best predictor of college entry among the 56 cognitive variables studied was the Mathematics Information Scale (R-106) with a correlation of .52. This scale, although less adequate in terms of range and ceiling, was highly related to Mathematics (320). Finally, the tape edit revealed that loss of cases due to missing data for these two variables would be slight. The edit sample is identical to analysis group 1 which is described in the section on design.

In order to form the moderator variable, Reading Comprehension and Mathematics scores were combined by use of linear standard scores. Means and standard deviation (35.304 and 9.597 for Reading Comprehension and 24.118 and 8.444 for Mathematics) were obtained on the editing run. These were used to obtain standard scores with a mean of 500 and a standard deviation of 100 for each variable. The composite moderator variable was the sum of the two standard scores.

As can be seen, an effort was made to construct an ability moderator that would include commonly used academic aptitude variables and, at the same time, would have a substantial relationship with the criterion to be predicted. A different approach might have involved the use of a moderator variable more or less unrelated to the criterion. The relative utility of the two approaches is unknown. However, with an item pool consisting of biodata, the criterion-relevant ability moderator appeared to be an ideal choice.

The biodata used in the study was obtained from the Student Information Blank (SIB), a questionnaire with items covering topics such as hobbies; club memberships; social activities; health; work experiences; study habits; school experiences; characteristics of family, home and community; parents' age, source of income, activities, and education; plans for college, military service, marriage, and career; etc. Of the 394 items, 148 were eliminated because they did not appear to fit the common definition of biographical data. These items chiefly covered study habits and plans of various kinds.

Since the exact nature of the 246 item biodata pool is not relevant to this study, the full, copyrighted instrument is not reproduced here. A copy can be found on pages 5-6 through 5-35 of the publication entitled The American High School Student (Flanagan, et al., 1964). Item numbers for the items that were used and the identification number assigned to them for purposes of this study can be found in Appendix A. Since 148 of the original items were not used and since some rearrangement of items was necessary for processing, the project item identification number differs in many cases from the SIB number. Including omits, a total of about 1650 response options were available for item analyses. Although item analysis data were obtained for omits, omits were not included on any scoring keys.

Of the 246 SIB items that were used in this study, 226 contained from two to six response options while the rest contained from 10 to 17 options. To facilitate analyses, these latter items were restructured so as to require no more than nine response categories. This involved combining response options that had low frequencies as reported by Project TALENT and/or that seemed to fit together on a logical basis. Because of the unusually large number and diversity of response options for SIB items 206 and 208, each of these two items was split into two parts with each part treated as a separate item. The analyses for each response involved all cases in the item analysis sample, not just those of students marking responses in a given part. A list of all combinations and the response option identification numbers used in the study is presented in Appendix B along with some minor alterations of response options for three other items. As can be seen in Appendix B, for purposes of this study, the response options were indicated by numbers rather than letters (as on the SIB). Thus, response option A was assigned to number 1; B was assigned 2, etc. in all cases except where recoding was required.

The criterion variable, college attendance, was obtained from the responses of students to item two on the 1961 follow-up questionnaire. A copy of this questionnaire is presented in Appendix H of the American High School Student (Flanagan, et al., 1964). Item two is reproduced in Appendix D of this report. Those students indicating that they had entered college as full-time students are included in the college attendance group. All others are in the nonattendance group. Thus, the students' responses to item two served to operationally define the criterion, college attendance. It is recognized that some students indicating college attendance may have dropped out after a short time and that others not indicating attendance may enroll in subsequent years. In addition, part-time students may become full-time students and vice versa. However, the above definition of college attendance proved adequate for purposes of this study.

Design and Analyses

As was already noted, the editing run on data supplied by Project TALENT revealed missing data in some of the student records. Obviously, if a student's response to item two of the follow-up questionnaire was not available, the rest of his data was lost to the study. In addition, scores for the two tests used in the moderator variable had to be present and within the possible raw score range. Finally, SIB responses had to be present for analyses to be worthwhile. Hence, in addition to the other checks, each record was checked, before processing, for the presence of responses for the first ten SIB items. If no responses were present, the record was ignored and hence was lost to the study. Approximately 400 records were lost for this reason. An additional 400 were lost because of lack of the required test scores. Finally, follow-up data for item two was not available for about 360 cases. The records of 20,367 students remained after these checks were made. For purposes of analyses, these records were randomly subdivided into four independent groups.

Before any analyses could take place, a procedure had to be developed for determining the optimum number of subgroups (and hence, moderated scoring keys) required for maximizing the predictive effectiveness of the item pool. (See primary objective number one.) The only feasible approach involved the subdivision of the item analyses sample into ability groups according to score on the moderator variable. Using Project TALENT figures on attend, nonattend percentages at various ability levels, the effect on criterion group size of various numbers of ability level groups was determined in light of the need to allocate the total sample into four subgroups for analyses. To be avoided were extreme attend, nonattend splits and low sample sizes within an ability level group. This, of course, would most likely occur at the extremes of ability.

Twelve ability level subgroups as nearly equal in size as possible were chosen as being appropriate to the purpose of the study and the data available. By combining response frequencies for adjacent groups, item analysis data based on six and three ability level groups could be obtained without need for a separate item analysis. In addition, data for the total sample could easily be obtained by combining data from the three ability level groups, each of which represented four of the original twelve groups. These were important practical considerations in light of the large sample size and number of response options to be analyzed.

By use of the 12 ability level subgroups, scoring keys could be formed for the total group and for 3, 6, and 12 subgroups. Thus investigation of the optimum number of moderated scoring keys required for maximizing the predictive effectiveness of the item pool was possible. Procedures for dividing the sample into twelve subgroups are described in the section on Group 1 analyses. The nature of the keys is also described in that section. Procedures for deciding on the optimum number of moderated keys are described in the section on Group 3 analyses.

Group 1 Analyses. The records of 10,183 students were randomly assigned to this group. A computer program was written to calculate the moderator variable score for each student in the group and to determine the score limits that divided the group into twelve subgroups as nearly equal in size as possible. Tie scores and the fact that 10,183 is not evenly divisible by 12 prevented the groups from being exactly equal. Score limits and frequencies for the twelve groups are shown in Table 1 along with the attend, nonattend split in each of the groups. These score limits were used in all subsequent analyses.

Since moderated scoring keys were to be formed at each of the twelve ability levels and combinations thereof, 21 moderated keys had to be developed. These plus the general key made a total of 22 keys. Four keys were appropriate to each of the 12 ability levels: the general key, the key specific to the ability level, the first order combined key formed with data from two adjacent levels and the second order combined key formed with data from adjacent first order combined keys. The keys appropriate to each of the 12 ability levels are shown in Table 2. As can be seen, key 22, the general key, was formed and scored on all students. Normal empirical keying techniques would have involved only this key.

Group 1 was used to obtain the item analysis data from wh ch all scoring keys were developed. For each response to an SIB item, a two-by-two contingency table was formed using marked-not marked and attend-nonattend dichotomies. Phi coefficients calculated from this data formed

Table 1

Moderator Variable Score Limits for the 12 Ability Level Groups

Ability level	Score limits	Frequency Attend	Frequency Nonattend	Total
1	401 - 718	106	744	850
2	719 - 810	162	681	843
3	811 - 872	228	628	856
4	873 - 928	275	566	841
5	929 - 976	341	511	852
6	977 - 1021	407	435	842
7	1022 - 1064	459	403	862
8	1065 - 1104	518	326	844
9	1105 - 1147	596	256	852
10	1148 - 1193	642	208	850
11	1194 - 1238	671	151	822
12	1239 - 1600	777	92	869
Total		5182	5001	10183

Note.--Limits and frequencies based on data from analysis group 1. Maximum and minimum possible moderator variable scores were 1600 and 400 respectively. No scores reached these limits.

Table 2

Keys Appropriate to Each of the Twelve Ability Levels

Ability level	Moderated scoring keys: Ability level keys	1st order combination	2nd order combination	General key
1	1	13	19	22
2	2	13	19	22
3	3	14	19	22
4	4	14	19	22
5	5	15	20	22
6	6	15	20	22
7	7	16	20	22
8	8	16	20	22
9	9	17	21	22
10	10	17	21	22
11	11	18	21	22
12	12	18	21	22

Note.--Ability level 1 represents the lowest ability group.

the basic item analysis statistics although chi squares were also obtained along with the proportion of the attend and nonattend groups that marked each response option. As previously mentioned, omits were treated as reponse options in the item analyses although this option was not included on any scoring key.

Before keys could actually be formed, a procedure for determining key length had to be developed. In the past, a common item analysis practice has been to key only those responses which show a relationship with the criterion that is statistically significant at some predetermined level, usually the five per cent level or the one per cent level. However, the results of studies by Appel and Kipnis (1954), Kuder (1951), and Clark (1961), suggest that the optimum significance level cut-off point is a complex function of the criterion, the size and nature of the sample, and the test or inventory being used. Most recently, Abrahams (1967) showed in a convincing fashion that optimum key length is specific to the item pool being keyed. Hence, it was necessary to develop keys of varying length for each ability level subgroup and the total group so that it would be possible to determine empirically the optimum key length upon cross-validation.

