MULTIVARIATE STATISTICAL PREDICTION/CLASSIFICATION OF STUDENTS WITHIN INSTRUCTIONAL LEVELS IN SELECTED NINTH GRADE SUBJECTS: A COMPARISON OF THE RELATIVE EFFECTIVENESS OF THE MULTIPLE REGRESSION AND DISCRIMINANT MODELS

By

RICHARD A. LALIBERTE

Bachelor of Arts
St. John's University
Collegeville, Minnesota
1948

Master of Arts
University of Minnesota
Minneapolis, Minnesota
1951

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF EDUCATION
May, 1969

MULTIVARIATE STATISTICAL PREDICTION/CLASSIFICATION
OF STUDENTS WITHIN INSTRUCTIONAL LEVELS IN
SELECTED NINTH GRADE SUBJECTS: A
COMPARISON OF THE RELATIVE
EFFECTIVENESS OF THE
MULTIPLE REGRESSION
AND DISCRIMINANT
MODELS

Thesis Approved:

Wm P Ewens
Thesis Adviser

Dan Wesley

Bernard R. Belden

John D. Hampton

David Glenday

D. D. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

The writer wishes to acknowledge appreciation to the chairman of his advisory committee, Dr. W. P. Ewens, for the continuous encouragement he gave the writer during his graduate program and for the careful reading and helpful suggestions which he, along with the members of the committee, made concerning the dissertation. The members of the committee are: Dr. B. R. Belden, Dr. D. Glenday, Dr. J. D. Hampton and Dr. D. Wesley.

Appreciation is also extended to the administrative, guidance and teaching staffs of the Hinsdale, Illinois Public Schools for their interest in and encouragement to the writer in his investigative efforts.

The author also wishes to express his gratitude to Dr. C. Spearman and Dr. A. Perrelli, former Superintendent and Assistant Superintendent respectively, of the Hinsdale Public Schools, for their personal interest and support of the investigation, and to the members of the Boards of Education, Districts 181 and 86, for their financial support of the data processing phases of the study.

In addition, gratitude is extended to Dr. R. E. King, Assistant Director of the Applied Mathematics Division, Argonne National Laboratory, Argonne, Illinois for his personal interest in and support of the investigation from its earliest stages; to Mr. J. W. Butler, mathematician in the Applied Mathematics Division, Argonne Laboratory, for his personal interest in and support of the study and for the arrangements he made in securing subvention for the last stages of the computer processing; and to Mrs. Hestand and Mrs. Fu, former programmer statisticians in the Argonne Laboratory, Applied Mathematics Division, for the special care and interest they manifested during the various stages of the data processing.

A special note of thanks is due Dr. F. W. Simpson, Professor of Education at the University of Tulsa, who made available to the writer his personal copy of Morton's Study (1961) as well as provided guidance in the application in this investigation of his formula (Simpson, 1957) for the adjustment of the discriminant equation.

Special thanks are also due to Dr. W. Shephard, Dr. K. J. Vander Horck and Dr. H. E. Miller, the writer's present colleagues, who have made it possible through their encouragement and support to bring this dissertation to the final stages of completion.

Appreciation is also due to Mrs. Kay Nettleton who, with Mrs. Karen Lynch and Mrs. Diane Celarier, guided the study through the final typing stages with patience and good humor while under considerable pressure of time.

Finally, the greatest share of appreciation and thanks go to my wife, Pat, who, along with typing drafts, including the final draft, managed to sustain her husband with her encouragement and confidence and at the same time help our seven children understand why their mother and Dad were always so busy on the 'long assignment.'

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

Table		Page

Table		Page

Table		Page
XLVIII.	Summary of the Results of the Single Classification Chi Squared Test for Over-all Departure of Correct Predictions/Classifications by the Multiple Regression Equations from those Expected by Chance, Junior High Group Check Sample.	160
XLIX.	Summary of the Results of the Goodness-of-Fit Chi Squared Test of the Normality Assumption as Applied to the Distributions of ITBS and Lorge-Thorndike Scores Used in the Multiple Discriminant Analyses, Junior High Group Validation Sample	162
L.	Summary of the Results of Bartlett's Chi Squared Test of the Homogeneity of Within-Group Variance Assumption as Applied to the Distribution of ITBS and Lorge Thorndike Scores Used in the Multiple Discriminant Equations, Junior High Group Validation Sample	164
LI.	Comparison with the a Priori Proportions, the Proportions of Students Predicted/Classified Within Achievement Levels of Two Subject Areas by Means of Two Versions of the Multiple Discriminant Equations, Junior High Validation Sample.	165
LII.	Comparison with the a Priori Proportions, the Proportions of Students Predicted/Classified Within Instructional Levels of Two Subject Areas by Means of Two Versions of the Multiple Discriminant Equations, Junior High Group Check Sample.	166
LIII.	Summary of the Single Classification Chi Squared Test of Significance of Differences Between Proportions of Students Classified by Means of Discriminant Equations and the Expected a Priori Proportions, at each Instructional Level, Junior High Group Validation and Check Samples	168
LIV.	Summary of the Comparisons with Total Chance Expectations of the Total Number of Students Correctly Predicted/Classified (Hits) in Two Subject Areas by Two Versions of the Multiple Discriminant Equations, Junior High Group Check Sample.	172
LV.	Summary of the Results of the Single Classification Chi Squared Test for Departure of Over-All Correct Predictions/Classifications by the Multiple Regression and by Two Versions of the Multiple Discriminant Equations from Those Expected by Chance, Junior High Group Check Sample.	173

Table		Page

CHAPTER I

INTRODUCTION

For the past fifty years in American secondary education, the administrative procedure for programming students into levels, tracks or groups within required or elective courses has been on the basis of previous grades, test scores, entrance tests, teacher and counselor recommendations, and frequently, student and parent choice. This procedure, although the subject of controversy and much inconclusive research concerning its merits (Eckstrom, 1961), is nevertheless used with increasing frequency in comprehensive schools large enough to support such programs (Anderson and Van Dyke, 1963). As an indication of the increasing trend of grouping, one study in 1955 (Crawford, 1955) found that 37% of the schools in five midwestern states were using some form of ability grouping. Another study four years later, conducted in four of these states (Van Dyke and Sparks, 1959), found 68% of the schools were employing some form of grouping. In a recent study of grouping practices in California's secondary schools (Thomas, 1966), based on an 86% sample of the state's school districts, more than 84% of the English classes in the sample studied followed an ability-group scheme compared with 57% of social studies classes, 47% of mathematics classes and 21% of the science classes.

The person who played the most prominent role in assigning students to their sections, according to the California study, was the counselor (in 50% of the schools). In 26% of the schools, a counselor and teacher worked together to determine students' assignments. The kinds of information used to determine the section or the instructional level in English to which the students were assigned varied widely but the most common combination of data, used by 33% of the schools, consisted of the prior teacher's recommendations, test scores (particularly reading subtests from batteries), and the last year's marks in English.

Perhaps in recognition of this widespread use of test and other data for, among other purposes, the tracking of students, participants at a conference on measurement and research (Traxler, ed., 1961) pleaded that counselors and teachers be taught how to interpret test scores in relation to all other data available on the same student. This is recognized as a noble goal but it is contended by Cooley (1964) that it is an unrealistic one since, even if there were available the follow-up data which would make predictive interpretation possible, people are just not able to process that much information reliably.

Even supposing that counselors were able to process all the information available on the same students as reliably as, for example, a computer-based prediction/classification system, the question remains whether or not this is the most efficient use of the counselor's time, given the shortage of school counselors that exists now and the even greater shortage likely for the projected school enrollments (Cooley, 1964). Although the question of the counselor's proper functions is the focus of much discussion and controversy (Fullmer, D. W. and Bernard, H. W., 1964), the latest policy statement by the American School Counselor's Association lists ten functions judged to be what should obtain, second among which appears the provision of the placement and appraisal services for students.

Typically, the placement or grouping process is as follows: the students may be sectioned into a level or group on the bases of an IQ score, the number of levels being determined partly on the basis of arbitrary two-or-more-level grouping or tracking. In some cases the average of achievement test score(s) which is thought to be relevant to the course level into which students are to be sectioned is used. In still other instances, the rather extensive amount of test data that may be available may be disregarded as irrelevant; in place of available data, the administration of additional test batteries is requested. In addition to or instead of one or more of the above procedures, the previous teachers' recommendations may be incorporated in the placement-decision process. The counselor and/or other school personnel processing any part or all of the above data components subjectively weights the information according to his set, preferences, biases. Seldom, if ever, is the data-collection,differential weighting, and placement process objectively validated against criterion performance. Validation of the subjective, 'clinical', variety, perhaps even of the self-fulfilling prophecy variety, undoubtedly occurs.

The methods used most often to validate objectively, when it is attempted, the above process are the zero order correlation and multiple correlation statistical procedures. As will be noted in Chapter II, REVIEW OF THE LITERATURE, the multiple regression model is less frequently used below the college level. These procedures have as their basic task to indicate whether or not positive, zero, or negative statistical relationships exist among the independent variables and, in the case of regression analysis, between independent or predictor variables and a criterion variable. This criterion variable is usually an overall grade point average (GPA) across subject areas, seldom the GPA within subject areas.

Cooley (1964) indicates that measurement specialists have for some time sought to educate counselors and other school personnel to the idea that a subtest score on an achievement battery, or one IQ score by itself, or any other single datum supplies insufficient information to infer much of anything about a student. Instead, these measurement specialists began promoting the people approach, dubbed the "parallel stalks model", where each stalk stands for a test and a line drawn from stalk to stalk represents a student's profile or score combination. This procedure gives the impression that all the most recent test score information available on a student is being considered simultaneously. Such an impression may be only an illusion. While the profile may be representative of all the most recent information, consideration of the profile simultaneously for even one particular prupose demands a considerable amount of information about the meaning of the test scores in combination. That is, a student may have high scores on tests of verbal aptitude, reading comprehension and computational skills but low scores on tests of reference skills and spelling, among others. In reviewing these scores in profiled form, the counselor might ignore the low scores and assign the student to a 'fast' section of mathematics. The counselor may have made a correct decision (the student achieved well) but may have erred in the case of another student. Thus, profile representation may be a source of disservice to a student since peaks and valleys in the profile may cause the attention of one looking at the profile to be glued to a peak or valley and, therefore, to over-interpret. Additionally, a profile of current data may obscure the relevance of longitudinal data--either it is not conveniently available or the task of interpreting the profiles obtained over time may demand more time and energy from the observer than is available.

There seems to be a need for a summary of all current and previously secured data which are relevant to particular educational and career plans. Cooley (1964) has described in some detail a multivariate procedure for summarizing these data by a computer-measurement system. Based on a factor conceptualization of human behavior, Cooley points out that personality has its locus in an _m_-dimensional space. That is, an individual's personality has its unique location in this space, the location predicted by the total pattern of _m_ behavioral measures (e.g. test scores) which are available for that individual. People who have similar patterns of test scores, for example, will occupy similar regions of this _m_-dimensional space. Therefore, people with similar interests, achievements, aptitudes, and personalities tend to make similar types of career plans, or have similar achievement test score patterns and succeed in class work, etc., in like fashion (Cooley, 1964). Once the regions of the personality space occupied by people who have made particular types of career plans or have achieved within particular tracks or groups have been defined, the probability that another person will make like career plans or have similar test score patterns can be estimated. Essentially, this approach, which is the multiple discriminant model approach, seeks to answer the question: to which group of people is an individual, with a given matrix of interests, achievement and aptitude test scores, most similar.

Need For The Study

As will be discussed in Chapter II, REVIEW OF THE LITERATURE, a number of studies have been conducted in which the correlation/regression and/or discriminant models have been employed for predicting academic success, but to the writer's knowledge, no study using pre-college samples has yet been conducted in which the regression and discriminant models have been compared for their efficiency in predicting academic success. This study represents an effort to fill that need. It has been structured on the need to investigate ways in which to classify students on the high school level into instructional groups such that two goals are optimally achieved: 1) that the student will be successful (as reflected by grades received) in terms of the achievement goals set for the group and 2) that the student will be 'fully challenged' by his participation in a psychometrically-formed group (i.e., the content and/or pace of instruction is geared to the presumed talents of the group numbers). Because the grade point average criterion is built into the regression model, the first goal is a regression-type task; because the homogenous proportioning is built into the discriminant model, the second goal is a discriminant-type task; yet it is to the detriment of the student to separate these two goals. If there are to be instructional levels, a predicted success measure (e.g., grades) is meaningful only in relation to a particular level; a grade in one instructional level is not directly comparable to a grade in another level. Conversely, classifying students into an homogenous group without including a success component, such as grade-getting motivation, might result in under-or over-challenging some students. Realistically these two goals must be considered simultaneously and it is the purpose of this study to employ two statistical prediction/classification models and to evaluate each model's performance in carrying out this dual role.

Statement of the Problem

Briefly, the major problem on which this investigation focuses is to study, within a framework of a computer-based measurement system ways in which to effectively group students into instructional levels. The dimensions of the problem are analyzed in four phases as follows:

1. The first phase studies the predictive validity of the relatively extensive number of test variables and a limited number of non-test variables available on the target population for estimating first semester ninth grade averages (GPA) within instructional levels of four subject areas using the multiple correlation/regression model.
2. In the second phase of the problem under study, the investigation focuses on the effectiveness of the multiple correlation/regression model for grouping (predicting/classifying) students within instructional levels in two subject areas against a priori and chance expectations.
3. The third phase is concerned with studying the effectiveness of two versions of the multiple discriminant model in grouping (predicting/classifying) students against a priori and chance expectations.
4. In the fourth phase, this investigation centers on a comparison of the relative effectiveness, against various criteria of the multiple correlation/regression model and two versions of the multiple discriminant model for grouping (predicting/classifying) students within instructional levels of two subject areas.

Hypotheses

The hypotheses to be tested in this study have been arranged into seven sets of major hypotheses, together with certain minor hypotheses. In general, Major and Minor Hypotheses I are concerned with testing the relative efficiency of the multiple correlation/regression equations on validation samples while Major Hypothesis II explores how the equations perform on the check samples.

Major and Minor Hypotheses III and IV are concerned with the assessment of the relative effectiveness of the multiple correlation/regression and two versions of multiple discriminant equations in performing a similar task, that of predicting/classifying students against a priori (see page 14 for definition) and chance expectations.

Major Hypotheses V, VI, and VII compare the relative efficiency of the multiple correlation/regression and discriminant equations in correctly predicting/classifying students within instructional levels of two subject areas.

All the hypotheses are stated in the null form. They are listed again in Chapter III, METHOD, along with the statistical procedures appropriate for testing them.

Major Hypothesis I

Within the combined and junior high validation samples, there is no statistically significant reduction in the error sums of squares as a function of the predictor variable entered at each step to a limit of six steps of the step-wise regression routine.

Minor Hypothesis I

Within instructional levels of selected subject areas of the junior high validation sample, there are no significant differences in the multiple correlation coefficients (multiple R's) at the sixth step of the step-wise multiple regression routines as a function of different sets of independent variables drawn from additional data sources.

Major Hypothesis II

Within instructional levels of selected subject areas there are no significant differences between the validation and check sample multiple R's of the combined and junior high groups.

Major Hypothesis III

There are no significant differences in the proportion of the junior high group check sample students predicted/classified within levels of two subject areas (English and mathematics) by means of multiple regression equations and the proportions expected a priori.

Minor Hypothesis III

The over-all proportion of junior high group check sample students correctly predicted/classified in two subject areas (English and mathematics) by means of multiple regression equations does not differ significantly from the proportion expected based upon the operation of chance.[1]

---

[1]This hypothesis is classified as minor in agreement with Cronbach and Gleser (1957, p. 31) who point out that when the a priori population is defined, as in this study, chance selection is not necessarily the alternative strategy with which prediction/classification procedures based upon test and other quantifiable data are to be compared. Cronbach and Gleser point out that the decision makers (counselor et al.) who may not base their decisions on statistically processed test and other data may use conferences with students and/or other information available to provide some basis for a decision. Nevertheless, as discussed previously in this chapter, since other information available is likely to be used by the same decision maker or others in classifying the same student in an unsystematic and, therefore, variable way, chance factors are operating. Thus, the hypothesis concerning the correctness of prediction/classification based on the operation of chance does merit some attention and so its status here as a minor hypothesis.

Major Hypothesis IV

The proportion of junior high group check sample students predicted/classified within each of three levels of two subject areas (English and mathematics) by means of two versions of multiple discriminant equations does not differ significantly from the proportion expected a priori.

Minor Hypothesis IV

The over-all proportion of the junior high group check sample students correctly predicted/classified in two subject areas (English and mathematics) by means of two versions of multiple discriminant equations does not differ significantly from the proportion expected based on the operation of chance.

Major Hypothesis V

There are no significant differences between the over-all number of junior high group check sample students predicted/classified in two subject areas (English and mathematics) by each statistical method and the total number succeeding in the groups in which they were registered.

Major Hypothesis VI

There are no significant differences between subject areas in the over-all number of junior high group check sample students correctly predicted/classified by each statistical method.

Major Hypothesis VII

There are no significant differences between statistical methods in the over-all number of junior high group check sample students correctly predicted/classified in two subject areas.

Limitations of Study

Concerning Sample Size

It will be observed in Chapter IV, RESULTS, that within certain subject area levels, the number of students comprising the sub-samples (i.e., instructional levels within subject areas) is quite small in both the validation and check samples. This limitation could have been alleviated to some extent by combining the validation and check samples. However, this modifcation of design would have precluded the highly important strategy of evaluating the regression and discriminant equations using new samples.

Concerning the Variables

It would have been interesting and possibly of predictive relevance to have employed other test and non-test variables in addition to or instead of those which were used (e.g., socio-economic status, motivation, divergent thinking, etc.). However, it was not the purpose of the study to add variables to an already extensive pool of data but rather, as stated previously, to analyze the data pool to find the most efficient set of predictive and classificatory variables.

The variables employed in this study, with one exception, were entered as input to the statistical models in the form in which they were available. Thus they were not represented by a common standardized scale. One consideration that contributed to the decision to use them 'as they were' was the desire to provide the school personnel with statistical procedures that could be utilized without the necessity of first converting the data to another scale. It was judged that the more the data had to be modified before they could be used with the procedures of this study, the less likely that they would be used.

Nevertheless, it is recognized that using data in their 'as is' form sets limits on the generalizability of the findings to other populations.

Concerning Prediction/Classification Models

One model not directly investigated in this study might be termed the parent-pupil-teacher-counselor (pupil et al.) model, or briefly, the quasi-clinical model.[1] Two factors precluded the use of this model as a definitive element: 1) the complexity of the model, since the section or level in which a student was scheduled was a function, in unknown ways, of the pupil et al. decision; 2) even if it were feasible to sort out the elements in this model, constraints of time availability prevented this investigator from studying this broadly-defined model in relation to the two statistical models employed in this study. Perhaps this quasi-clinical model may be judged to have been studied in selected subject areas inferentially, since the groups of students as scheduled into each level within subject areas can be assumed to have been formed as a result of the operation of this model. In studying the efficiency of the regression and discriminant models, reference is made to the groups constituted by the quasi-clinical model, defined in this study as the a priori proportion.

Concerning Longitudinal Data

A complete computer-based academic prediction model would include test and non-test data from the students' earlier years in school. The validity of such data for prediction/classification purposes would need

---

[1]Quasi-clinical as compared with the clinical model which is more rigorously defined, for example, by Meehl (1954).

to be assessed. The task of marshaling such a longitudinal data record would be sizeable indeed. The feasibility of such a longitudinal record would depend upon many factors, not the least of which is ready access to computer facilities for storing, retrieving and updating the data. The value of including longitudinal data as a part of the data source for a prediction/classification scheme is recognized by the investigator; however, the inclusion of all such data was beyond the scope of this study.

Concerning the Sex Factor

The development of prediction equations separately for boys and girls is recognized as a desirable procedure and perhaps requisite in some settings. Separate equations for boys and girls were not developed in this study for two reasons: 1) the sample size for some instructional levels in some subject areas was too small for further breakdown; 2) the processing costs would have been nearly doubled. Nevertheless, the lack of separate equations for boys and girls is acknowledged as a limitation.

Concerning the Sample Characteristics

For a number of reasons (sample size, possible uniqueness of curricular offerings and grading practices to mention a few), generalizations of the findings beyond the target population of this study must be made with considerable caution. To be sure, one of the reasons for caution resides in the nature of the target population itself where the mean values on academic ability measures tend to hover around the eighty-fourth centile (national norms). Such a finding is not unusual in suburban communities surrounding large metropolitan areas. To some extent the finding of this study, if generalized, would be limited to student populations possessing similar characteristics.

Definition of Terms

The following is a brief list of words and/or phrases used in this study which are either not too frequently found in the literature or have been found to be used interchangeably with their synonyms, or which are defined here for purposes of this study.

1. Actuarial: see statistical prediction method.
2. A priori proportion: the proportion of the total junior sample assigned by school staff to instructional levels of a subject area.
3. Characteristic score: the score made by the hypothetical typical members of a group on the discriminant equation.
4. Check sample: the new sample (in this study, the high school class of 1967) on whom the regression and discriminant equations, developed in the validation sample, are applied. This definition follows that employed by Ludlow (1962) and Simpson (1957). Also referred to in the literature as the cross-validation sample (Hampton, J. D., personal communication).
5. Classification equation (adjusted): the classification equation whose constant term has been adjusted (modified) to permit classification in proportion to the original (a priori) probabilities.
6. Classification equation (basic): the unadjusted equation resulting from the application of the discriminant function to a set of test and non-test score variables in the validation sample.

7. Correctly predicted: in this study, defined to mean that the junior high group check sample students are correctly assigned in the ninth grade by means of the regression and discriminant equations to the level within English and mathematics in which they are enrolled and succeeding (i.e., their first semester GPA in that level of the subject area is 1.5 or greater when 1 = F; 5 = A).
8. Grade point average (GPA): the simple arithmetic mean of each student's grades; defined numerically (A = 5, F = 1) within each subject during each semester.
9. Grouping: see predicted/classified.
10. Hits/misses: see correctly predicted.
11. Instructional level: that level within a subject area into which students are grouped by the quasi-clinical method based upon presumed ability to achieve in that level.
12. Multiple discriminant function analysis (also, discriminant analysis, dispersion analysis): a method of determining statistically in which one of three or more groups a student seems to belong in terms of his similarity (in test scores or other measures) to others of that group who have successfully achieved in the academic tasks set for the group.
13. Predicted/classified: the procedure of assigning students to instructional levels within subject areas by statistical methods.
14. Quasi-clinical: method of assigning students based upon the subjective weighting of test and/or non-test variables by one or more school personnel.

15. Statistical prediction method: the method as employed in this study of predicting/classifying students into groups through use of the multiple regression and multiple discriminant models; used interchangeably with acturial.

16. Validation sample: comprised of the high school class of 1966 (graduation year) upon which the statistical prediction procedures were developed. (See check sample for references). Also referred to as the development sample (Hampton, J. D., personal communication).

CHAPTER II

A REVIEW OF THE LITERATURE

The review of the literature will begin with a brief introduction of the varieties of techniques, and the rationale behind them, for selecting, classifying, and/or predicting academic achievement. There will follow a brief review of surveys by other authors of studies concerned with the prediction of general academic achievement below the college age level.

Next will be presented a review of selected studies of a) general academic achievement below the college age level which employed the correlation and/or regression model; b) general and/or subject area academic success below the college age also employing the correlation and/or regression models; and c) prediction of academic success below college age level employing a discriminant model.

Finally, the review will present studies concerned with comparing techniques (models) for predicting academic success or for studying ways of classifying students into curricular groupings.

Varieties of Prediction/Classification Techniques

Although presented in a broad context of statistical methods of educational research, Hoyt and Johnson (1954) provide what they term as the first discussion, for the Review of Educational Research, of regression and correlation as a separate topic. A few years earlier, Travers (1949) had reviewed more than twenty years of ever-increasing activity in the prediction of academic success and while Travers (1939) himself had introduced to the American literature another technique, discriminant function, for the prediction task, he observed no use of it up to 1949. Two years later, Rulon and his colleagues (1951) introduced the multiple discciminant function to American researchers, a statistical procedure described by Nunnally (1967) as one that is employed when three or more groups of persons are defined a priori and the purpose of the analysis is to differentiate the groups from each other on the basis of the score profiles of the group members.

Tatsuoka and Tiedman (1954) attributed the virtual disuse of the discriminant technique to the negative attitudes of some psychologists and educators toward its use (e.g. Garrett, 1943; Wherry, 1947). Wherry's negative influence, at least, can be noted in Guilford and Michael's (1949) publication dealing with predicting categories from measurements. Wherry's influence on Guilford persisted even to the fourth edition of his well-known text on statistics (Guilford, 1965) although Guilford did choose to use the discriminant model in a recent prediction study (Guilford, 1965) reviewed below. Tatsuoka (1957) was quite critical of the rather negative treatment given the discriminant model by French (1955). However, in fairness to French, it should be indicated that he sought to map out the areas of application which he judged to be appropriate for the discriminant and regression models. Perhaps on the basis of space allotment in his article, French seemed to favor the regression model as the most generally appropriate one for the problem of differential prediction.

Concerning the discriminant model, Helmstadter (1964, pp. 216-217) observed:

> At first glance, this technique appears to be ideal. . . While under certain circumstances, this procedure is one of the most satisfactory, there are a number of disadvantages which make it less useful than often supposed.

As disadvantages, Helmstadter points out that when aptitude measures comprise the profile variables there is no information as to whether those aptitudes which optimally distinguish among groups are essential for success; further, Helmstadter contends that no evidence is yielded, once a person has been classified, concerning the degree to which he is likely to be successful. Helmstadter stresses that it is only assumed by those who employ the discriminant analysis for personnel classification purposes that if an individual is in an occupation with people like himself, he will be successful. The discriminant technique, Helmstadter maintains, does not assist individuals in being placed in any more appropriate job(s) than are in the criterion groups, a situation which merely perpetuates the successes and mistakes that have been made in the trial and error process that led to the formation of the criterion groups. Helmstadter ends his critque of the discriminant model on a positive note by stating that

> . . .the multiple discriminant function provides us with the best possible answer to the classification problem when no continuous criterion measure (e.g. GPA's in the school setting) of the degree of success is available. Thus, if the only criterion data available is grouping (e.g., trichotomizing) by occupations, by academic areas, by socioeconomic levels. . .and so forth, then the multiple discriminant function provides the appropriate technique to use.