Of several possible approaches, the one chosen involved the use of five phi values as cutoffs for keys of five different lengths. These phi values were .03, .05, .07, .10, and .15. Thus the longest subkey at a given ability level would be made up of item responses with phi coefficients of .03 or higher. The shortest subkey would use .15 as the cut off. This was done for each of the 22 keys. A total of 110 subkeys were formed, 5 for each of the 22 keys.

Obviously, with the volume of item analyses data involved, some mechanical means of forming keys was required. No attempt to identify logical or illogical response validity data was made. Instead, a computer program was written to form subkeys. It was applied in the same manner to item analysis data for each of 22 keys.

Of the 110 subkeys formed in this way, those appropriate to keys that were actually scored after preliminary analyses are presented in Appendix C. For each key and subkey length, the item and response number is identified along with the direction in which the response is to be scored, i.e., +1 or -1. Thus Appendix C really represents an index to item responses with phi coefficients falling in five different intervals of absolute value: .03-.049, .05-.069, .07-.099, .10-.149, and .15 or greater. This manner of presenting item response validity data rather than a complete listing of item analysis data was chosen because the latter would require more than 600 pages.

Anyone interested in which items had a given level of relationship with the criterion should refer to Appendix C. It is this data, along with the results obtained in determining optimum key length and number of moderated scoring keys, which could be used in constructing an instrument to identify students who are likely or who are not likely to attend college.

Group 2 Analyses. The records of 3387 students were randomly assigned to this group. Each student was placed in one of the 12 ability level groups on the basis of his moderator variable score. The SIB responses of the students were then scored on four keys (see Table 2), each with a total of 5 subkeys of varying length. Thus twenty scores were obtained for each student. All scoring was done by computer and at the same time, data was accumulated that allowed the calculation of point biserial key-criterion correlations for each of the 110 subkeys (22 keys, 5 subkeys per key). The trend in these point biserials across the five subkeys scored for a given key was used to estimate optimum key length for that key. Only this key length was used in subsequent scoring.

Group 3 Analyses. Analyses on this group had three primary purposes: (a) to determine the optimum number of moderated scoring keys appropriate to the data; (b) to develop a common score scale on which key results could be reported; and (c) to develop equations for use in predicting college attendance versus nonattendance for students in Group 4, the cross-validation group. As explained below, it was possible to accomplish steps 2 and 3 at the same time. The records of 3393 students were randomly assigned to Group 3 analyses.

In order to determine the optimum number of moderated scoring keys, some way to compare the effectiveness of different numbers of keys had to be devised. Comparison of the point biserial correlations obtained for the sets of 12, 6, and 3 keys was not satisfactory because the attendance, nonattendance split differed across the sets. For example, the split in group 12, the highest ability group, would not be the same as the split in groups 9 through 12 combined. As discussed by Guilford (1956), the point biserial correlation coefficient is sensitive to the criterion split. As the split becomes more extreme the values reached by the coefficient tend to be depressed, other things being equal. Hence, the point biserial coefficients could not be used in any simple fashion for interkey comparisons.

It was appropriate, however, to score each higher order key combination on the next lower order combination. Thus, the general key could be scored on the ability groups appropriate to keys 19, 20, and 21. If the development of moderated scoring keys is of no particular value, then key 22 should do as well on the groups appropriate to keys 19, 20, and 21 as the keys do themselves. In fact, it should do better because it is based on an item analysis sample three times as large as the second order key combinations. At the same time, the effect of criterion group split is controlled because all point biserial comparisons are made on the same group.

Using this principle, the following rules were devised for determining optimum key length.

1. The general key is scored on the groups appropriate to the second order combination keys, *i.e.*, keys 19, 20, and 21, and will replace any of these three keys unless the difference in point biserial correlation coefficients is at least .01 units in favor of the second order keys.

2. The second order combination keys resulting from step one are scored on the groups appropriate to the first order combination keys, i.e., keys 13 through 18. Each second order key is scored only on the two first order combination groups appropriate to the key. Thus key 19 would be scored in key groups 13 and 14 separately. If the point biserials for a second order key are higher than those of both first order keys, the second order key will be used instead of the first order keys. If the point biserials are lower than those of both first order keys, the first order keys will be retained. If the results are mixed, i.e., second order biserial higher than one of the first order key biserials but lower than the other, a judgment will be made as to which difference is greater. For the sake of economy in scoring time, this decision will generally be made in favor of the second order keys. If the decision is in favor of the second order key, it will replace both first order keys. If not, both will be retained.

3. The first order combination keys resulting from step two are scored on the ability groups appropriate to those keys. Thus key 13 (or its replacement) would be scored separately on ability groups 1 and 2. Decision rules paralleling those in step two are then applied.

Application of these rules would appear to provide a flexible as well as workable procedure for determining the optimum number and combination of moderated scoring keys. It would be possible for the general key to replace all of the moderated keys or for some combination of keys involving the general key, second order keys, first order keys, and ability group keys to be selected. For example, ability groups 1--4 might be scored on the general key, groups 5--8 on key 20, groups 9--10 on key 17, and groups 11 and 12 on their own moderated scoring keys. This would be an example of a case where moderated scoring keys are unimportant for low ability groups but become more and more important as ability increases.

Once the optimum set of keys was determined, a common score scale had to be developed along with equations for predicting attendance versus nonattendance. Both of these problems were solved through use of the classification procedure described by Cooley and Lohnes (1962). This procedure allows one to determine the similarity of an individual to one or more groups on the basis of one or more scores for the individual. In addition to differences in group means, differences in group size and dispersion are taken into account. In this case there were two groups, college attenders and nonattenders, for which a student's similarity was to be estimated. These estimates of similarity were in the form of probabilities ranging between .00 and 1.00.

Prediction equations were developed from group 3 data for each of the keys in the best combination of moderated keys, for the general key, for the ability moderator, and for the general key and ability moderator in optimally weighted combination. The estimates of attendance group similarity obtained from the equations for moderated key scores formed a common score scale to which the moderated keys could be anchored.

Group 4 Analysis. Analyses conducted on this group were for the purpose of obtaining information bearing directly upon the general and subsidiary hypotheses. The records of 3404 students were randomly assigned to these analyses.

Scores for each student were obtained on the moderated scoring key appropriate to his ability level, the general key, and the moderator variable. Prediction equations developed in Group 3 were applied to these scores so that for each student estimates of group similarity based on the three scores mentioned above plus the last two in optimally weighted combination were obtained. Thus, there were four attendance group similarity estimates for each student. These estimates, in the form of probabilities, were based on (a) the student's score on the moderated scoring key appropriate to his ability level; (b) his score on the general key; (c) his score on the general key optimumly combined with his academic aptitude score; and (d) his academic aptitude score alone. Whenever the probability exceeded .50, "attend" was predicted. Otherwise, "nonattend" was predicted.

Records of hits and misses according to what the students had actually done were accumulated for each of the four predictors as the scoring was done. More important, a record of the joint hit-miss data for the moderated keys and each of the other three predictors was maintained. In this way it was possible to determine the number of cases for which the general key and moderated key, for example, performed the same, i.e. both "hit" or both "missed." More important, it was possible to determine the number of cases for which one predictor hit when the other missed. If the moderated keys correctly predicted many of the cases which the general key missed but the reverse did not happen, then the moderated key would clearly be superior. A statistical test described by McNemar (1962) for data of this kind was used to obtain a *z* value for the information of those interested in the traditional test of statistical significance.

The same procedure was followed for each of the subsidiary hypotheses. In addition, phi coefficients expressing the relationship between predicted criterion group membership and actual criterion group membership were obtained for each of the four predictors.

Results

Group 1 Analyses

As explained in the section on design, data for this group were used in the formation of ability group limits (see Table 1), in the item analysis, and in the construction of scoring keys. Scoring keys are presented in Appendix C. The number of items in the five subkeys for each of the 22 keys is presented in Table 3. There is a wide range in subkey length for all keys thus providing ample opportunity for the appearance of trends in the relationship between key length and validity. No conclusions relevant to project objectives can be made from Group 1 analyses.

Table 3

Key Lengths for Each of the 22 Keys

Key	No. of responses scored per phi cutoff .15	.10	.07	.05	.03
1	4	52	153	347	671
2	11	56	199	391	697
3	22	116	283	502	785
4	14	92	251	455	780
5	20	89	237	457	762
6	20	105	281	491	815
7	20	115	293	517	834
8	30	126	297	506	808
9	15	76	197	399	700
10	21	100	265	458	765
11	1	67	224	435	744
12	9	36	170	338	627
13	4	35	110	260	572
14	9	81	218	406	678
15	16	82	198	395	733
16	19	99	259	445	730
17	13	67	186	348	650
18	2	31	136	289	601
19	6	45	156	303	580
20	17	74	210	385	676
21	8	52	156	307	590
22	45	159	326	507	804

Group 2 Analyses

Analyses on Group 2 involved the determination of optimum key length for each of the 22 keys. The results of these analyses are presented in Table 4. As can be seen, optimum key length and peak pattern varies from key to key. Although much of this variation may be due to chance effects in item and group sampling, trends for the larger groups should be relatively stable. For the type of data used in this study, most key validities appear to reach a plateau rather early. In this connection, Abrahams (1967) has noted that for an item pool with a given range of response validity, there is very little increase in key validity after 100 to 150 item responses have been combined. Key lengths used in all subsequent scoring are underlined in Table 4.