A review of Helmstadter's own research (1957) concerned with comparing the discriminant with other statistical and non-statistical methods in a classification task is presented later (page 34).

More recently, Nunnally (1967, pp. 399-400) states:

> The wisdom of applying discriminatory analysis depends on the problem. Potentially, discriminatory analysis is most useful in applied psychology, where, for example, it might be used in assigning persons to jobs, students to courses of training, and patients to diagnostic groups. There are, however, some logical difficulties in employing discriminatory analysis for that purpose. One logical difficulty is in deciding how to designate the members of groups prior to performing discriminatory analysis. For example, in discriminating different professional groups on the basis of score profiles, should the group of engineers include all engineers, only highly successful engineers, or some model group with regard to success? As another logical problem, in employing discriminatory analysis in applied psychology, it usually is assumed that all persons will be assigned to one of the groups, which is a poor strategy. In addition to these and other logical difficulties in employ- discriminatory analysis in applied psychology, it simply has not worked very well in studies to date. The amount of overlap between groups in the discriminant space tends to overshadow the separation between groups. It is wishful thinking to hope that all engineers will differ markedly from all physicians, or that all schizophrenics will differ markedly from all neurotics, on any collection of variables.

Nunnaly goes on to delineate the basic psychology areas in which he judges the discriminant model to be relevant. In essence he contends that this model's greatest contribution is in understanding the major differences between groups than it is for classifying people into groups.

Quite at variance with the lack of enthusiasm for the discriminant model that pervades the statements of the above authors is the strong support given to the model by Rulon and associates (1967) for its applicability to the same tasks for which it was not recommended by Helmstadter (1964) and Nunnally (1967). Earlier, in a symposium presented at the 1950 American Psychological Association convention, and published a year later (Tiedman et al., 1951) the applicability of the discriminant model to educational and psychological areas of investigation was presented by Tiedman. Rulon (1951) elaborated on the distinctions between the regression and discriminant models in terms of the prediction/classification questions being asked. Bryan (1951) presented computational routines for the multiple (i.e., more than two groups) discriminant model.

Pickrel (1958) observed that a comprehensive review of classification problems and techniques for handling them was not available in the literature up to that date. He proceeded to discuss the rationale behind such procedures as selection, classification and differential prediction, pointing out that the term selection is appropriately applied to the single job (or admission/non-admission; pass/fail; etc.) categroy while the term classification applies to two or more job (curricula, levels within curricula) categories. Pickrel continued with a brief discussion of the discriminant, regression, multiple cutoff, unique pattern and factor analytic models.

In a more penetrating analysis of the rationale lying behind these models, Tatsuoka (1956) categorized differential prediction and classification procedures into two groups: 1) the regression model and 2) the profile similarity models. The latter group includes intuitive approaches which leave it to the subjective judgment of the guidance person or personnel worker to determine which of several group profiles a given individual's profile most closely resembles. Tatsuoka points out that attempts to provide more objective methods for judging profile similarity led to the development, among others, of the distance function which, according to Tatsuoka, has the basic defect of forcing a multi-dimensional phenomenon (Cooley's /1964/ M-dimensional space referred to in Chapter I, page 5) into a uni-dimensional framework (i.e., a summation of the square of the differences of each pair of variables in the data matrix). Tatsuoka indicates that the multiple discriminant model represents a culmination of the class of methods concerned with profile-similarity.

With this brief review of the varieties of techniques available for predicting and classifying individuals against some criteria (e.g., GPA, a priori instruction levels, fields of concentration), attention will now be given to a review of surveys by other investigators concerned with the prediction of general academic achievement below the college level.

Surveys By Other Investigators of Studies Concerned With the Prediction of General Academic Achievement at the High School Level

McLaughlin (1950) reported in his review of the literature that there were only three published studies concerned with predicting success in high school from information obtained prior to high school entrance.

Owen (1956) cited Travers (1949) as noting that the prediction of success in high school has been of greater concern in Europe than in America. Owen didn't speculate on a possible explanation for this greater concern in Europe but it is probably readily explained by the more prevalent administrative provision for streaming, as grouping or tracking is referred to in Europe, principally in Britain and other countries under the Crown. In her review, Owen (1956) cited only one study, Layton's (1954), relevant to prediction of success in high school, a study that was concerned with predicting success at the twelfth grade from ninth grade predictor variables.

Scannell (1958) in his review of the literature cited two studies, one of them an unpublished one by Fahnle (1942) which was a master's thesis completed at the State University of Iowa; and the other, a published one, Wellman's (1957). Owen's own dissertation which was concerned with predicting success at the high school level was not cited by Scannell. Lavin (1965) included in his 1953-1961 survey some nine studies not cited above which employed intellective factors as predictors of academic achievement.

The above listing of surveys of studies are representative reviews of studies in which the prediction of general high school academic achievement was the focus. In the group of studies reviewed next, specific attention will be given to the findings reported by investigators who studied the academic success prediction problem below the college level.

Selected Prediction Studies of Academic Achievement Below the College Level

Studies of General Academic Achievement Employing the Correlation and/or Regression Model

Layton (1954) found zero order correlations of .63 to .82 when he correlated ninth grade scholastic aptitude and English achievement test variables with similar variables and centile rank in class obtained (on these students) in the twelfth grade.

Fahnle (1942) compared eighth and ninth grade Iowa Every Pupil Tests of Basic Skills subtests and found that the intercorrelations between subtests bearing the same name were, not surprisingly, the highest. He further observed that the eighth grade Work Study Skills subtest correlated highly with all the ninth grade subtests.

In his study, McLaughlin (1950) investigated the relationships between subtests on the eighth grade Iowa Tests of Basic Skills (ITBS) and the twelfth grade Iowa Tests of Educational Development (ITED). The zero order correlation between the ITBS and ITED Composite scores was .80; among the subtests of the two batteries, the zero order r's ranged from .61 to .76; the correlation between the eighth grade ITBS Composite and four-year GPA was .59.

Scannell (1958), as a part of his study, obtained inter-correlations between grades four, six, eight ITBS and grade twelve ITED Composities which ranged from .72 to .78. When he correlated the ITBS Composities with high school GPA, he found zero order r's ranging from .53 (grade four) to .61 (grade eight).

When the criterion of academic success was an end-of-ninth-grade GPA computed on a varying number and kinds of courses taken, Gibbons (1962) concluded that the total score on either or both of the eighth grade scholastic aptitude or the achievement tests were as effective or more effective predictors as any one or both of the non-cognitive variables (study methods, personality questionnaire).

Jacobs (1959) concluded from his study of predicting general academic success that an arithmetic proficiency test was the best single predictor when the criterion was a three or four year GPA.

A few generalizations concerning the studies cited above which dealt with the task of predicting general academic achievement of high school students are these:

1. The predictor variables employed were largely of the scholastic aptitude and achievement variety.
2. The zero order correlation model was the preferred technique.

3. Global types of criterion variables such as total high school GPA or GPA within a grade level are attractive to use because of the presumed greater stability associated with them. It easier to secure a respectable sample size using GPA over all subject areas within and/or across grade levels. The information yielded in these studies has some relevance to decision-making processes in the schools; however, since the prediction task at hand is usually to estimate likely success within specified subject areas, review of studies with subject area focus is presented in the next section.

Studies Predicting General and/or Subject Area Academic Success Below the College Level

Wellman (1957) using combined ninth and tenth grade GPA as one set of criteria, found correlations of .82 and .83 when the Otis and Primary Mental Ability Test (PMA) total or the Otis and PMA factor scores respectively, were the predictor variables. When the predictor variables were the Otis Intelligence Test and two PMA subtests (Verbal Meaning and Number Scores) and the criterion variables were English and science GPA, multiple R's of .81 with each criterion were obtained. When the predictor variables were Otis, PMA Space and Number factors, the multiple R with mathematics GPA was .76.

Foreign Lanuages

Pimsleur et al. (1962), studying the differential predictability of traditional (grammar-reading) and more recent concepts of achievement (oral-aural) in first year French, by means of a step-wise regression procedure, obtained a multiple correlation coefficient of .65 using six tests of Cooperative French Test battery; a multiple R of .41 using five tests to predict aural comprehension; and a multiple R of .41 using five tests to predict aural comprehension; and a multiple R of .41 using five tests to predict speaking proficiency. Among the most durable variables predicting the various criteria was a measure of verbal I.Q.

Carroll (1962), reviewing his own and other studies of the prediction of success in intensive language courses, stated, among the conclusions reached, the following:

1. Foreign language learning aptitude is not specific to particular languages or particular groups of languages--the same battery of tests predict success in languages as diverse as German and Chinese with approximately the same degree of validity.
2. Language aptitude as measured by tests seems to consist of at least four identifyable abilities: a) phonetic coding; b) grammatical sensitivity; c) rote memorization; d) the ability to infer linguistic forms, rules, etc., from new linguistic content.
3. A relatively small fraction of the general population, perhaps one-third to one-half, has a good chance of success (achieving satisfactory grades) in these courses--practically the full range of the general population in regular school classes can succeed in less intensively paced foreign language courses.

Hascal (1959), using the simple zero order correlation model, set out to study the validities of several variables for predicting, among various criteria, success in the study of foreign languages. Of the several conclusions reached, the following are relevant:

1. The relative predictive validity of the several variables varies as a function of sex.
2. Variables tended to demonstrate greater predictive validity for boys.
3. The best predictors for both sexes with the criteria used (GPA in specific foreign language and end-of-year Cooperative Foreign Language test scores) were previous English GPA and certain Differential Aptitude Test (DAT) subtests (especially the Sentence subtest) and Stanford achievement scores. Other predictor variables such as a measure of academic aptitude, another achievement test battery, and Kuder interest test scores were recommended by Hascall (1959) for elimination as predictor variables.

Mathematics

Guilford, et al., (1965) using his structure-of-the-intellect model as a referent, investigated the construct, predictive, and classificatory validities of some forty-four variables, thirty-four of which grew out of Guilford and associates' own earlier research on the model. The criterion measures were GPA and standardized mathematics achievement levels of test scores in four 9th grade mathematics groups (basic mathematics, non-college algebra, regular algebra, accelerated algebra).

In the context of this study, Guilford's group used factor analysis, step-wise regression and two-group discriminant analytic techniques to study the operations of the several variables for predicting against test criterion (step-wise regression model) and classifying students against their group membership (discriminant model). Here he concluded that the same test battery could not be best used for doing the double duty of classifying students and also predicting achievement in the two kinds of courses. As a result of this investigation, Guilford and his associates reached the following of a number of general conclusions:

1. Batteries of factor scores were better predictors of achievement than two of the standard-test combinations, significantly so in Algebra.
2. A composite of 13 factor-test scores gave increased prediction when added to each of the three standardized test combinations significantly so in the algebra courses.
3. Combinations of factor-test scores discriminated between successful (above-median) algebra students and general mathematics students with an accuracy close to 90%.

Dinkel (1959) found a multiple R of .86 between algebra achievement and a series of seventh and eighth grade predictor variables, including previous grades, intelligence and prognostic and achievement tests. Barnes and Asher (1962) found that, out of eleven variables studied for predicting ninth grade GPA, the best single predictor was the eighth grade GPA in mathematics. The only other variable that significantly increased the multiple R was an eighth grade arithmetic achievement test.

Wellman (1957) obtained an optimally significant multiple R of .76 using three of six predictor variables (Otis, PMA Space and Number factor scores) and ninth and tenth grade mathematics GPA's (algebra and plane geometry).

English

Diederich (1957) reported studies which indicated that the vocabulary sections of intelligence tests are highly relevant to the prediction of composition skill.

Wellman (1957) obtained an optimally significant multiple R of .81 between three of six predictor variables (Otis, PMA Verbal-meaning and Number scores) and English grade nine and ten GPA's, respectively.

In summary, the above review of studies of estimating academic success in general and/or selected subject areas appear to yield the following generalizations:

With the exception of the studies by Pimsleur et al. and Guilford and his associates, the zero order and multiple correlation models are the ones most typically employed. The studies by the Pimsleur and Guilford groups employed the multiple regression and multiple discriminant models. Guilford's group utilized the regression and discriminant models sequentially. They were not concerned with comparing the relative efficiency of either model in their study.

Carroll's statement that the same battery of tests predict success in a variety of languages should be noted in view of the present writer's findings in Chapter IV, RESULTS, pages 76 ff., relative to the kinds of predictor variables that appear to be associated with success in various foreign languages.

Guilford's study is impressive in its extensive exploration of the domain of predictor variables. It represents a basic research effort out of which there might hopefully emerge, in time, new sets of independent variables that might have practical relevance.