As can be seen in Table 4, results for 110 subkeys were obtained. The SIB responses of each student were scored on 4 keys (20 subkeys). These four keys included the general key scored on all students and three keys selected according to the student's ability moderator score. For 4 of the 20 subkeys, more than 700 item responses were often scored. And yet, with the FORTRAN program that was written, scoring proceeded at the rate of less than one second per student on an IBM 360 Model 44. Data on key statistics were accumulated as the scoring was done. Hence, the type of analyses performed would appear to be well within the limits of practicality.

No conclusions relevant to project objectives can be made from Group 2 analyses.

Group 3 Analyses

As described in the section on general design and analyses, rules were formulated for determining the optimum number and combination of moderated scoring keys required for maximizing the predictive effectiveness of the item pool. These rules were applied to the point biserial correlations presented in Table 5. As can be seen, the general key was retained for scoring at eight of the twelve ability levels. Keys 17 and 18, both first order combination keys, were retained at ability levels 9-10 and 11-12, respectively. Thus, it would appear that ability serves as a moderator variable only for the more able students. Even for this group, the effect is not great.

When applied to real data, the procedures and rules developed for determining the optimum number of moderated scoring keys seem to work well. It would appear that both the procedures and the rules can be readily applied to other data if sample sizes are substantial. Hence, on this basis, it is concluded that primary objective number one has been achieved.

Results at this stage of the analyses make it doubtful that subsidiary hypothesis S1 will be confirmed. However, Group 4 analyses are required before a conclusion can be reached.

Classification procedures used for developing a common score scale in the form of probabilities of college attendance group membership were

Table 4

Relationship Between Key Length and Key Validity for Each of the 22 Keys

Key	Sample size: Attend	Sample size: Nonattend	$r_{pt\ bis}$ per phi cutoff: .15	.10	.07	.05	.03
1	32	258	.197	.164	.157	.185	.180
2	45	220	.233	.254	.258	.272	.270
3	59	185	.334	.348	.345	.344	.329
4	96	187	.405	.444	.443	.448	.432
5	117	168	.460	.455	.457	.463	.466
6	155	144	.333	.375	.378	.367	.392
7	147	136	.330	.413	.407	.399	.398
8	154	108	.434	.479	.457	.4[illegible]5	.428
9	206	80	.343	.360	.341	.334	.326
10	229	78	.367	.332	.387	.382	.377
11	230	55	.135	.321	.330	.315	.324
12	261	37	.295	.312	.317	.325	.294
13	77	478	.26[illegible]	.251	.254	.241	.274
14	155	372	.341	.386	.404	.412	.405
15	272	312	.387	.408	.418	.428	.426
16	301	244	.347	.447	.460	.438	.436
17	435	158	.294	.359	.357	.378	.369
18	491	92	.154	.346	.341	.349	.343
19	232	850	.363	.350	.365	.361	.398
20	573	556	.394	.443	.456	.449	.452
21	926	250	.347	.377	.390	.396	.389
22	1731	1656	.540	.576	.566	.565	.558

Note.--The key length that was used in subsequent scoring is underlined.

Table 5

Determination of Optimum Number of Scoring Keys for Group 3

Key group	Point biserial correlations: Appropriate key	Higher order key: 22	21	20	19	18	17	Sample size: Attend	Nonattend
				Step 1[a]					
21	.372	.357						931	242
20	.454	.404						538	593
19	.406	.408						242	847
				Step 2[a]					
18	.278		.249					461	76
17	.410		.415					450	166
16	.423	.466						303	261
15	.457	.479						235	332
14	.474	.460						156	378
13	.244	.275						86	469
				Step 3[a]					
12	.240					.248		259	27
11	.258					.271		222	49
10	.300						.346	261	73
9	.469						.457	189	93
8	.359	.421						169	125
7	.455	.508						134	136
6	.454	.478						131	158
5	.416	.470						104	174
4	.503	.470						77	193
3	.417	.456						79	185
2	.137	.218						55	233
1	.251	.336						31	236

Note.--Keys chosen at each step are underlined.

[a]Refers to steps for determining optimum key length. See section on design.

described in the section on general design and analyses. These procedures require data on criterion group sizes, means, variances, and, where two predictors are used in combination, covariances. This data is presented for analysis Group 3 in Table 6. Standard deviations are shown instead of variances. In addition, t test data is presented for each of the point biserials. As can be seen from these t values, it is highly unlikely that any of the correlations could have been obtained by chance.

The common score scale used in this project would appear to satisfactorily meet the requirements of primary objective two. In principle, it involves the use of criterion predictions as a common basis for anchoring the various moderated scoring key score distributions. This procedure should work equally well whether criterion group membership or criterion scores are being predicted. In the latter case, regression rather than classification procedures could be used to form the common score scale.

Under the suggested system, the score reported for an individual on a moderated key is essentially a prediction directly relevant to the purpose for which the key was constructed and is to be used. Whenever data are available on the actual predictive validity of an instrument for the purposes and group on which it is to be used, it should be possible to develop a common scale for reporting moderated scoring key results. Hence it is concluded that primary objective two has been achieved.

Group 4 Analyses

The main purpose of the Group 4 analyses was to determine the accuracy of attendance, nonattendance predictions so that the relative effectiveness of the four predictors developed in this study would be revealed. The results of these analyses are directly relevant to achievement of the secondary objective of the project, and hence to testing the general and subsidiary hypotheses.

Data summarizing the relationship between predicted and actual status on the criterion are summarized in Table 7 for each of the four predictors. It is immediately apparent from the hit rates and phi values that, in this particular situation, use of moderated keys hold no particular advantage over use of the general key. The phi coefficient (.54) and hit rate (77%) for the moderated key set were both only slightly higher than the phi coefficient (.52) and hit rate (76%) obtained with the general key. Thus, academic aptitude does not appear to be an effective moderator variable for the predictor and criterion under study.

The same data are presented in a different form in Table 8 where joint hit, miss rates are shown. The most interesting aspects of this data are the instances in which one predictor made a correct prediction (a hit) while the other yielded a miss. For example, the moderated keys correctly predicted 135 of the general key misses while the general key

Table 6

Descriptive Data for the Optimum Set of Moderated Scoring Keys for Group 3

Ability group	Key used	Attend			Nonattend			$r_{pt\ bis}$	t
		N	$\bar{X}$	SD	N	$\bar{X}$	SD		
1- 8	22	780	91.1	10.6	1440	78.2	10.6	.50	27.3
9-10	17	450	188.9	15.6	166	173.1	15.4	.41	11.1
11-12	18	481	186.7	13.0	76	175.7	13.4	.28	6.8
1-12	K22	1711	96.1	11.3	1682	80.0	11.5	.58	41.1
1-12	MV	1711	1096.6	145.3	1682	917.0	171.6	.48	32.0

Note.--For purposes of information, data is also presented for the general key (K22) and for the moderator variable (MV). The covariance for K22 and MV in the attend group was 811.6. For the nonattend group the covariance was 921.4.

Table 7

Comparison of Accuracy of 4 Predictors in the Prediction of College Attendance for Group 4

Actual status	Predicted status: Non-attend	Predicted status: Attend	Hit rate	Phi	X^2
	Moderated Keys				
Attend	386 (11%)	1343 (39%)			
Non-attend	1273 (37%)	402 (12%)			
			77%	.54	981.1
	General Key				
Attend	417 (12%)	1312 (38%)			
Non-attend	1278 (38%)	397 (12%)			
			76%	.52	926.6
	Moderator Variable				
Attend	372 (11%)	1357 (40%)			
Non-attend	1056 (31%)	619 (18%)			
			71%	.42	602.6
	Best Combination of General Key and Moderator Variable				
Attend	379 (11%)	1350 (40%)			
Non-attend	1264 (37%)	411 (12%)			
			77%	.54	976.8

Note.--Per cent of total N given in parentheses is rounded. Hit rate represents a separate rounding of totals.

Table 8

Joint Hit, Miss Rates for Moderated Keys
Versus Other Three Predictors

Moderated key	Miss	Hit	z[a]
	General Key		
Hit	135 (4%)	2481 (73%)	
Miss	679 (20%)	109 (3%)	-1.66
	Moderator Variable		
Hit	493 (14%)	2123 (62%)	
Miss	498 (15%)	290 (9%)	-7.26
	Best Combination of General Key and Moderator Variable		
Hit	137 (4%)	2479 (73%)	
Miss	653 (19%)	135 (4%)	-0.12

[a] *z* test for nonindependent proportions performed according to McNemar (1962, pp. 52-56).

correctly predicted 109 of the moderated key misses. Although this difference in favor of the moderated key is statistically significant at the five per cent level (one-tailed test), it is certainly not great. On the other hand, the difference in favor of the moderated keys is substantial when compared with the moderator variable. Finally, it can be seen that there is very little difference between the joint hit, miss rate for the moderated keys and the general key in combination with the moderator variable.