In the next section, the three studies reviewed will be those that have explored the efficiency of the discriminant or the regression model for predicting/classifying or grouping pre-college samples of students within subject areas or grade levels.

Studies of the Prediction/Classification of Students Within Grade or Instructional Groupings by Means of the Multiple Correlation/ Regression or the Discriminant Model

Leton and Anderson (1964) focused on the application of the discriminant model for the grouping of students, grades 4-12, on the basis of achievement characteristics as compared with the conventional age-grade administrative arrangements. Random samples of 150 sutdents from each grade level comprised the study populations. The variables used in developing the classification equations were the appropriate levels (i.e., elementary, etc.) of the 1957 edition of the California Achievement subtests. The validation groups consisted of three grade-sets: a) 4, 5, 6; b) 7, 8, 9; c) 10, 11, 12. The classification equations were developed for each of the three sets. The efficiency of the classifications was evaluated in relation to each of the age-grade sets. The model's efficiency was generally most apparent in the elementary (4-6) and junior high (7-9) sets--correctly placing about two-thirds of the students at grades four and six and seven and nine. Fifty percent correct classifications were made for students in grades five and eight. At the tenth and twelfth grades, 50% of the students were correctly placed while at the eleventh grade classification, in terms of the age-grade criterion, was at the chance level. The discriminant equations were not evaluated on either cross-validation or check samples.

Morton (1961) studied the feasibility of employing Simpson's adjustment (1957, see page 36 of this REVIEW for a brief review of Simpson's study) of the discriminant function equations utilizing I.Q. (Otis) and Composite ITED scores to classify tenth grade students into three homogenous groups; 22.5% high, 55% medium and 22.5% low.

The validation sample criterion groupings (a priori proportions) were based on the pooled judgments of English and history teachers. The classification equations developed on the validation sample were then applied to a check sample, yielding a total 90% correct placement over the three groups.

Schusler (1964) employed the step-wise multiple regression procedure to study the predictive/classificatory validity of 30 variables for estimating end-of-year GPA's of 342 sophomores in English, mathematics, social studies and foreign languages. Letter grades for courses of different levels of difficulty were converted to a common numerical scale by the method of equal-appearing intervals, a procedure involving the subjective judgment of the person(s) performing the conversion. Substantial correlations were obtained between actual and estimated GPA's as converted.

In a separate phase of her study, Schusler scheduled 10th grade students into either the high or low level English course based on regression equations developed on still another sample. Again, substantial correlations were obtained between estimated and actual converted GPA's. Schusler reports that the English teachers did not know which were the students in their classes who,if they had been scheduled (sectioned) according to the usual method (counselors et al. deciding), would have been placed in a different level than that in which they were assigned by the computer-based method. Among the conclusions she reached, Schusler states that scheduling students by an actuarial method through the use of a computer represents a cost that is less than was currently being spent (presumably for the counselors' time) for the same general purpose.

In summary, the three studies just reviewed represent efforts to explore new ways to assign students by actuarial methods to grade levels (Leton and Anderson, 1964) and to group students within instructional levels of selected subject areas of a grade (Morton, 1961; Schusler, 1964). In these studies, the existing grade or instructional level groupings (designated the a priori groups) are used as the criterion for assessing the efficiency of the statistical model employed. The objective these investigators had was to study the extent to which the test variables, as operated on by the discriminant or regression models, would indicate how much alike were students with similar test score or converted GPA patterns. A basic assumption involved here is that by estimating the degree of similarity existing between a student's own test profile and that of the a priori or criterion group and then making the assignment, the student is therefore scheduled or grouped with other students most like himself. The statistically determined grouping procedure that is represented by the multiple discriminant model is viewed here as analogous to the multiple correlation procedure of grouping students on the basis of their estimated GPA (or any other criterion). This analogy is relevant to the writer's own investigation because it provides the basic rationale by which the multiple discriminant and multiple correlation/regression models can be compared.

In the next set of studies reviewed, the results of investigations concerned with comparing the relative effectiveness of certain actuarial prediction/classification models employed for various purposes will be presented.

Studies Comparing the Effectiveness of Selected Statistical Methods for Predicting/Classifying Students

The group of studies which follow are concerned with comparisons among methods (models) for predicting/classifying students against various criteria (e.g., GPA, field of concentration). Except for Owen's study, the investigations cited below were conducted with college-age samples or with artificial data.

Owen (1956) investigated the effectiveness of four approaches to predicting the GPA of a sample of 832 high school seniors. The methods studied were: 1) prediction from a single ordered variable; 2) prediction from centile ranks (expectancy tables); 3) prediction from multiple regression using five predictor variables; 4) prediction from pattern analysis. After studying the stability of the four methods on a cross-validation sample, Owen concluded that when due consideration was given to the accuracy of prediction, computational and interpretive ease, the following three methods were rated about equal in predictive stability: 1) prediction from a single factor (reading achievement), 2) pattern analysis, and 3) multiple regression. Owen judged the prediction from a single factor cutting score as the simplest, computationally, and better than the other approaches for interpretive purposes. Although not stated specifically, it would appear that Owen would recommend the use of the ordered single variable cutting score (reading achievement) as the optimally most useful method.

Helmstadter's (1957) search of the literature yielded 15 apparently different indices which were recommended by previous investigators for use in the estimation of the similarity of test profiles. Helmstadter defined the problem of assessing the similarity between two profiles as that of determining the degree of similarity between two sets of numbers of which the profiles are constituted. Helmstadter categorized the 15 indices into five groups. Three of the groups, correlation methods (rank, product moment and intra-class); sum of the differences between pairs in the data matrix, squared, and a distance function (which assumes a chi squared distribution) Helmstadter used as quasi-experimental methods in this investigation. As comparison (control) methods, Helmstadter used the indices he categorized in the remaining two groups: the subjective judgment and the linear discriminant function methods. Helmstadter applied the 15 methods to artificial data: 270 geometric solids, one-third of which were spheres, one-third cylinders, one-third tetrahedrons. The data were modified to simulate real data in terms of error components and distribution characteristics. Helmstadter did not define the manner in which the subjective or 'clinical' method was performed except to indicate that three judges were employed for the task. Using chance expectations as the criterion with which to assess the hits/misses record of each of the 15 methods, Helmstadter found no significant differences between methods compared pair-wise. Since insufficient information is provided by Helmstadter to enable one to judge the precise task with which the judges were faced in classifying the profiled artificial data, inferences to real-life-clinical-judgment-decision-making situations are very much limited. Helmstadter provides a rough estimate of the time in minutes (and proportions thereof) needed by each method to classify each profile in the data set. The range was from about twenty seconds ('clinical' method) to about 20 minutes (product moment correlation method). The 'clinical' judges' average rate and success record over that achieved by the statistical methods are indeed impressive. However, since Helmstadter does not indicate by what means the statistical methods were processed, one must assume that a desk calculator was employed. If this was the case, the tabulated rates are so unrealistic as to be hardly worth reporting.

Helmstadter's study is certainly unique in its use of artifical data adjusted to simulate real-life situations. One must assume that his techniques of simulation do in fact approximate reality. His methodology would seem to be most relevant to a basic research approach to mapping the areas in which clinical-actuarial methods might be compared. But a test of the degree to which the simulated data do correspond to actual data would appear to be needed before any generalizations to the real world can be made.

Markwardt (1960) compared the efficiency of predicting graduation from college by pattern analysis, single variable cutting score and linear multivariate (discriminant function) approaches. Using a sample of 400 female elementary education graduates, he randomly divided the total sample into validation and cross-validation sub-samples. He included as predictor variables, weighted achievement, interest and personality variables along with the high school rank (HSR). The efficiency of each approach was determined by the number of correct classifications (graduation vs. non-graduation) compared with the base rate number. Markwardt found that the pattern analysis approach yielded significantly better than the base rate classifications for both the validation and cross-validation samples. For the cross-validation sample, pattern analysis and single variable approaches yielded better classifications over base rate than the discriminant function.

Lawshe and Schuckler (1959) set out to examine whether those trying to predict success against some criterion tend to be needlessly sophisticated in their statistical weighting of test scores for multiple group prediction. Three elements were the focus of their study: 1) weighting method (four methods varying in statistical complexity); 2) sample size (three different N's used--20, 40, 90); and 3) magnitude of predictor inter-correlations (three average correlations ranging from low to medium). The weighting methods ranged in sophistication from multiplication of each predictor score by its least squares regression coefficient to simple addition of predictor raw scores. Lawshe and Schuckler concluded that differential weighting by the methods included in their study is no more significant for multiple group prediction than the simple addition of raw scores when N = 100 or less (no differences were found among sample sizes under test). It was further noted that there was a trend toward less efficient prediction as test inter-correlations increased--a trend that persisted regardless of weighting method or sample size.

Simpson (1957), using a priori proportions as the criterion, studied the effectiveness of three versions of the multiple discriminant model in classifying students into fields of concentration. The three versions were the basic or unadjusted discriminant equations; by equations corrected according to a procedure by Rao (1952); and, finally, by equations adjusted by a formula devised by Simpson (see Chapter III, METHODS, page 63 for brief details of this formula; also see the Appendix, page 201 for an illustration of the computational procedures involved in applying Simpson's formula). The major conclusion reached by Simpson is that the adjusted and corrected classification equations were found to be superior to the basic classification equations against a priori proportions in placing (classifying) students into fields of concentration.

Ludlow (1962) studied the effectiveness of multiple regression and multiple discriminant function in predicting college freshmen grade point average. Ludlow defined his task as that of studying the performance and efficiency of discriminant analysis when there is set for it the objective of assuming the same role regression analysis has, that of classifying or predicting success against a criterion. The major conclusion reached by Ludlow was that when he applied Simpson's (1957) adjustment to the basic discriminant equations, these equations, as adjusted, yielded on check samples correct placements (predictions against a criterion of GPA) that were not significantly different from predictions based on regression analysis; if, in addition, it is desired to predict achievement in line with a priori probability proportions (e.g., selection to a criterion of a specified group size) then the adjusted discriminant equations surpass the regression model.

In contrast to Ludlow's study (1962), Dunn (1959) set up a study, selection of a college major, which required the regression model to assume the role of the discriminant model in indicating with which group students were most like. This assigned role for the regression model had the expectation that if the validation group is large enough, if the predictor variates are actually related to success in the field, and if the resulting multiple R is high enough, then the regression equation can be used to estimate concentration field areas of other students for whom the same data are available. Dunn noted that variables selected by the discriminant model were generally quite different from those selected by the regression analysis. Among the findings reported, Dunn observed the unexpected appearance of a mathematics achievement variable in the regression equations predicting achievement in Modern Languages.

The findings emerging from Dunn's study led her to conclude that characteristics which, by means of the discriminant model, separate the groups are better guides for predicting group membership for new students than are the abilities defined by the educational success patterns established by the regression model. Dunn added that this finding is particularly well demonstrated for concentration fields which normally attract smaller numbers of students. Thus, predictions for students, based on success criteria (e.g., GPA or achievement test results) would result in the highest probabilities being assigned to those fields in which the validation criterion measures (i.e., GPA) are greater. Therefore, Dunn contends, the use of the regression model might lead individuals to select a goal more readily attained (i.e., concentrate in a field in which it would be easier to obtain high grades) than to select fields that may be of greater challenge to their talents. Dunn, in a final summary, writes that the results of her study suggest that the use of the multiple regression model for the guidance of students in making field of concentration decisions (at the college level) is questionable.

Tatsuoka (1955) compared the effectiveness of the regression, discriminant and the joint probability (a system for combining the regression and discriminant models as developed by Tatsuoka) models in predicting: 1) field of concentration, 2) graduation from college, and

3) dropping out of college (by choice or by requirement because of low grades). Employing eleven test variables (concentrated in scholastic aptitude and achievement areas), high school rank, GPA, type of high school, and applying the equations developed on validation samples to check samples, Tatsuoka concluded that: 1) the joint probability model was somewhat superior to the discriminant model and definitely superior to the regression model in predicting field of concentration; 2) that the discriminant and joint probability models were of equal efficacy in predicting graduation with the regression model falling below both of these models; 3) that the discriminant model surpassed the joint probability model in predicting dropouts with the regression model again following these two models in effectiveness.

In the studies just reviewed, the following conclusions appear to be supported:

1. Pattern analysis when compared with other methods emerges as an efficient model for estimating academic success.
2. When the relative effectiveness of the multiple regression and multiple discriminant models were compared against various criteria, the discriminant model appeared to be a promising technique in the kinds of prediction/classification problems it was asked to solve.
3. Use of artificial data with which to investigate the comparative effectiveness of clinical and actuarial classification methods appears to be a promising approach for basic research purposes.
4. A technique that seemed to be the most promising was

that devised by Tatsuoka which combined the multiple correlation/regression and multiple discriminant models for predicting/classifying students against various criteria.

As was mentioned in Chapter I, page 6, the present investigation represents an extension of the efforts of other investigators to study the efficiency of a number of general and specific aptitude and achievement variables for estimating GPA criterion performance within selected subject areas. Further, this investigation represents an effort to evaluate on a pre-college sample (as Ludlow's /1962/ investigation did on a college sample) the relative effectivenss of the multiple regression and two versions of the multiple discriminant equations for predicting/classifying students within instructional levels, since this search of the literature revealed no studies which have attempted this on pre-college samples. This is not surprising in view of the relative newness of the multiple discriminant model which, according to Nunnally (1967), was formally introduced to American researchers in 1950 by Tiedman and associates at the American Psychological Association convention symposium. Additionally, the employment of multivariate statistical procedures such as the multiple regression and multiple discriminant models have been made feasible by the relatively recent availability of large scale electronic computers.

CHAPTER III

METHOD

To reiterate, the major problem on which this investigation focuses is to study, within a framework of a computer-based measurement system, ways in which to effectively group students into instructional levels (Chapter I, page 7). The method followed in studying the problem is presented in this chapter, beginning with descriptions of the validation and check samples.

Samples

Validation Sample

The total validation sample (referred to in this study as the combined group) was comprised of all students of the Hinsdale Township High School (Illinois) ninth grade class of 1966 (graduation year) for whom there were complete data on all the variables included in this study. This combined group validation sample consisted of 544 students. This number represented 82% of the total ninth grade class.

For the purposes of this study, one-half of the combined group validation sample (referred to in this study as the junior high group, N = 272) was separately analyzed since the prediction/classification equations developed in this study could be computed on independent variables that were available on the junior high group during the fall and winter of their eighth grade year; whereas the variables included for the combined group were not available until the spring of their eighth grade year or at the start of their ninth grade year. Thus, the results of the application of the statistical prediction equations for the junior high group could be made available to counselors and their staff while students were still in the eighth grade. This junior high group consisted of all those ninth grade students in the combined validation sample who had attended the village of Hinsdale's junior high school and for whom complete data on all the variables included for analysis were available. The total N of the junior high group validation sample represented 76% of the total junior high eighth grade class who attended the Village of Hinsdale Junior High School as eighth grade students.

Check Sample

The total check sample (again defined as the combined group) consisted of all students of the Hinsdale Township High School ninth grade class (also referred to as the Class of 1967 - their year of high school graduation) for whom there were complete data on the same variables employed in the validation sample analyses. This combined group check sample consisted of 604 students, representing 91% of the total ninth grade class.

The junior high group check sample (defined in the same way as for its validation sample counterpart) consisted of a subgroup of the class of 1967 for whom there were complete data on the same sets of variables as were utilized in the junior high group validation sample analyses. This junior high group check sample consisted of 259 students representing 74% of the total population of the junior high group who had attended the Village of Hinsdale Junior High School as eighth grade students.

Community Setting[1]

The combined group was made up of two groups of students, half of whom lived in the Villages of Hinsdale and Clarendon Hills. The other half came from surrounding areas which were a part of Hinsdale Township, having attended grades K-8 in one of seven elementary school districts.

In order to give a brief picture of the socio-economic level of the families living in Hinsdale Township, it might be well to focus on two communities, Hinsdale-Clarendon Hills, from which the total junior high group came, and Westmont, which represented the largest percentage of students in the remaining half of the combined group. The data recorded in Table I are as compiled in 1961.

Westmont, made up of a slightly younger age group, had a median school year completed of 10.6, 13.3% professional workers, and a median family income of $7,600. Hinsdale ranked ninth and Clarendon Hills 13th among 75 Chicago communities and 175 suburban communities, based on a 1960 composite of family income, years of completed schooling and percentage of professional workers. In combination, Hinsdale-Clarendon Hills had a median school year completed of 13.1, 43.9% professional workers and a median family income of $11,000. The remainder of the Township lies on a socio-economic scale between Westmost and the Villages of Hinsdale and Clarendon Hills.

---

[1]Sources for the information included in this section were the following: Suburban Factbook, Chicago: Northeastern Illinois Metropolitican Area Planning Commission, 1962; Renstrom, R.T., Hinsdale: Community Profile, Omnibus and Chicago FM Guide, August, 1965.

TABLE I

COMPARISON OF SOCIO-ECONOMIC FACTORS ASSOCIATED WITH THE COMMUNITY SETTING OF THE STUDY SAMPLE

	Hinsdale-Clarendon Hills	Westmont
Population	19,700	3,300
Median age	33.5	30.0
Median school year completed	13.1	10.6
Median family income	$11,000	$ 7,600
Median house value	$28,000	$15,200
Professional workers	43.9%	13.3%

Hinsdale-Clarendon Hills are communities which have attempted to retain their past traditions and small-town atmosphere despite their closeness to and association with a large metropolitan area. They are commuter suburbs, 18 miles and 22 express-minutes west of the Chicago Loop. The educational system is made up of eight public elementary schools, one junior high and Hinsdale Township High School from which 80% of the graduates continue on to colleges and universities; the pupil-teacher ratio is 25-1 and the school district spends an average of $470 per pupil per year, figures which are at the median when compared to other Chicago suburbs.

Subject Areas

The subject areas listed below with their corresponding levels were the focus of this study and are briefly described.

English

On the bases of the eighth grade teachers' recommendation and whatever eighth grade test and grade data were available plus the results of the Cooperative Reading Test (see page 48 for description), the high school counselors and members of the English department pooled their judgments in sectioning the incoming ninth grade students into one of four of the following levels of the English course: reading (about 11%), basic composition (about 12%), composition or speech (about 63%) and accelerated composition (about 13%). When a student was sectioned into the largest section - composition or speech - it was entirely a matter of scheduling factors that determined whether a student would take composition or speech for the first semester. If it happened to be composition for the first semester, the student automatically took speech the second semester and vice versa. Since neither the reading nor the basic composition courses enjoyed a prestige status among the students and their parents, it happened in a number of instances that parental insistence resulted in the student being placed in either the composition or speech section. No provision was made for parental request that their son or daughter be placed in the accelerated or honors composition section.

Owing to the fact that the reading course was dropped from the curriculum during the second year, and because students who were grouped in composition the first semester automatically took speech the second semester, the study was limited to an evaluation of the regression and discriminant models using the three levels of composition within the English subject area.

Mathematics

Eighth grade math teacher recommendations along with teacher grades and available test data formed the basis upon which the high school counselors and math department members sectioned incoming ninth grade students into one of four required levels: remedial math (3%), general math (25%), algebra (59%) and honors algebra (13%). Remedial math was not included in this study because of the extremely small sample size and also because it was dropped from the curriculum during the second year of the study.

Social Studies

Incoming ninth grade students were sectioned by the high school counselors and social studies department staff members into one of two required year-long courses, world geography and western civilization (about a half in each), again on the bases of eighth grade teacher recommendations, grades and available test data. The world geography course was the one into which the less academically able student was sectioned.

Foreign languages

Incoming ninth grade students who were encouraged by their parents, eighth grade teacher and/or counselor recommendations (partly based on data obtained from the high school-administered Modern Language Aptitude Test, described on page 49), had the option of electing to continue in the second year of French or Spanish if they were taking it in the eighth grade or of enrolling in the beginning course of any one of five foreign languages: French, Spanish, Latin, German or Russian. About two-thirds of the incoming ninth grade students so elected. Latin and Russian were eliminated from the study because of their small enrollment.

Independent Variables, Combined Group Validation Sample

Iowa Tests of Educational Development (ITED)

The ITED, Form SL-4, was administered to the incoming ninth grade students during the third week of September. This battery includes nine objective tests designed to measure educational development which is the result of several factors such as experiences provided by the school, home, community and individual initiative. The sub tests, each of which were included as independent or predictor variables, comprising the battery are the following:

1. Understanding of Basic Social Concepts (Bkd SS)
2. Background in the Natural Sciences (Bkd NS)
3. Correctness and Appropriateness of Expression (Corr Expr)
4. Ability to do Quantitative Thinking (Q Thkg)
5. Ability to Interpret Reading Materials in the Social Studies (Rdg SS)
6. Ability to Interpret Reading Materials in the Natural Sciences (Rdg NS)
7. Ability to Interpret Literary Materials (Rdg Lit)
8. General Vocabulary (Gen Vocab)
9. Use of Sources of Information (Ref)
10. Composite (Compos)

Students recorded responses on an answer sheet purchased from the test publisher. The scoring facilities of the publisher (Science Research Associates) were used. Raw scores were converted to the standard score scale (national norms) established for the battery (ranging from 0 to 40). These data were punched on Hollerith cards as a part of the scoring service. As punched, the data were employed later in the analyses.

The ITED battery, in its several editions since the original publication in 1942, has been extensively reviewed. In one review (Page, 1965, p. 50) the following statements are made concerning the predictive validity and split-half reliabilities of the ITED:

> Predictive validity (of the ITED) rests upon the correlation of prior editions of the ITED with later school and college success (which range) from the .40's to the .70's or higher between ITED composite scores and rank in high school graduating class, average high school grades in specific courses, high school grade-point averages and college freshmen grades. . .
>
> Within-grade split-half reliabilities for all tests. . . (range) in the .80's and .90's, with reliabilities for the composite reading. . . .98 or .99. .
>
> In summary, the ITED is a modern battery of subject area tests designed in conformity with good canons of test construction, supplied with high quality norms and statistical information. . .With some reservation about profile differences, it measures what it measures very well.

<u>Cooperative Reading Tests</u>

The Cooperative Reading Test (Co-op Rdg), 1960 Revision, Lower Level, C, Form A, was administered at the request of the high school English department by the high school counselors to incoming ninth grade students during March of their eighth grade year. The subtests included are: Vocabulary, Level of Comprehension, Speed of Comprehension. Students recorded responses on IBM answer sheets; scoring was accomplished on a locally available IBM Model 026 scoring machine. Raw scores on each subtest were converted to the standard score scale (National norms) developed by the publisher, a scale ranging from 200 to 800. A review of this test (Fleming, 1965, pp. 1084-1085) summarizes validity and reliability information as follows:

> Validity coefficients obtained from a number of studies are presented. . .not only for total reading, but also for Vocabulary, Level of Comprehension and Speed of Comprehension. These studies are based on earlier editions of the test, but it is claimed, not unreasonably, that the latest edition is enough like these that the

> findings may be considered relevant. . .Reliability coefficients between alternate forms are presented for different parts of the tests and for different grade groups. These are satisfactorily high.

<u>Modern Language Aptitude Test</u>

The Modern Language Aptitude Test (MLAT), Form A (1959), long form, is a battery consisting of two parts, including five subtests, two subtotals, and a total score as follows:

Part I: Number Learning, Phonetic Script, Part I Sub-total (also called Side A)

Part II: Spelling Clues, Words in Sentences, Paired Associates, Part II Sub-total (Side B)

Total for the battery.

IBM answer sheets were used and scored on a locally available IBM Model 026 scoring machine. Raw scores for each of the five subtests, along with the total, were key punched onto Hollerith cards for later processing. Among one of the most perceptive reviews of tests to be found anywhere is the one prepared by Fisher and Masia (1965) on the MLAT battery:

> Tests designed to predict achievement in the study of a foreign language invariably reflect two rather serious weaknesses: (a) they are not rooted in psychological studies of language and language behavior, and (b) they do not indicate the specific language learning outcomes or instructional objectives the test is designed to predict.
>
> The first weakness has given rise to validity coefficients between predictor and criterion variables which are not significantly higher than coefficients obtained when general scholastic ability tests are used as predictors. . .The second weakness gives rise to a range of predictive validity coefficients when the prognostic instrument is used with foreign language classrooms of different teachers and in different schools. Since error and bias in the criterion variables, particularly when it is represented by teacher grades, is most likely random across teachers and schools, variability in predictive validity coefficients may in large measure be associated with variations in instructional objectives.

Against the setting of these two major weaknesses of foreign language aptitude test, Fisher and Masia (1965) analyze how well the MLAT fares. They point out that predictive validities show a Pearson _r_ coefficient increment of .20 over general ability and intelligence tests in predicting success in foreign language study (high school level, combination of grades and sex; teacher grade, the criteria). In their overall evaluation of the MLAT battery, Fisher and Masia indicate that the test content suggests to them that the MLAT test measures the student's ability to recode English and that as such the test may well be a better predictor of English grades than of grades in a foreign language course.

In summary, a total of twenty-one scores drawn from the ITED, Co-op Reading and MLAT test batteries, comprised the data sources for the independent or predictor variables for the combined group.

Independent Variables, Junior High Validation Sample

In addition to the three sets of variables listed above (ITED, Co-op Rdg., MLAT) for the combined group (of which, to repeat, the junior high group was a part), the three sets of variables described below were also entered into the predictor variable matrix for the junior high group. Some of these same test data were available for some of the other half of the combined group; however, these tests were not administered to those students at the same time as was the case for the junior high group and for this reason were not included.