On the basis of these results, it can be concluded that subsidiary hypotheses S1 and S2 are not supported. Thus the general hypothesis relevant to the secondary objective of the project is also not supported. While subsidiary hypothesis S3 is supported, the overall conclusion with respect to the secondary objective of the project is that, while the moderated scoring key techniques can be generalized to the data under investigation, their effectiveness with this data does not warrant their practical application to it. In short, academic aptitude does not appear to be an effective moderator variable for the data under study.

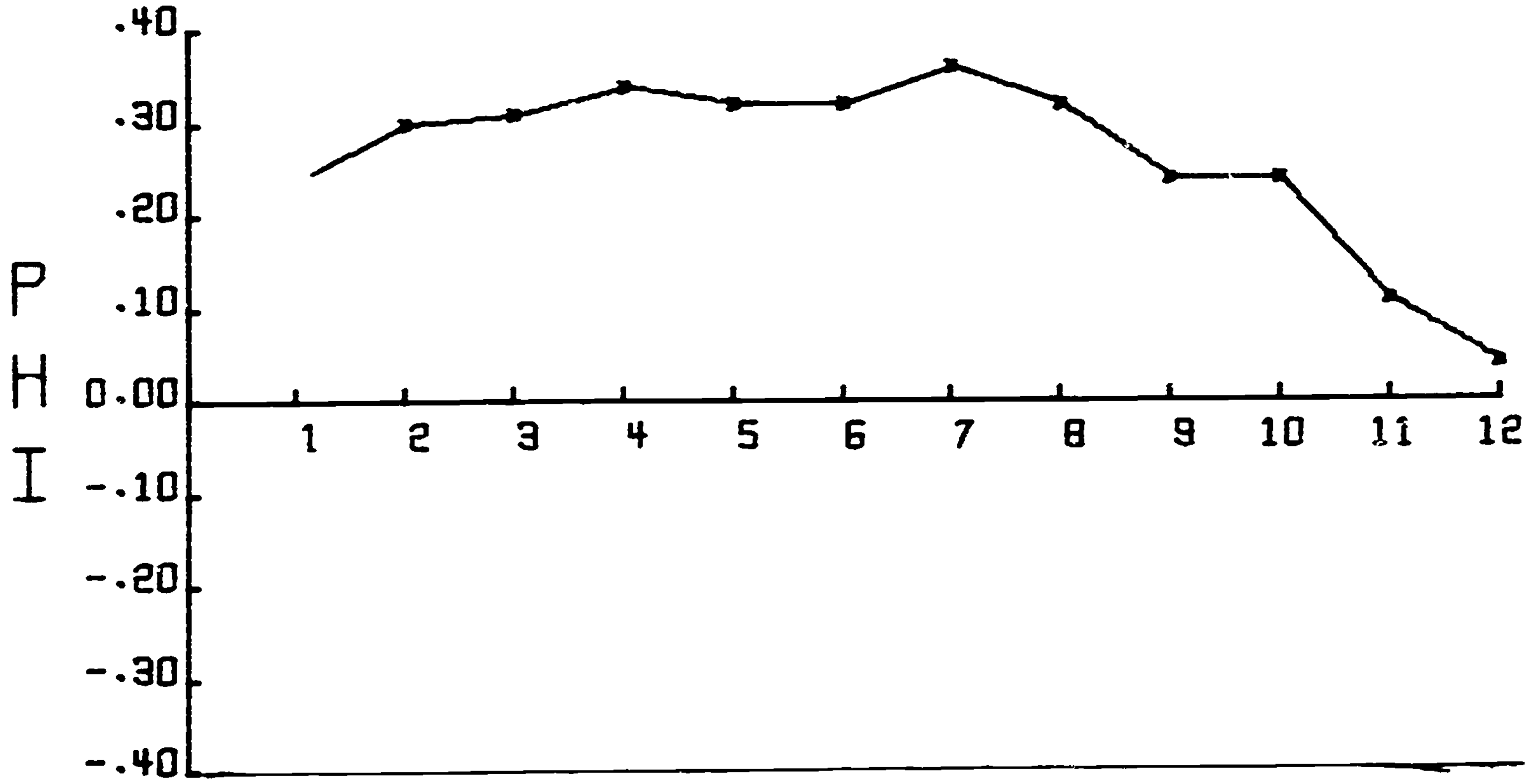
It should be pointed out that the scoring techniques developed in this study exist apart from the data to which they are applied. While it may be that the techniques will be of no practical value in any setting, such a conclusion is not warranted on the basis of the results obtained in this study. It would appear that the application of moderated scoring techniques to settings in which moderator variables have been known to be operating would be required to determine their actual value.

Although academic aptitude was shown to be ineffective as a moderator variable when applied to the SIB as a whole, it may still be effective with a few of the items in the SIB. That is, the level of relationship between selected biodata items and college attendance may, itself, be related to the ability moderator. Figures 1, 2, and 3 are presented as evidence of this relationship. In each, the phi coefficients for selected item responses are shown for the twelve ability levels. Definite trends can be seen in each plot. It should be noted that these plots were selected to illustrate a point. Detailed analysis and discussion of plot data is beyond the scope of this report. However, the general type of relationship shown in Figures 1 through 3 must be present for a number of item responses in order for moderated scoring keys to be effective with a given set of data.

One of the interesting aspects of the results presented in Tables 7 and 8 is the level of accuracy achieved in the predictions. A 77% hit rate in the prediction of a gross and arbitrarily defined criterion such as college attendance is seldom found, especially with a base rate similar to the one in the sample under study. While the level of relationship expressed by a phi of .54 may not seem high, one must remember that the phi coefficient is smaller than the Pearson product moment correlation coefficient in situations where it is possible to compute both (Guilford 1956). It is also smaller than the coefficient obtained when the point biserial correlation technique is used. This is apparent when the Group 4 phi coefficients of .52 and .42 for the general key and moderator variable are compared with the point biserial correlation coefficients of .58 and

Figure 1

Phi Trends Across 12 Ability Levels
for Item 65, Response 2[a]

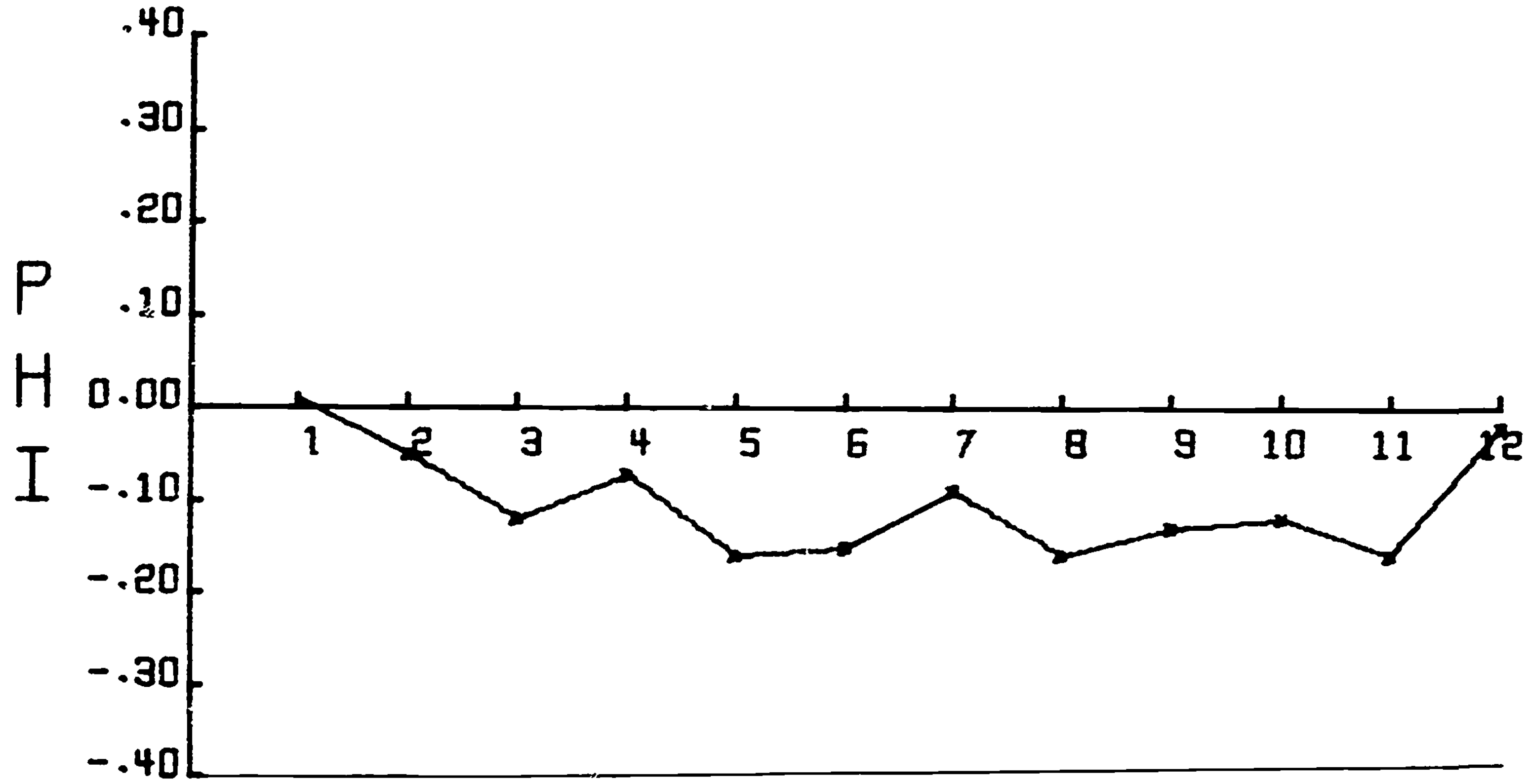


Note.--The horizontal axis represents the 12 ability levels.