Iowa Tests of Basic Skills

The Iowa Tests of Basic Skills (ITBS), Form I, 1955 multi-grade edition, is a battery designed to test the functional skills of students, grades 3-9, in the following areas: reading comprehension, vocabulary, language skills, work study skills and arithmetic. The test battery was administered to the students during the first week of October in their eighth grade year over a period of four half-days. The students recorded their responses on the Measurement Research Center (MRC) answer sheets secured from the test publisher and were scored at the MRC scoring facilities, Iowa City. The results were reported in grade equivalents and centile ranks (national norms). Local norms were developed by the writer from these reported distributions. These local centile norms were the basis used for converting the grade equivalents to normalized T scores, using the conversion table found in Cronbach (1960), for each subtest and subtest total and battery total score. Fifteen normalized T scores were thus derived from the subtest, subtest total and overall composite grade equivalent scores. The subtest titles and their test label code used in this study are as follows:

V	Vocabulary	W_1	Maps
R	Reading	W_2	Graphs
L_1	Spelling	W_3	References
L_2	Capitalization	W_t	Work Study Skills Total
L_3	Punctuation	A_1	Arithmetic Concepts
L_4	Usage	A_2	Arithmetic Problem Solving
L_t	Language Total	A_t	Arithmetic Total
		C	Composite Total for battery

The junior high group students comprising this sample had two previous experiences in grades four and six with this form of the test at the appropriate grade level. In his review of the ITBS battery, Herrick (1959) observed that:

> This battery cannot be considered as an achievement test in the usual sense of measuring knowledge in the common content areas. . . (It is a battery) of generalized achievement. A major strength of this new battery is its curricular validation. . . (Split-half) reliabilities range from .84 to .96 for the major test (areas) and from .70 to .93 for the subtests. . . Intercorrelations among the various subtests range from .37 to .83, with the average ranging from .60 to .70. . . The tests of vocabulary and reading comprehension have the highest intercorrelation will all other subtests, indicating a heavy loading of all subtests with vocabulary and reading skills.

Another reviewer of ITBS (Morgan, 1959), also published in the fifth edition of Buros, notes the absence of predictive validity data. A 1964 edition of the ITBS administrator's manual (Lindquist and Hieronymus) reports the results secured by Scannell (1958) and reviewed in Chapter II, REVIEW, of this study. Briefly those results were: a zero order _r_ of .73 between grade eight ITBS and grade twelve ITED composites and an _r_ of .61 between grade eight ITBS composite and high school GPA.

Lorge-Thorndike Intelligence Tests

The Lorge-Thorndike Intelligence Tests (LTSA), Verbal and Non-verbal, Form A, Level 4 (grades 7-9), 1957 multi-level edition, was described by the authors in the revised Technical Manual (Lorge and Thorndike, 1962) as:

> . . . a series of tests of abstract intelligence covering the range from kindergarten to college freshmen. Abstract intelligence is defined as the ability to work with ideas and the relationships among ideas. The tests are based on the premise that most abstract ideas with which the school child or working adult deals are expressed in verbal symbols, so much so that verbal symbols are the appropriate medium for the testing of abstract intelligence. Nevertheless, they take account of the fact that for some - the young, the poorly educated, or the poor reader - printed words may constitute an inadequate basis of appraising an individual's abilities. Consequently, a parallel set of nonverbal tests is provided to accompany the basic verbal series.

The tests were administered to the students in February of their eighth grade year (they had taken the appropriate level of this form of the tests in the sixth and fourth grades). The students recorded their responses on the MRC answer sheets which were scored at the MRC scoring facilities, Iowa City. The results were reported and punched as deviation standard scores (national norms) on Hollerith cards by the scoring service. Three scores, Verbal, Non-verbal and Total (the total is an arithmetic average of the Verbal and Non-verbal scores) were used as the three predictor variables from this test. Freeman (1959) had the following to say about the LTSA test:

> This 1957 version of the Lorge-Thorndike Intelligence Tests is among the best group tests available, from the point of view of the psychological constructs upon which it is based and that of statistical standardization. . .
>
> Evidence of reliability of the scales is presented in several ways. Alternate forms correlate rather well (.76 to .90) at all levels, but the verbal scales for levels 3, 4 and 5 yield the highest coefficients, namely .90, .86 and .86. All these coefficients are all the more significant since, in each instance, they were computed on the population of a single grade. . .The odd-even reliabilities are very high (.88 to .94). . .(Relative to predictive validity) a correlation of .67 between (the LTSA) given at the beginning of grade 9 and the "average achievement" of 214 pupils at the end of the grade (was reported). . .
>
> (Concurrent validity evidence is as follows): the correlation between (LTSA) IQ's and Stanford (achievement) grade equivalents in reading was .87,. . .(and) arithmetic was .76 for 171 sixth grade pupils. . .Congruent validity (indicated by) correlations. . .with four other group tests as well as with the Binet and WISC (can be observed by) coefficients (which) were .60 or higher. . .

The test authors (Lorge and Thorndike, 1962) report correlations of the LTSA with subtests and composite of the ITED administered at grade ten that range from .54 to .69 (Non-verbal LTSA) and .70 to .86 (Verbal LTSA) and .70 to .86 (Total LTSA). They further report on a study of a correlation of .67 between the LTSA administered at the beginning of grade nine and GPA at the end of grade nine.

First Semester Eighth Grade GPA's in Five Subject Areas

As the final set of independent variables used in selected subject areas of the junior high group, the GPA (A = 5, F = 1) in each of the following subject areas were employed: English, mathematics, social studies, science and foreign language. The GPA in each of these five junior high subject areas was a simple arithmetic mean of the grades assigned in each subject. The foreign language grade column in the Hollerith cards was deleted when eighth grade GPA data were included as independent variables for predicting a non-foreign language GPA criterion variable, since not all students of the junior high group were taking a foreign language during their eighth grade year.

Criterion Variable, Combined and Junior High Validation Samples

For each level of each subject area included in this study, the criterion variable was the arithmetic average of the grades received by the student during the first semester of the ninth grade where A = 5; F = 1. This arithmetic average was called the grade point average (GPA). The first semester GPA within levels of subject areas was chosen as the criterion variable for two reasons: 1) this was the period during which considerable counselor, teacher and student time was expended on the rescheduling of students who had been assigned to achievement levels in the manner described in Chapter I, page 3, and who, because they may have been placed in a level too difficult for them and were failing, or the level was judged by someone (parent, teacher, student) as not challenging enough; and 2), it was assumed that the success pattern reflected in the GPA which the student established during the first semester would persist during the second semester despite curricular and sometimes staff changes occuring during the second semester.

Statistical Procedures

Correlation - Regression Model

As noted in Chapter I, a basic purpose of this study is to answer the question: "In which level of a ninth grade course (English, mathematics, social studies, foreign language) will the student perform best?" The relevant task is to select the five or six test data variables and, in certain instances, GPA variables, that will best answer this question, using GPA in the level of the course as the criterion. The statistical model appropriate to seeking an answer to this question is the multiple correlation/regression model. Since the objective is to select five or six variables that optimally answer this question, the procedure known as step-wise regression is the appropriate one. A computer program[2] developed for this purpose was used. As stated in the BMD user's manual (Dixon, 1964).

> This program computes a sequence of multiple linear regression equations in a stepwise manner. At each step one variable is added to the regression equation. The variable added is the one that makes the greatest reduction in the error sum of squares. . .it is the variable which, if it were added, would have the highest F value. In addition, variables can be forced into the regression equation and automatically removed when their F values become too low.

---

[2] The majority of the multiple correlation-regression analyses were computed using the BIMD 34 program as written at the Health Science Computing Facility, Department of Preventive Medicine and Public Health, School of Medicine, University of California, Los Angeles. The user's description of this program is published in the BIMD Manual, 1959 edition (Dixon, 1964). In the later phases of the data processing, a slightly revised program was developed at the UCLA Health Sciences Computing Facility. This revised program (January 1, 1964),BMD02R, replaced the BIMD 34. The computer facility at the Argonne National Laboratory, Applied Mathematics Division, Argonne, Illinois, was used for this processing. Since the BIMD34/BMD 02R programs were written

Because the number of the independent variables for the combined group was twenty-one and for the junior high group was forty-four, the number of steps in the step-wise regression program was set at six for three practical considerations: 1) although the variance in the dependent variable, GPA, might not be maximally accounted for by six variables, perhaps not even by all the variables in the data sources used, it was desired to have a set of variables whose number could be managed, for example, by the school personnel (and the writer) in application to new samples; 2) it was anticipated that by the sixth step, little, if any, of the remaining error sums of squares would be reduced further at a statistically significant level by adding variables to the regression equation; and, 3) constraints on availability of funds for computer processing argued for a limit of six variables.

The regression equations, developed within the junior high validation sample subject areas, were then applied to a junior high check sample which yielded proportions that were compared with the actual numbers (a priori) of students comprising the groups at the end of the

---

for IBM 704/7094 computers, they had to be slightly modified so that they could be run on the Argonne Lab's Control Data Corporation (CDC) Model 3600 computer. The initial modification on the BIMD 34 program was made by Mrs. Hustand, programmer-mathemetician in the Applied Math Division. Later, Mrs. Fu, programmer-statistician in the Applied Math Division, modified the BMD 02R program for use on the CDC 3600 and, like Mrs. Hustand earlier, worked closely with the writer in the data processing procedures.

The IBM 7094 computer facilities on the campus of the University of Chicago were used for calculating univariate and bivariate statistics for the total combined and junior high group samples. Certain sample statistics such as skewness and kurtosis could be calculated through use of computer programs available at the University of Chicago campus.

4. The standard deviations of the estimated GPA's in each subject area were adjusted to the Level 1 and 3 a priori proportions in each subject area.
5. Students were assigned to Level 1 or Level 3 in each subject area according to whether their GPA (as estimated by the Level 2 prediction equation) was smaller or larger than the adjusted standard deviation.
6. Students whose estimated GPA fell within the plus and minus one adjusted standard deviation were assigned to the Level 2 group within each subject area.

Testing Assumptions

Prediction equations computed on the basis of the multiple regression technique are generalizable to other similar populations for estimating a criterion variable to the extent that the assumptions underlying the statistical model being employed have been tested. The importance of testing the assumptions would appear to be especially relevant when the size of the samples employed is relatively small, as is the case for the junior high group in this study, and when different statistical models are being studied by themselves and are being compared for their relative efficiency in answering the questions posed. According to Johnson (1949, pp. 240-245), the basic assumptions underlying the use of the multiple correlation/regression model are:

1. That the distribution of the variables in the total sample, of which the levels within subject areas are a part, are normal. Statistical procedure: chi squared goodness-of-fit test (Guilford, 1965, pp. 243-247).

2. That homoscedasticity of the criterion score distributions obtains; i.e., that the spread of the array (scatter) of criterion (first semester GPA) scores for each predictor variable, within each level of the English and mathematics subject areas, does not depart significantly from the mean of the array. Statistical procedure: Welch-Nayer L_1 Criterion test (Johnson, 1949, pp. 240-242).

3. That a linearity of the multiple regression line obtains; i.e., that the mean of each vertical array (criterion variable) should not depart significantly from the regression line; or, there is a linear relation between the independent variable and the criterion within each level of the English and mathematics subject areas. Statistical procedure: analysis of variance (Johnson, 1949, pp. 240-244).

These assumptions underlying the inferential use of the multiple correlation/regression model will be partially tested for the combined group. Tests of the normality of the distribution of the independent variables will be provided for the combined group. The normality assumption will be evaluated in terms of the skew (spread) and kurtosis (peakedness) of the test score distribution. The reasons for this variation in testing the assumptions are as follows:

1. Since the administration of two out of the three test batteries (Co-op Reading and MLAT) were discontinued by the high school personnel, and,

2. Since the third battery (ITED) was administered to the students during the fall of their ninth grade, the data was thus secured too late for predictive (classification) use by the counselors and other school personnel,[3] it was decided to limit the testing of the assumptions underlying the multiple regression and discriminant models to the junior high group.

Multiple Discriminant Function

Since a corrollary purpose of this study is to answer the question "Within a subject area (English and mathematics), with which group (track, level) of students is an individual most like," the statistical procedure that is appropriate for answering this question is the multiple discriminant analysis. Because a further objective in this study was to compare the efficiency of the multiple regression and multiple discriminant function models in classifying/predicting students within each of three levels of two subject areas (English, mathematics) and because no step-wise discriminant procedure comparable to the step-wise regression procedure was available at the time the data of this study were processed,[4] the following criteria were used in selecting the six

---

[3]It should be noted, however, that the statistical prediction/classification procedures would nevertheless be available for counselor et al. use in classifying students enrolling later from outside the district; further, consideration was being given to administering the ITED battery just prior to the opening of school. Whether these or other application possibilities were present, demonstration of the relevance of the statistical approach was of primary consideration.

[4]A step-wise discriminant analysis computer program, BMD 07M, developed also by the UCLA Health Sciences computing facility, is now available.

independent variables for use in the discriminant model:

1. They would be selected from among the variables selected at the sixth step of the step-wise regression program.
2. They would be those variables which appeared in the regression equations of all levels of each subject area, English and mathematics.
3. They would be those variables whose means in each level of a subject area exhibited the largest difference among groups while their variances displayed the most consistent homogeneity.

A computer program[5] developed for calculating the multiple discriminant function equations for several groups was used. According to the BMD User's Manual (Dixon, 1964):

> This program directs the computation of a set of linear functions for the purpose of classifying an individual into one of several groups. . .The group assignment procedure followed is derived from a model of multivariate normal distribution of observations within groups such that the covariance matrix is the same for all groups. An individual is classified into the group for which the estimated probability density is largest. The equivalent computational procedure followed evaluates the computed linear function corresponding to each of the groups and assigns an individual to the group for which the value is the largest.