[a]Project item code refers to SIB item number 91, response B.

Figure 2

Phi Trends Across 12 Ability Levels
for Item 71, Response 2[a]

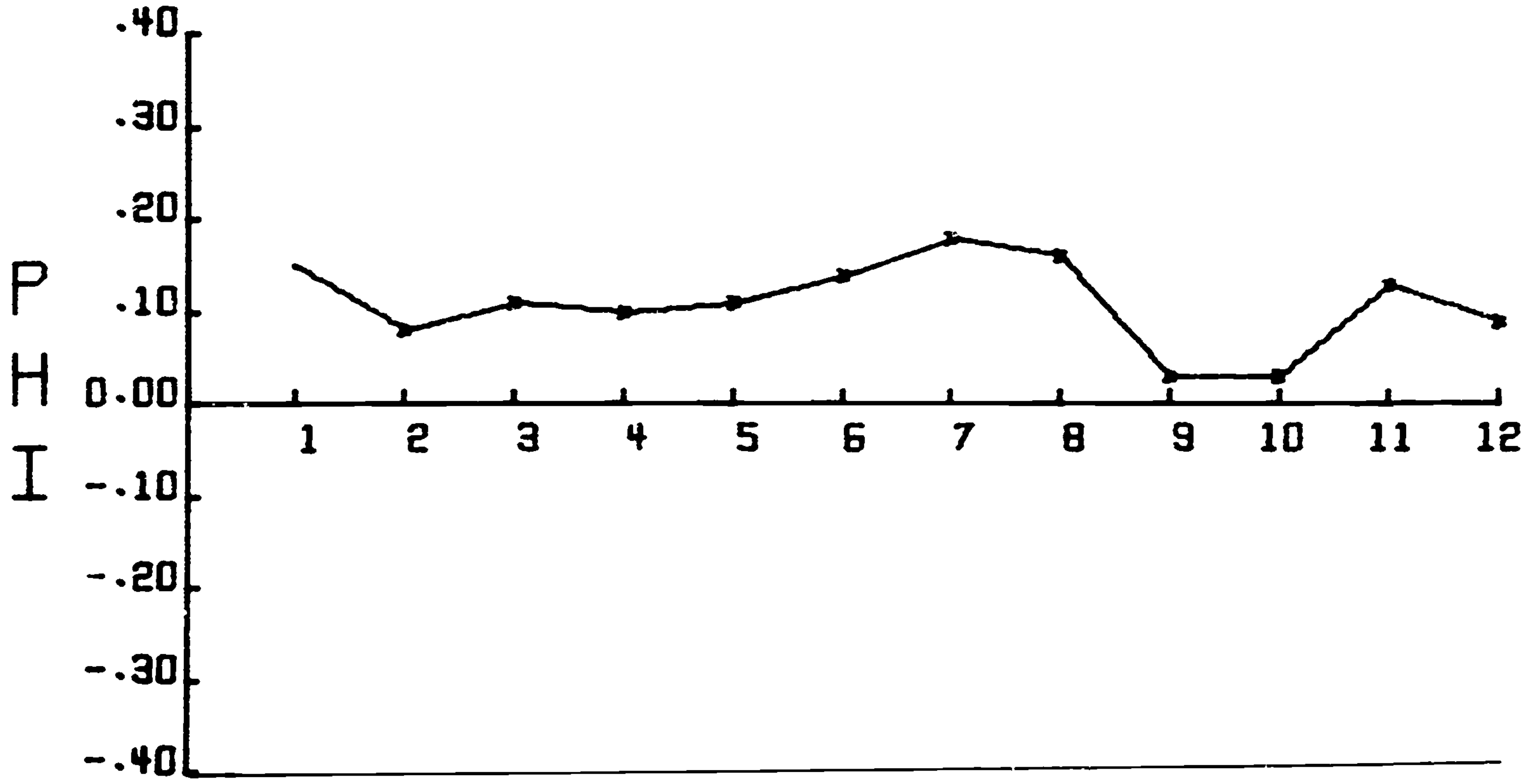


Note.--The horizontal axis represents the 12 ability levels.

[a]Project item code refers to SIB item number 97, response B.

Figure 3

Phi Trends Across 12 Ability Levels
for Item 96, Response 6[a]



Note.--The horizontal axis represents the 12 ability levels.

[a]Project item code refers to SIB item number 122, response F.

.48 obtained independently for raw scores versus criterion status. The phi coefficient for the general key and the moderator variable in weighted combination was .54. The equivalent point biserial correlation was .60. These point biserial correlations also underestimate the Pearson product moment coefficient that would apply if the criterion (college attendance, or tendency toward same) could be considered to be continuous and normally distributed in the population under study. In this latter case, biserial correlation would be appropriate. The biserial coefficient equivalent to a point biserial coefficient of .60 is .75, a value which may more accurately represent the level of relationship and predictive accuracy achieved in this study.

In any case, it is interesting to note that Project TALENT (Flanagan, et al., 1964) reports a multiple point biserial correlation of only .566 between 51 cognitive predictors and college attendance for a sample highly similar to the one used in this study. Although the nature of the criterion differs slightly from that of the present study, there is sufficient similarity to reveal the power of biodata as a predictor.

Conclusions

Conclusions already stated in the section on results are summarized below as they pertain to the primary and secondary objectives.

Primary Objective

The primary objective of the project was essentially the exploration and refinement of techniques for developing moderated scoring keys. The two specific tasks which were set are separately discussed below.

(a) Development of procedures for determining the optimum number of subgroups (and hence, moderated scoring keys) required for maximizing the predictive effectiveness of an inventory.

The procedures developed involve the formation of as many subgroups of equal size as is warranted by the size of the available sample. It is considered important to maintain the natural criterion distribution in each subgroup rather than achieve some arbitrary success-failure split. Sets of moderated scoring keys differing widely in number per set can easily be formed from item analysis data on each subgroup and by combination of data from pairs of subgroups. Thus separate item analyses for each set of moderated keys is avoided. Rules were formulated for judging when a more general moderator key could replace a specific key. These rules involve the comparison of key-criterion correlations for different keys scored on the same group. The procedures and rules when applied to actual data through use of a medium-scale computer prove workable and economically feasible. (In fact, the total analyses for this project took slightly over seven hours on an IBM 360, Model 44. FORTRAN programs written especially for the project, were used in all analyses.) Hence, it is concluded that the first task under the primary objective has been achieved. The second task was as follows:

(b) Development of a single scale for reporting the scores obtained from a set of moderated scoring keys, thus facilitating the practical application of moderated keys to prediction problems.

In order to develop an anchor scale, it is proposed that criterion estimates be used as the basis for equating scores from different moderated scoring keys. In this way a common scale can be formed which, in addition, has direct relevance to the purpose for which scores are obtained. The application of criterion group membership probabilities to the data under study proved practical and economically feasible. The procedure should work equally well whether criterion group membership or criterion scores are being predicted. In the latter case, regression rather than classification procedures could be used to form the common score scale. Hence, it is concluded that the second task under the primary objective has been achieved.

Secondary Objective

The secondary objective of the project can be described as the investigation of the generalizability of moderated scoring key techniques to a sample different from that on which they were first developed. The general hypothesis was that the contribution of biodata to the prediction of college attendance is greater, upon cross-validation, for scoring keys developed <u>within</u> ability level subgroups (<u>e.g.</u>, <u>moderated keys</u>) than for a key developed on the total group (<u>e.g.</u>, <u>the general key</u>). For purposes of evaluation, this general hypothesis was restated in the form of the first of the three subsidiary hypotheses listed below.

<u>Subsidiary hypotheses</u>. Moderated scoring key predictions of college attendance-nonattendance for high school students are more accurate than predictions based on: (a) normal empirical keying procedures used in forming the general key; (b) the above, plus academic aptitude data; and, (c) academic aptitude data alone. Although a statistically significant difference (p .05) in favor of the moderated keys was obtained, the practical implications of this difference were slight. Hence, it is concluded that support for the first subsidiary hypothesis, and thus, the general hypothesis is not warranted by the data. Support for the second subsidiary hypotheses is also unwarranted. Predictions were as accurate but not more accurate. Finally, support for the third subsidiary hypothesis is warranted. The moderated scoring keys produced a substantially higher percentage of correct predictions.