Essentially this involves computing a set of weights which will maximize the ratio of the between-means variance to the within-levels variance. For the problem in this study (see Appendix, p.206, for an

---

[5]This program, coded BMD 05M, has a sample size limitation of 150 cases in any one group. This limitation was modified by Mrs. R. Fu, statistical programmer at the Argonne National Laboratory where the data was processed. The modification allowed for a maximum N of 175 for any one group. This computer program, like the step-wise regression, BIMD34 and BMD 02R referred to earlier, was also modified to be run on a Control Data, Model 3600, computer.

illustration of the application of the discriminant equations to selected data), three sets of six weights were computed for each subject area (one set for each instructional level, and since six variables were used, there were six weights in each set). After the weights (the coefficients of the discriminant equation) were obtained, each person's score on the six test variables included in the equation for each subject area was multiplied by each of the six weights in each set and summed over each. The resulting three sums of products for each person are called the discriminant scores. A person is classified in that level for which his discriminant score is the largest. Thus, if a person's a priori membership in English is Level 2 (Composition) and his largest discriminant score is for Level 3, he would be assigned (classified in) to Level 3 by the discriminant procedure. The three sets of weights together with the three constants (one for each set of weights which are analogous to the constants computed in the regression equations) are called the basic discriminant equations. The classifications of individuals which results from the processing of each individual's score through the three discriminant equations are such that the individuals classified are more like each other with respect to their test scores than they are like the members of other levels.

While the groups thus classified may be desirable in terms of the test score homogeneity, their size may not fit the administrative requirements of the setting (school in this study). The problem then is to maintain the essential nature of the group separation (homogeneity) while at the same time to effect a modification of the group size so as to conform to administrative requirements. A technique for modifying the basic discriminant equation so as to take into account the unequal sizes of the groups (i.e., the a priori proportions) upon which the discriminant equations were developed was devised by Simpson (1957). In essence, Simpson's adjustment method involves adding a constant, K, to the discriminant equation's constant so that when the students' input variables have been operated on by the adjusted discriminant equations, the resulting discriminant scores will classify the students into homogeneous instructional groups in proportion to the size limitation set administratively for each level. Simpson's formula and an illustration of its use, together with associated statistics (means, standard deviations and discriminant equations) are presented in the Appendix (page 201).

Assumptions underlying the use of the multiple discriminant function model are:

1. That the variables which are entered as input to the discriminant model are normally distributed. Statistical procedure: chi squared goodness-of-fit test (Guilford, 1965, pp. 243-247).
2. That the within-group homogeneity of variances obtains. Statistical procedure: Bartlett's test of homogeneity of variance with unequal degrees of freedom (Edwards, 1960, p. 127).
3. That equal probability densities prevail in each of the groups. This assumption does not obtain. As indicated above, Simpson's (1957) formula is employed for adjusting the basic discriminant equations to allow for the unequal densities in each level or group within a subject area.

Statistical Procedures Employed for Testing the Hypotheses

For the reader's convenience, the major and minor hypotheses under test in this study, as presented in Chapter I, pages 8ff., are repeated here along with the statistical procedure appropriate for testing each one. The probability for determining whether to accept or reject each hypothesis, stated in the null form, was set at $\alpha = .05$.

1. To test MAJOR HYPOTHESIS I, which states that within the combined and junior high validation samples there is no statistically significant reduction in the error sums of squares as a function of the predictor variable entered at each step to a limit of six steps of the step-wise regression routine, the statistical procedure is: analysis of multiple regression (Wert et. al., 1954, pp. 237-249).
2. To test MINOR HYPOTHESIS I, which states that within instructional levels of selected subject areas of the junior high validation sample there are no significant differences in the multiple R's at the sixth step of the step-wise multiple regression routine as a function of different sets of independent variables drawn from additional data sources, the statistical procedure is: Hotelling's F test for the significance of difference between correlated sample multiple R's (Wert et al., 1954, p. 299).
3. To test MAJOR HYPOTHESIS II, which states that within instructional levels of selected subject areas there are no significant differences between the validation and check sample multiple R's of the combined and junior

high groups, the statistical procedure applied is the $\underline{z}$ test for the significance of difference between independent sample multiple R's (Wert et. al., 1954, p. 296).

4. To test MAJOR HYPOTHESIS III, which states that there are no significant differences in the proportion of the junior high group check sample students predicted/classified within levels of two subject areas (English and mathematics) by means of multiple regression equations and the proportions expected a priori, the statistical procedure is: the single classification chi squared test (Dixon and Massey, 1957, p. 222).
5. To test MINOR HYPOTHESIS III, which states that the overall proportion of junior high group check sample students correctly predicted/classified in two subject areas (English and mathematics) by multiple regression equations does not differ significantly from the proportion expected based upon the operation of chance, the statistical procedure is: the single classification chi squared test (Dixon and Massey, 1957, p. 222).
6. To test MAJOR HYPOTHESIS IV, which states that the proportion of the junior high group check sample students predicted/classified within three levels of two subject areas (English and mathematics) by means of two versions of multiple discriminant equations does not differ significantly from the proportion expected a priori, the statistical procedure is: the single classification chi squared test (Dixon and Massey, 1957, p. 222).

7. To test MINOR HYPOTHESIS IV, which states that the over-all proportion of the junior high group check sample students correctly predicted/classified in two subject areas (English and mathematics) by means of two versions of multiple discriminant equations does not differ significantly from the proportion expected based on the operation of chance, the statistical procedure is: the single classification chi squared test (Dixon and Massey, 1957, p. 222).

8. To test MAJOR HYPOTHESIS V, which states that there are no significant differences between the over-all number of junior high group check sample students correctly predicted/classified in two subject areas (English and mathematics) by each statistical method and the total number succeeding in the groups in which they were registered, the statistical procedure is: McNemar's chi squared test for two correlated samples (Siegel, 1965, pp. 63-67).[6]

9. To test MAJOR HYPOTHESIS VI, which states that there are no significant differences between subject areas in the over-all number of junior high group check sample students correctly predicted/classified by each statistical method, the statistical procedure is: the *z* test of the significance of

---

[6]It should be noted that in these comparisons against a priori proportions, each pair of total 'hits' are regarded as being drawn from correlated samples. This view of the a priori 'hits' proportions with which each statistical prediction method is compared differs from the way in which the a priori proportions were regarded in the test of Major Hypotheses III and IV where they are the expected proportion for which an independent sample single classification chi squared test was appropriate.

differences between non-correlated proportions (Walker and Lev, 1953, p. 78).

10. To test MAJOR HYPOTHESES VII, which states that there are no significant differences between statistical methods in the over-all number of check sample students correctly predicted/classified in two subject areas, the statistical procedure is: McNemar's chi squared test for two correlated samples (Siegel, 1956, pp. 63-67).

CHAPTER IV

RESULTS

Initially, this chapter will present the findings that relate to <u>phase one</u> of the problem set for this investigation: to study the predictive validity of the relatively extensive amount of test data and certain non-test data available on the target population for estimating first semester ninth grade GPA's within instructional levels of four subject areas through use of the step-wise multiple correlation/regression technique. These results are presented in two parts, first for the combined sample and then for the junior high sample. There will then be presented the results derived from the focus on <u>phase two</u> of the problem: to study the effectiveness of the multiple correlation/regression model for predicting/classifying students into instructional levels of two subject areas, English and mathematics, against various criteria. <u>Phase three</u>, parallel to phase two, will study the effectiveness of two versions of the multiple discriminant model for predicting/classifying students. In the <u>fourth phase</u> of the problem, attention is directed to a comparison of the effectiveness of the statistical models, singly and in pairs, in predicting/classifying students within instructional levels of two subject areas.

Following is PHASE ONE of the problem: The Multiple Regression Model for Studying the Predictive Validity of the Variables Used in Estimating GPA within Instructional Levels of Four Subject Areas.

Combined Sample

Before the results of the regression analyses are presented for the combined sample, a report is given of the normailty of distribution assumption which underlies the use of the regression model. This assumption is tested for all of the ITED, Co-op Reading and MLAT subtest score distributions of the validation sample.

Test of normality of distribution assumption: The means, standard deviations and their standard error, along with indications of the skew and kurtosis of the distribution of scores comprising these data sources for the multiple regression analyses are presented in Table II.

A value of zero for skewness demonstrates symmetry in the score distribution. A positive skew indicates an excess in the number of scores smaller than the mean value while a negative skew denotes a piling up of scores above the mean value. Similarly, a zero value for kurtosis demonstrates symmetry, i.e., there is no excess in the number of scores at the mean of the distribution. A positive value for kurtosis indicates an excess of scores at the mean and toward the tails of the distribution, while a negative kurtosis value indicates that the distribution of scores are not centered at the mean; rather, that the scores are distributed in a manner such as to give a picture of a flattened curve. Taken together, the kurtosis and skew values recorded in Table II are used as indices for departure of a particular test score distribution from the normal distribution. To evaluate whether the observed values of skew and kurtosis depart significantly from the values expected based upon normal distribution, the ratio of the observed skew and kurtosis values to the standard errors is calculated. The resulting quotients are referred to the *t* distribution.

TABLE II

MEANS, STANDARD DEVIATIONS AND THEIR STANDARD ERRORS, SKEW AND KURTOSIS OF THE DISTRIBUTION OF THE IOWA TESTS OF EDUCATIONAL DEVELOPMENT (ITED), COOPERATIVE READING (CO-OP) AND MODERN LANGUAGE APTITUDE TESTS (MLAT)

Combined Group Validation Sample

N = 544

Variable	Mean	Error	Std.Dev.	Error	Skew	p.01	Kurtosis	p.01	Score Range Min	Score Range Max
ITED										
Bkd Soc St	15.89	0.24	5.61	0.16	0.05		-0.24		1.00	30.00
Bkd Nat Sc	16.92	0.25	5.71	0.15	-0.29		-0.44		1.00	29.00
Corr Expres	14.91	0.21	4.88	0.15	-0.12		-0.07		1.00	28.00
Quan Think	14.06	0.27	6.37	0.19	0.38	<	-0.09		1.00	32.00
Rdg Soc St	14.03	0.26	6.09	0.16	0.35	<	-0.43		1.00	21.00
Rdg Nat Sc	15.25	0.28	6.45	0.18	0.09		-0.39		1.00	32.00
Rdg Lit	15.37	0.23	5.45	0.15	0.03		-0.34		1.00	29.00
Vocabulary	17.06	0.21	4.82	0.16	-0.03		0.26		1.00	30.00
Composite	16.05	0.25	5.76	0.18	0.23		0.10		1.00	35.00
Use Ref	16.38	0.28	6.60	0.16	-0.14		0.78	<	1.00	31.00
Co-op										
Vocabulary	51.48	0.34	7.96	0.27	-0.12		0.51		27.00	74.00
Level Compre	50.24	0.40	9.21	0.26	-0.52	<	-0.23		21.00	69.00
Speed	50.03	0.38	8.86	0.25	0.19		-0.32		29.00	74.00
MLAT										
Number Learn	23.96	0.43	9.98	0.23	-0.12		-0.81	<	1.00	44.00
Phonet Script	20.86	0.18	4.15	0.13	-0.05		0.02		5.00	31.00
Side A	44.74	0.54	12.64	0.31	-0.01		-0.66	<	5.00	74.00
Spell Clues	9.79	0.24	5.52	0.15	0.60	<	-0.39		1.00	26.00
Wds in Sent	15.20	0.26	6.06	0.21	0.67	<	0.60	<	1.00	37.00
Prd Assoc	13.43	0.24	5.54	0.12	0.27		-1.00	<	1.00	25.00
Side B	38.38	0.56	12.97	0.37	0.54	<	-0.21		10.00	78.00
Total	88.16	0.99	23.05	0.60	0.30	<	-0.53		31.00	143.00

Skew Std. Error = 0.11 Kurtosis Std. Error = 0.21

A review of the ITED skew and kurtosis data in Table II indicates that in two score distributions, Quantitative Thinking and Reading Social Studies Materials, there is evidence of significant departures from zero in the skew values while a significant _t_ value was obtained for the negative kurtosis value of subtest 10, Use of References, indicating a platykurtic-type distribution of the scores for this subtest. Further insepction of Table II yields information that a number of the distributions of MLAT scores depart significantly from the normal distribution.

The inferences to be drawn concerning the findings just reported are as follows:

1. Use of the multiple regression model for the situations where the ITED battery served as the data source for the predictor variables was felt to be justified on the basis of the rationale offered by Remple (1960) which is that if the multivariate analysis equations are applied to new samples (as they were in this study, see page 100) an evaluation of the relative effects of the departures from the normality assumption (very few departures for the ITED data) is thus made available.
2. As for the appropriateness of the multiple regression model for the situations where the MLAT and Co-op Reading Test batteries were included along with the ITED battery in the data source, the regression equations developed with predictor variables drawn from these sources must be used, according to Remple, with considerable caution.