Byproduct of Study

Biographical data were shown in this study to be powerful predictors of college attendance. The fact that a single biographical inventory produced correct predictions of college attendance for 77% of the cases is especially surprising in view of the arbitrary way in which the criterion was defined. Equally surprising, in light of previous research, is the relative performance of biographical data and academic aptitude as predictors of college entrance. The predictive accuracy of the former substantially exceeded that of the latter. In a reverse of their usual roles, academic aptitude added little to the predictive accuracy obtained through use of biographical data alone. Finally as noted in the section on results, the correlation between biographical data and the criterion was substantially greater than that achieved by means of a multiple regression equation combining [illegible] variables (Flanagan, et al., 1964). Hence, biographical data would appear to have considerable promise as predictors of who will attend college. With some attention to age-related items, the scoring key data presented in Appendix C might be used to develop an inventory for this purpose.

A final point needs to be made. Although the moderated scoring key techniques developed in this project were not found to produce more valid results than normal keying methods, the techniques themselves have not been shown to be without merit. Similar, though less elaborate techniques have been shown to work in a different setting (Prediger, 1966). It well may be that moderator variables are situation specific as Ghiselli and Sanders (1967) suggest.

The procedures developed in this study could easily be applied to other settings, provided that a large base of data is available for analysis. It would appear that the next step would most profitably involve application of moderated scoring key techniques in a situation where moderator variables have already been shown to be operating.

References

Abrahams, N. M. SVIB Keylength: dissident data. J. appl. Psychol., 1967, 51, 266-273.

Appel, V. and Kipnis, D. The use of levels of confidence in item analysis. J. appl. Psychol., 1954, 38, 256-259.

Barger, B. and Hall, E. The interaction of ability levels and socio-economic variables in the prediction of college drop-outs and grade achievement. Educ. psychol. Measmt, 1965, 25, 501-508.

Berdie, R. J. and Hood, A. B. How effectively do we predict plans for college attendance? Personnel Guidance J., 1966, 44, 487-493.

Brown, F. G. and Scott, D. A. The unpredictability of predictability. J. educ. Measmt, 1966, 3, 297-302.

Clark, K. E. The vocational interests of nonprofessional men. Minneapolis: University of Minnesota Press, 1961.

Clark, K. E. and Campbell, D. P. Manual for Minnesota Vocational Interest Inventory. New York: Psychol. Corp., 1965.

Cooley, W. W. and Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Dunnette, D. A modified model for test validation and selection research. J. appl. Psychol., 1963, 47, 317-323.

Flanagan, J. C.; Davis, F. B.; Dailey, J. T.; Shaycroft, M. F.; Orr, D. B.; Goldberg, Isadore, and Neyman, C. A., Jr. The American high-school student. Pittsburg, Pa.: Project TALENT Office, University of Pittsburgh, 1964.

Flanagan, J. C. and Cooley, W. W. (Edit.) Project TALENT one-year follow-up studies. Pittsburgh, Pa.: University of Pittsburgh School of Education, 1966.

Ghiselli, E. E. The prediction of predictability. Educ. psychol. Measmt, 1960, 20, 3-8.

Ghiselli, E. E. Moderating effects and differential reliability and validity. J. appl. Psychol., 1963, 47, 81-86.

Ghiselli, E. E. and Sanders, E. P. Moderating heteroscedasticity. Educ. psychol. Measmt, 1967, 27, 581-590.

Gleser, Goldine C. and DuBois, P. H. A successive approximation method for maximizing test validity. Psychometrika, 1951, 16, 129-139.

Goodstein, L. D. and Heilbrun, A. B., Jr. Prediction of college achievement from the Edwards Preference Schedule at three levels of intellectual ability. J. appl. Psychol., 1962, 46, 317-320.

Guilford, J. P. Fundamental statistics in psychology and education (3rd edition). New York: McGraw-Hill, 1956.

Hakel, M. D. Prediction of college achievement from the Edwards Personal Preference Schedule using intellectual ability as a moderator. J. appl. Psychol., 1966, 50, 336-340.

Heilbrun, A. B. Jr. Personality factors in college dropout. J. appl. Psychol., 1965, 49, 1-7.

Hobert, R. and Dunnette, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. J. appl. Psychol., 1967, 51, 50-64.

Korman, A. K. Self-esteem variable in vocational choice. J. appl. Psychol., 1966, 50, 479-486.

Korman, A. K. Self-esteem as a moderator of the relationship between self-perceived abilities and vocational choice. J. appl. Psychol., 1967, 51, 65-67.

Kuder, G. F. A comparative study of some methods of developing occupational keys. Educ. psychol. Measmt, 1951, 11, 16-22.

McNemar, Q. Psychological statistics (3rd edition). New York: Wiley, 1962.

Prediger, D. J. Prediction of persistence in college. J. counsel. Psychol., 1965, 12, 62-67.

Prediger, D. J. Application of moderated scoring keys to prediction of academic success. Amer. educ. Res. J., 1966, 3, 105-111.

The Project TALENT Data Bank. Pittsburgh, Pa.: Project TALENT Office, University of Pittsburgh, 1965.

Rock, D. A.; Barone, J. L.; and Linn, R. L. A FORTRAN computer program for a moderated stepwise prediction system. Educ. psychol. Measmt, 1967, 27, 709-713.

Schoenfeldt, L. F. The grouping of subjects into homogeneous subsets--a comparison and evaluation of two divergent approaches. Unpublished doctoral dissertation, Purdue University, 1966.

Appendices

A. MSK-SIB Item Equivalents

B. Changes in SIB Response Option Coding

C. Scoring Keys

D. Item Two of Project TALENT 1961 Follow-up Questionnaire

Appendix A

MSK-SIB Item Equivalents[a]

MSK No.	SIB No.	MSK No.	SIB No.	MSK No.	SIB No.
1	1	48	48	95	121
2	2	49	49	96	122
3	3	50	50	97	123
4	4	51	51	98	124
5	5	52	52	99	125
6	6	53	53	100	126
7	7	54	54	101	127
8	8	55	55	102	128
9	9	56	56	103	129
10	10	57	57	104	130
11	11	58	58	105	131
12	12	59	59	106	132
13	13	60	60	107	133
14	14	61	61	108	134
15	15	62	62	109	135
16	16	63	63	110	136
17	17	64	64	111	137
18	18	65	91	112	138
19	19	66	92	113	139
20	20	67	93	114	140
21	21	68	94	115	141
22	22	69	95	116	142
23	23	70	96	117	143
24	24	71	97	118	144
25	25	72	98	119	145
26	26	73	99	120	146
27	27	74	100	121	147
28	28	75	101	122	148
29	29	76	102	123	149
30	30	77	103	124	150
31	31	78	104	125	151
32	32	79	105	126	152
33	33	80	106	127	153
34	34	81	107	128	154
35	35	82	108	129	155
36	36	83	109	130	156
37	37	84	110	131	157
38	38	85	111	132	167
39	39	86	112	133	168
40	40	87	113	134	169
41	41	88	114	135	170
42	42	89	115	136	171
43	43	90	116	137	172
44	44	91	117	138	173
45	45	92	118	139	174
46	46	93	119	140	175
47	47	94	120	141	176

MSK No.	SIB No.	MSK No.	SIB No.	MSK No.	SIB No.
142	177	186	256	230	208
143	178	187	257	231	213
144	179	188	258	232	214
145	180	189	259	233	215
146	181	190	260	234	216
147	182	191	261	235	227
148	183	192	262	236	218
149	184	193	263	237	219
150	185	194	264	238	220
151	186	195	265	239	221
152	187	196	266	240	222
153	188	197	267	241	223
154	189	198	268	242	226
155	190	199	269	243	230
156	191	200	270	244	231
157	192	201	271	245	224
158	193	202	272	246	225
159	194	203	273	247	228
160	195	204	274	248	229
161	196	205	275		
162	197	206	276		
163	198	207	277		
164	199	208	278		
165	200	209	279		
166	201	210	280		
167	202	211	281		
168	203	212	282		
169	204	213	283		
170	205	214	284		
171	241	215	285		
172	242	216	286		
173	243	217	287		
174	244	218	288		
175	245	219	289		
176	246	220	290		
177	247	221	291		
178	248	222	292		
179	249	223	293		
180	250	224	294		
181	251	225	295		
182	252	226	296		
183	253	227	206		
184	254	228	206		
185	255	229	208		

[a]The MSK No. refers to the item identification number used in this project. The SIB No. refers to the equivalent Student Information Blank number. For copy of SIB, See Flanagan, et. al., (1964).

Appendix B

Changes in SIB Response Option Coding

Project item No. and responses	SIB item No. and responses	Project item No. and responses	SIB item No. and responses
51: 6 1-5	51: A B-F	238: 1-6 7 8,9	200: A-F G,H I,J
53: 3-6 2,1	53: A-D E,F	239,240 241,242 243,244: 1-7 8 9	221,222 223,226 230,231: A-G H,I J-L
222: 6 1-5	292: A B-F	245: 1-6 7 8 9	224: A-F G,H I,J K,L
227,229: 1 2-9 228,230: 1-8	206,208: Q A-H I-P	246,247: 1 2-8 9	225,228: A-C D-J K,L
231: 1-5 6 7,8,9	213: A-E F&I G,H,J	248: 9 1 2-7 8	229: A B,C D-I J,K
232: 1-4 5 6-9	214: A-D E,G F-J		
233 234,235: 1 2-9	215 216,227: A,B C-J		
236&237: 1 2-9 0	218&219: A,B C-J K		

Appendix C

Scoring Keys

General Explanation

Twenty-two scoring keys were formed as a result of item analyses conducted on Group 1. Five key lengths were scored for each key. The items and responses for the general key and keys 17 and 18 are presented separately for responses scored +1 (indicating a tendency toward college attendance) and -1 (indicating a tendency toward nonattendance). Data for the other keys have been omitted since they were not included among the final scoring keys.

In the table on the next and following pages the symbols listed below are used.

K: key identification number

17 is the key scored on ability levels 9 & 10
18 is the key scored on ability levels 11 & 12
22 = general key

L: key length index. If a negative sign appears before L, the responses are scored -1. If no sign appears, they are scored +1.

1 = item responses with phi coefficients equal to or greater than .15.
2 = item responses with phi coefficients equal to or greater than .10, <u>including</u> responses already listed for the key.

3 = item responses with phi coefficients equal to or greater than .07, <u>including</u> responses already listed for the key.

4 = item responses with phi coefficient equal to or greater than .05, <u>including</u> responses already listed for the key.

5 = item responses with phi coefficients equal to or greater than .03, <u>including</u> responses already listed for the key.

Items and responses are represented as one integer. The units digit represents the response number. All other digits represent the item number. Thus, 2284 would stand for project item number 228, response 4.

K	L	ITEMS AND RESPONSES														
17	1	652														
17	2	11	31	125	132	301	311	534	702	771	811	831	843	873	886	996
17	2	1522	1604	1882	2376											
17	3	12	21	22	23	61	62	106	112	234	235	271	302	615	715	733
17	3	736	786	801	812	821	823	833	841	842	872	926	1006	1074	1232	1375
17	3	1385	1394	1415	1425	1474	1482	1575	1585	1596	1791	1801	1981	1991	2224	2286
17	3	2287	2326	2366	2374	2392	2401	2447	2448	2483						
17	4	24	33	91	113	114	115	126	151	162	164	205	215	225	263	281
17	4	342	371	373	394	405	453	475	502	511	572	606	714	716	726	756
17	4	796	832	871	912	1014	1065	1121	1133	1156	1161	1162	1173	1192	1193	1202
17	4	1204	1224	1231	1273	1302	1354	1366	1384	1451	1473	1484	1506	1556	1645	1662
17	4	1703	1731	2192	2212	2213	2252	2262	2284	2307	2315	2335	2344	2357	2367	2375
17	4	2411	2446	2449												
17	5	34	35	46	56	92	94	95	104	141	152	153	173	252	294	321
17	5	355	392	393	414	431	464	491	492	493	522	531	552	583	603	604
17	5	616	623	625	631	634	661	671	701	725	735	764	813	822	851	861
17	5	862	884	896	931	966	976	1004	1015	1026	1063	1064	1085	1091	1112	1131
17	5	1134	1144	1145	1163	1182	1184	1194	1201	1222	1263	1284	1285	1303	1305	1333
17	5	1341	1402	1403	1405	1423	1426	1435	1436	1441	1475	1483	1486	1492	1495	1496
17	5	1503	1505	1523	1526	1533	1542	1545	1565	1576	1583	1584	1595	1612	1613	1615
17	5	1621	1623	1644	1753	1755	1774	1781	1872	1941	1971	2011	2021	2102	2142	2143
17	5	2153	2174	2186	2216	2231	2244	2283	2285	2292	2304	2328	2353	2365	2368	2369
17	5	2381	2391	2423	2445	2457	2469	2482								
17	-1	26	36	111	536	651	834	835	845	874	881	922	2441			
17	-2	16	66	121	231	306	376	601	654	704	712	731	815	844	854	875
17	-2	991	1001	1136	1226	1236	1382	1481	1521	1531	1571	1572	1581	1602	1663	1664
17	-2	1881	2361	2413	2442	2489										
17	-3	96	101	155	366	395	496	516	521	544	576	611	636	664	706	713
17	-3	722	724	732	755	765	773	774	775	791	805	814	826	846	882	911
17	-3	962	1061	1081	1102	1166	1196	1206	1306	1361	1381	1392	1404	1471	1491	1501
17	-3	1553	1616	1641	1683	1702	1761	1792	1802	1982	1992	2275	2279	2281	2298	2371
17	-3	2372	2383	2398	2399	2429										
17	-4	131	165	185	193	221	232	254	262	292	313	336	346	375	402	471
17	-4	545	555	621	694	705	711	763	776	782	783	836	864	865	893	895
17	-4	921	934	963	1011	1072	1095	1113	1146	1276	1286	1336	1346	1353	1371	1372
17	-4	1391	1411	1421	1422	1444	1453	1541	1551	1552	1564	1582	1591	1601	1626	1673
17	-4	1674	1723	1733	1734	1776	2071	2182	2211	2234	2251	2271	2274	2299	2303	2308
17	-4	2319	2348	2396	2403	2405	2417	2428	2431	2452						
17	-5	42	43	52	102	103	122	136	143	171	172	201	212	222	223	233
17	-5	273	274	304	312	314	316	334	343	353	382	401	415	455	462	481
17	-5	501	514	525	556	561	584	586	591	624	644	655	665	666	672	674
17	-5	675	692	693	721	723	734	741	743	751	752	781	784	785	816	825
17	-5	855	866	933	935	944	956	961	971	972	992	1021	1023	1024	1042	1075
17	-5	1093	1126	1153	1176	1185	1186	1195	1214	1215	1266	1296	1314	1325	1343	1344
17	-5	1352	1376	1412	1433	1443	1445	1456	1465	1472	1561	1563	1592	1594	1624	1625
17	-5	1634	1642	1666	1684	1694	1735	1751	1772	1782	1871	1942	2051	2061	2101	2141
17	-5	2191	2195	2201	2223	2241	2264	2272	2277	2278	2293	2296	2314	2321	2325	2329
17	-5	2332	2337	2342	2346	2347	2351	2352	2358	2362	2363	2373	2382	2387	2394	2397
17	-5	2402	2404	2408	2414	2415	2426	2427	2464	2479	2486					

K L	ITEMS AND RESPONSES														
18 1															
18 2	405	811	831	966	976	1604									
18 3	21	61	235	311	475	522	652	716	733	771	801	802	821	822	841
18 3	861	871	872	1006	1026	1224	1273	1375	1384	1415	1451	2143	2287	2369	2446
18 3	2449														
18 4	31	34	56	92	106	115	133	175	195	225	255	362	371	435	445
18 4	465	474	482	491	534	541	571	584	635	715	786	794	796	812	886
18 4	911	926	931	996	1132	1156	1201	1232	1385	1401	1416	1424	1441	1461	1475
18 4	1482	1522	1556	1575	1586	1731	2162	2204	2224	2245	2252	2284	2326	2334	2344
18 4	2366	2376	2377	2448	2468	2483									
18 5	11	12	13	14	22	24	32	46	86	93	94	112	113	114	124
18 5	125	126	144	145	152	161	174	205	212	215	222	245	265	271	275
18 5	294	301	302	334	335	355	356	372	393	394	425	472	492	512	513
18 5	552	582	592	596	602	642	661	671	683	702	704	726	754	756	761
18 5	823	832	842	843	892	923	941	955	986	1036	1041	1062	1103	1104	1111
18 5	1121	1122	1134	1144	1162	1163	1171	1173	1174	1184	1193	1202	1221	1222	1234
18 5	1244	1291	1292	1293	1303	1304	1305	1341	1363	1364	1394	1405	1434	1473	1492
18 5	1493	1503	1523	1524	1525	1533	1542	1565	1584	1612	1621	1644	1645	1662	1703
18 5	1711	1715	1753	1764	1872	1991	2042	2052	2082	2142	2144	2152	2153	2163	2172
18 5	2176	2192	2196	2205	2231	2278	2302	2315	2316	2317	2328	2335	2367	2381	2392
18 5	2423	2438	2439	2445	2451	2453	2462	2473	2484						
18-1	834	874													
18-2	26	121	376	471	651	732	804	805	814	816	825	826	844	845	1061
18-2	1136	1382	1521	1663	2141	2361	2387	2441							
18-3	16	36	66	102	111	171	211	231	306	314	431	441	454	496	525
18-3	526	654	656	666	672	712	713	731	755	765	773	783	791	815	853
18-3	864	895	922	961	962	972	973	1001	1023	1025	1102	1126	1154	1166	1226
18-3	1236	1276	1296	1306	1392	1404	1462	1471	1501	1603	1641	1673	2151	2161	2171
18-3	2221	2236	2241	2251	2324	2331	2341	2362	2371	2372	2398	2399	2442	2443	
18-4	96	141	143	191	221	253	262	401	434	504	543	576	601	611	625
18-4	653	673	682	703	721	723	785	835	862	875	881	956	963	965	991
18-4	1034	1042	1043	1071	1096	1105	1155	1176	1225	1255	1336	1361	1362	1371	1372
18-4	1412	1414	1442	1443	1453	1456	1531	1541	1553	1564	1571	1572	1581	1601	1634
18-4	1642	1665	1674	1732	1735	1871	2191	2201	2275	2279	2281	2298	2299	2304	2314
18-4	2316	2329	2339	2342	2346	2394	2408	2414	2425	2464	2466	2482			
18-5	33	43	53	55	75	82	91	105	202	213	224	233	251	254	273
18-5	283	292	316	322	345	392	402	403	423	483	515	521	546	556	[illegible]
18-5	614	636	641	655	664	674	684	691	705	706	711	741	753	764	772
18-5	782	824	836	846	854	855	865	916	934	936	952	974	[illegible]	[illegible]	[illegible]
18-5	1002	1011	1021	1031	1053	1091	1113	1146	1246	1266	1295	1312	1321	1322	[illegible]
18-5	1345	1353	1356	1383	1406	1411	1413	1422	1431	1436	1446	1452	1454	1463	[illegible]
18-5	1481	1486	1491	1496	1515	1545	1552	1562	1591	1593	1602	1616	1651	[illegible]	[illegible]
18-5	1712	1724	1744	1755	1775	1782	1792	1802	1812	1832	1852	1861	1881	1892	1902
18-5	1982	1992	2041	2081	2122	2211	2222	2226	2232	2233	2271	2295	2312	2322	[illegible]
18-5	2333	2338	2347	2382	2389	2403	2404	2412	2415	2416	2429	2444	2452	2456	[illegible]
18-5	2477	2479	2487												

K L	ITEMS AND RESPONSES														
22 1	225	652	726	733	756	771	786	811	831	886	926	966	976	996	1026
22 1	1565	1604	2287												
22 2	31	56	61	215	235	271	311	405	616	715	716	725	735	736	761
22 2	801	821	822	823	833	841	842	843	872	873	1006	1063	1074	1156	1375
22 2	1335	1403	1415	1441	1461	1522	1556	1645	1662	1731	1991	2231	2307	2326	2365
22 2	2366	2369	2375	2376	2392	2401	2411	2446	2448	2449					
22 3	11	12	21	22	23	32	92	106	112	113	124	125	126	255	294
22 3	301	302	303	342	371	425	475	491	501	522	534	541	606	702	714
22 3	745	802	812	813	832	861	884	891	925	986	1041	1065	1073	1135	1232
22 3	1256	1351	1366	1384	1394	1473	1502	1533	1575	1584	1595	1703	1781	1791	1801
22 3	1832	1851	1882	1932	1962	1981	2022	2102	2112	2122	2142	2152	2172	2181	2206
22 3	2252	2283	2284	2285	2286	2315	2367	2374	2381	2439	2445	2447	2483	2484	
22 4	24	46	86	91	114	123	132	234	321	553	596	603	613	615	631
22 4	641	691	746	762	796	803	824	871	885	906	924	931	995	1004	1005
22 4	1036	1064	1103	1121	1131	1132	1134	1161	1162	1163	1221	1231	1316	1333	1341
22 4	1374	1401	1414	1416	1423	1424	1425	1451	1474	1482	1511	1532	1542	1576	1585
22 4	1644	1661	1672	1704	1712	1741	1774	1812	1822	1841	1901	1911	2002	2012	2143
22 4	2162	2196	2212	2216	2224	2244	2288	2302	2317	2335	2344	2357	2364	2368	2377
22 4	2423	2438	2467	2469											
22 5	33	62	71	93	115	144	151	152	153	162	163	164	174	195	205
22 5	214	245	263	264	295	442	451	465	492	511	512	563	565	592	593
22 5	594	602	604	614	662	676	701	765	792	794	946	951	994	1012	1013
22 5	1014	1015	1016	1062	1065	1104	1122	1133	1144	1182	1193	1194	1201	1202	1222
22 5	1234	1272	1273	1291	1292	1293	1301	1302	1303	1365	1393	1402	1405	1432	1472
22 5	1475	1503	1523	1555	1566	1594	1596	1615	1633	1671	1682	1722	1763	1773	1862
22 5	1872	1891	1922	1941	1971	2032	2042	2062	2132	2161	2173	2192	2202	2213	2214
22 5	2223	2245	2261	2326	2334	2343	2391	2393	2424	2437	2454	2457	2466	2468	2482
22-1	26	121	306	601	611	651	654	712	722	731	781	782	815	835	845
22-1	875	881	922	991	1001	1061	1136	1471	1531	1602	2361	2441			
22-2	16	36	66	111	221	222	231	316	376	536	556	561	591	653	655
22-2	721	742	753	755	763	773	776	783	805	826	834	846	854	855	865
22-2	921	961	962	972	973	1075	1105	1126	1166	1236	1361	1392	1404	1411	1412
22-2	1521	1552	1562	1563	1581	1601	1663	1664	1702	2275	2319	2371	2372	2442	
22-3	52	53	96	155	165	211	212	223	232	291	314	346	401	402	496
22-3	504	525	656	704	705	706	711	741	752	754	764	765	766	774	775
22-3	791	806	816	836	864	874	882	963	971	1011	1021	1023	1071	1226	1336
22-3	1371	1372	1381	1382	1391	1406	1421	1443	1463	1501	1551	1564	1571	1572	1591
22-3	1631	1641	1665	1673	1674	1683	1776	2271	2272	2279	2299	2329	2362	2399	2403
22-3	2431	2463	2489												
22-4	54	75	252	253	275	305	366	424	524	526	543	555	636	646	684
22-4	693	694	713	723	743	751	814	866	893	933	974	1022	1024	1031	1042
22-4	1043	1196	1206	1253	1352	1353	1362	1376	1386	1442	1444	1446	1456	1461	1541
22-4	1553	1561	1582	1593	1642	1646	1666	1675	1676	1733	1734	1792	1831	1852	1992
22-4	2031	2041	2211	2251	2281	2291	2314	2341	2342	2383	2387	2394	2395	2396	2397
22-4	2398	2404	2405	2413	2414	2428	2429	2443	2451	2462	2464	2488			
22-5	35	42	51	55	83	84	85	101	102	103	105	131	171	185	191
22-5	192	193	213	254	262	265	273	274	313	315	336	345	352	385	395
22-5	403	422	423	434	444	454	462	471	503	505	513	514	515	516	521
22-5	544	545	546	576	581	621	644	666	673	674	692	695	732	804	825
22-5	844	853	856	876	892	903	934	942	964	975	982	992	1002	1046	1066
22-5	1081	1086	1115	1146	1151	1152	1153	1154	1155	1251	1252	1254	1276	1296	1306
22-5	1312	1314	1326	1344	1345	1346	1356	1413	1436	1445	1455	1462	1464	1465	1466
22-5	1494	1512	1513	1516	1573	1592	1616	1636	1653	1654	1655	1684	1686	1694	1716
22-5	1724	1735	1761	1771	1782	1802	1821	1842	1881	1931	1961	2001	2021	2061	2101
22-5	2111	2121	2165	2182	2201	2221	2233	2234	2235	2236	2273	2274	2295	2296	2321
22-5	2324	2331	2332	2338	2339	2347	2349	2351	2355	2382	2385	2386	2389	2406	2408
22-5	2412	2415	2416	2421	2426	2427	2432	2461	2471	2479	2486	2487			

Appendix D

Item Two of Project TALENT 1961 Follow-up Questionnaire

Have you attended college since leaving high school?

1. Yes, as a full-time student.
2. Yes, as a part-time student.
3. Yes, I entered but have dropped out temporarily.
4. Yes, I entered but dropped and do not plan to return.
5. No, but I plan to enter college within a year or two.
6. No, but I plan to enter college eventually; I have no idea when.
7. No, and I have no plans to do so.

APPENDIX D.--ERIC REPORT RESUME

ERIC REPORT RESUME

Resume date	02-20-68
Title	New Procedures For Scoring Psychological Measurements (Development of Moderated Scoring Keys for Psychological Inventories) Final report
Personal authors	Prediger, Dale J.
Institution (source)	University of Toledo, Toledo, Ohio, College of Educ.
Report/series no.	
Other source	
Other report no.	
Other source	
Other report no.	
Publ. date	20-Feb-68
Contract/grant number	OEC 3-7-070030-2871
Pagination, etc.	43 pages
Retrieval terms	Moderated scoring keys, inventory keying, empirical keying, scoring keys, moderator variables, item analysis, test scoring, biographical data, prediction of college attendance
Identifiers	Project TALENT Student Information Blank

Abstract

The three major project objectives were as follows: (1) Development of procedures for determining the optimum number of subgroups (and hence, moderated scoring keys) required for maximizing the predictive effectiveness of an inventory; (2) Development of a single scale for reporting the scores obtained from a set of moderated keys; and, (3) Determination of the accuracy of moderated scoring key predictions of college attendance as compared with predictions obtained from conventional keying techniques.

Basic data consisted of biographical inventory responses and academic aptitude test scores for approximately 20,500 high school senior boys. Scoring keys were formed for each of 12 ability level subgroups and various combinations of these subgroups. The keying procedures developed as the primary objective of the project work well when applied to actual data.

Although a statistically significant difference in favor of the moderated keys was obtained, academic ability was not found to be an effective moderator variable. However, a hit rate of 77 per cent was achieved by biographical data as a predictor of college attendance versus nonattendance. This rate and the equivalent point biserial correlation coefficient of .60 were substantially higher than the corresponding figures for academic aptitude used alone.