In summary, the data presented in Table II indicates that the normality assumption underlying the multiple regression model is met fairly well by the ITED score distributions and less well by the Co-op Reading and MLAT tests. Attention was given to the implications that stem from the finding that the normality assumption was not fully met by all the test score distributions included in the predictor data sources.

Predicting Achievement Within Levels of Four Subject Areas

With the ITED battery as the data source, the results of applying the multiple regression equation to the combined validation sample for the purpose of predicting first semester ninth grade GPA within levels of four subject areas will be presented in the following order: social studies, foreign languages (with additional data sources), English and mathematics. At the end of the report of the results for each area, the disposition of Major Hypothesis I will be given.

This hypothesis states that there is no statistically significant reduction in the error sums of squares as a function of the predictor variable entered at each step to a limit of six steps of the step-wise regression routine. It should be noted that the disposition of the hypothesis in each instance and the inferences made concerning the optimally efficient prediction equation are data and situation relevant, that is, the inferences made from the disposition of the hypothesis are valid to the extent that the sample characteristics as well as characteristics associated with the predictor and criterion variables continue to be present. Because changes do occur in any one or more of the components upon which the prediction equations were validated, it is expected that new equations would be computed about every three years barring any drastic changes in the components at an earlier point.

Social Studies

World Geography: A summary is given in Table III of the regression equation used for predicting first semester ninth grade GPA in social studies Level 1, world georgraphy. Since this form of table will be used to present all the statistics of the regression equation for each subject level, it would be pertinent at this point to describe the salient features.

The predictor variables included in the step-wise regression routine are listed in the order in which they were entered in the equation. An indication of the statistical significance of the reduction of the errors of estimation due to the variable entered at each step is given by the F value. This indication is equivalent to saying that, in terms of the step-wise regression routine, the variable entered at each step had the highest F value for selection at that step among those not yet entered. When considering the significance of the F values, it is to be noted that the critical F values at the .05 and .01 probability levels are indicated by asterisks in the tables which follow. This information is relevant to an interpretation of the recorded multiple R values as well as to a decision one might make concerning the optimum number of variables needed for an efficiently predictive regression equation. Herein lies one of the considerable advantages of the step-wise regression procedure; it not only enables one to examine the relative contribution of each of several variables but also indicates which one or more are statistically significant in estimating the GPA or any other $\hat{Y}$ value. If one wished only to use the significant variables, the step-wise procedure will display at each suceeding step the regression equation unique to the set of variables at that step.

TABLE III

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTEIN PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Social Studies (Level 1, World Geography)

N = 292

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Compos	13.33	4.81	.057	.70#	.74	282.83**
2	ITED Ref	13.58	6.14	.045	.72	.73	13.97**
3	ITED Bkd NS	14.60	5.34	.030	.73	.72	5.92*
4	ITED Q Thkg	11.60	5.47	.025	.73	.72	4.52*
5	ITED Rdg SS	11.62	4.94	-.017	.73	.72	1.65
6	ITED Bkg SS	13.20	4.81	.018	.73	.72	1.30
			constant	.726			

Source of Predictor Variables: Iowa Tests of Educational Development

*p<.05
**p<.01
df=1,(N-1)-K
#Note that in this and all of the subsequent tables displaying the results of the step-wise regression routine for the combined and junior high validation samples that the multiple R listed step one is the zero order correlation of the variable entered at step one with the criterion.

The standard error of estimate is recorded in the tables primarily to enable one to see by how much it is reduced over the six steps. An interpretation of the standard error is given later in conjunction with the report of the application of the regression equation to new (check) samples.

The beta weights, rounded to three decimals, are the coefficients by which an individual's score on the variable listed on a particular step is multiplied. The sum of all the products of the individual variables is adjusted by a constant (recorded at the foot of the beta weights column) to yield, at the last step, the estimate of $\hat{Y}$ (or GPA in this study).

To look again, specifically, at the world geography equation, it is apparent that the ITED Composite score makes the most statistically signficant contribution to a reduction of the error sums of squares. While the second variable entered, Use of Sources of Information, does have a face validity relationship with the presumed objectives of the course and significantly reduced the error of estimate, the two variables more obviously relevant to the course, Reading Material in Social Studies and Understanding of Basic Social Concepts, listed at the fifth and sixth steps respectively, do not make a significant contribution to an accounting of the criterion variance. A finding such as this can be valuable to an instructional staff which may wonder about the components of this level of the course and may wish to search for possible reasons that may contribute to an explanation of the apparent irrelevance to the world geography course of these two social studies-oriented tests in the ITED battery.

In review, then, since the first four variables listed in Table III contribute to the reduction of the error sums of squares at a statistically significant level, Major Hypothesis I is rejected at each of the first four steps and is accepted for the last two steps. Thus, an optimally efficient regression equation for estimating the criterion variable, World Geography GPA, would include the first four variables in Table III.

Western Civilization: The regression equation for western civilization (Table IV) again included the ITED Composite score as the variable accounting for the largest proportion of the criterion variance, 49% (i.e., multiple R squared), while, as in the regression equation for World Geography, the subtest, Use of Sources of Information, was the second most significant variable. Although not significant in its contribution to a reduction of the error sums of squares, the test relevant to the social studies subject area, Understanding of Basic Social Concepts, at least approached significance. The six variables included in this equation account for the same proportion of the criterion variance (53%) as is accounted for by the same predictor variables in the world geography regression equation.

The disposition of Major Hypothesis I is as follows: it is rejected at steps one, two and four while it is accepted at steps three, five and six. Since the increment in the multiple R between steps two and four is so small (.71 to .72) and the standard error of estimate reflects no decrease, an optimally efficient prediction equation would be limited to the first two variables entered.

Foreign Language

To estimate the criterion variable, GPA, within levels (year) of three foreign languages, German, French and Spanish, independent

TABLE IV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Social Studies (Level 2, Western Civilization)

N = 232

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Compos	19.07	5.02	-.028	.69	.74	206.21**
2	ITED Ref	19.62	5.40	.038	.71	.72	12.35**
3	ITED Bkg SS	18.83	4.64	.064	.71	.72	3.12
4	ITED Corr Expr	17.13	4.28	.044	.72	.72	4.03*
5	ITED Q Thkg	16.88	6.12	.032	.72	.71	3.14
6	ITED Rdg SS	16.56	6.04	.025	.73	.71	2.30
			constant	.091			

Source of Predictor Variables: Iowa Tests of Educational Development

*$p<.05$
**$p<.01$
df=1,(N-1)-K

variables were drawn from the following data sources: ITED, Co-operative Reading and MLAT test batteries. Since the French II combined sample was identical with the junior high sample, junior high grades were added to the data sources just listed. The findings are presented first for French I and then followed by French II, Spanish I and II and finally German I.

French I: The multiple R's and regression coefficients, along with related statistics are tabulated for French I in Table V. The predictor variable making the most statistically significant contribution to the estimation of the criterion variable is the MLAT Total score, which alone accounts for 48% of the criterion variance. A subtest of the MLAT battery was included in the regression equation at the third step and was the final variable of the six entered to account for the criterion variance at a statistically significant level. The inclusion of the MLAT Total score as the single best predictor of French I GPA offers corroborative evidence of the predictive validity of that battery. Additional evidence of its predictive validity is derived from the inclusion of the Words in Sentences Subtest of this battery as one of the three statistically significant predictor variables.

In summary, since the F values associated with each of the first three variables entered in the regression equation are statistically significant, the null hypothesis is rejected at each of the last two. This finding provides evidence for the statement that as effective a prediction of the criterion variable is achieved with the first three variables as with all six. Corroborative evidence for this conclusion can be noted in the multiple R's beyond the third step which increase very little and in the standard errors of estimate at steps three

TABLE V

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA
WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Foreign Language (French I)

N = 55

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	MLAT Total	86.35	19.77	.028	.70	.68	49.91**
2	ITED Composite	17.33	4.19	.132	.75	.63	8.52**
3	MLAT Wds inSent	15.85	6.03	-.042	.77	.62	4.19*
4	ITED Corr Expr	16.16	4.04	.041	.78	.61	1.97
5	ITED Rdg Lit	16.75	5.88	-.042	.79	.60	1.30
6	ITED Bkd NS	17.84	4.76	-.042	.79	.60	1.37
			constant	-.298			

Source of Predictor Variables: Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test

*p<.05
**p<.01
df=1,(N-1)-K

through six which exhibit little or no decrease.

French II: The results of the regression analysis for this level are presented in Table VI. At this level, the best predictor of the criterion variable is the first semester junior high French I GPA, a finding that is similar to that reported by Hascall (1959) and cited in Chapter II, REVIEW, page 26. Evidence is again demonstrated for the predictive validity of the MLAT Total score by its inclusion in this equation as the second most significant variable. Next in order of statistical significance is the first semester junior high English GPA. It is interesting to note that the students comprising the sample were more varied in the first semester French I grades compared with the English grades received. The mean of the GPA is also lower in junior high language grades than it is for the English GPA. As can be noted in Table XLII, page 150, the mean of the criterion GPA is 3.6. One might surmise that the standards by which grades were assigned in French I in the junior high differed from those operating in the French II course in the ninth grade.

The disposition of Major Hypothesis I is as follows: the null hypothesis of no significant reduction in the errors of estimation at each of six steps of the step-wise regression routine due to the variable entering at each step is rejected over the first three steps and accepted over the last three. Since some increase in the multiple R between steps three and six occurs (from .72 to .77) while a corresponding decrease in the standard error of estimate (from .64 to .61) can be observed, the decision whether to limit the prediction equation only to the first three signficant variables would be determined by such local factors as the clerical and data processing costs associated with

TABLE VI

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Foreign Language (French II)

N = 52

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	JrHi Lang GPA	3.30	1.28	.420	.62	.72	31.77**
2	MLAT Total	104.49	16.76	.010	.68	.67	8.50**
3	JrHi Eng GPA	4.33	.87	-.244	.72	.64	5.71*
4	Co-opRdg Vocab	59.09	6.86	.034	.74	.63	3.16
5	ITED Bkg SS	21.04	4.58	-.064	.75	.62	1.65
6	ITED Rdg NS	21.59	5.32	.048	.77	.61	3.77
			constant	.576			

Source of Predictor Variables: Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test
Junior High Subject GPA

---

*$p<.05$
**$p<.01$
df=1,(N-1)-K

adding the last three variables. The additional 7% of the variance explained by the last three variables appears to be substantial but the lack of statistical significance associated with the last three variables indicates that this 7% may not be reliable enough to justify the costs of inclusion.

Spanish I: The predictive validity of the MLAT Total score is again demonstrated in the findings reported in Table VII for Spanish I. Among the variables selected by the step-wise regression routine for predicting the criterion variable, the MLAT Total clearly makes the most statistically significant contribution to the reduction in the errors of estimation (42% of the variance accounted for). Two other variables from the MLAT battery (Words in Sentences and Phonetic Script), while not significant in their account of the criterion variance, were among the six out of a pool of twenty-one variables to be selected, a finding that offers further evidence for some predictive validity of these subtests in the MLAT battery. The ITED subtest, Correctness of Expression and the Co-op Reading Comprehension test, together with the MLAT Total score emerge as the three variables of the six entered that contribute, to a statistically significant degree, to the reduction in the errors of estimation.

The disposition of Major Hypothesis I is as follows: the hypothesis of no significant reduction in the error sums of squares at each of six steps of the step-wise regression routine is rejected at steps one through three and is accepted at steps four through six. Therefore an optimally efficient regression equation that draws variables from the indicated sources would include the first three variables entered for estimating the Spanish I criterion variable.

TABLE VII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Foreign Language (Spanish I)

N = 138

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	MLAT Total	83.88	21.46	.024	.65	.88	97.14**
2	ITED Corr Expr	14.58	4.06	.120	.72	.81	26.46**
3	Co-op RdgCompre	49.88	7.91	-.029	.73	.79	6.59*
4	MLAT Wds in Sent	14.89	5.95	-.023	.74	.79	1.87
5	ITED Rdg NS	14.51	5.65	.020	.74	.79	1.39
6	MLAT Ph Script	20.93	3.75	.027	.74	.79	1.18
			constant	.292			

Source of Predictor Variables: Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test

---

*p<.05
**p<.01
df=1,(N-1)-K

Spanish II: Although entered as the second variable in the step-wise regression routine, the MLAT Total, as can be noted in Table VIII, again demonstrates its relevance as a predictor of success in foreign languages. The MLAT Total score, along with the ITED subtest, Use of Sources of Information, account for 53% of the criterion variance. A subtest of the MLAT battery, Words in Sentences, again appears among the six variables selected although its contribution to the reduction of the error sums of squares is not statistically significant.

The disposition of Major Hypothesis I is as follows: the hypothesis of no significant reduction in the error sums of squares at each of six steps of the step-wise regression routine is rejected at steps one and two and is accepted at steps three through six. Therefore, the most efficient regression equation which employs variables from the sources indicated would include the first two variables entered for estimating the Spanish II criterion variable.

German I: In Table IX are displayed the statistics for predicting the criterion variance in German I. It can be observed that two variables from the ITED battery, Correctness of Expression and General Vocabulary, along with one from the MLAT data source, Paired Associates, comprise three out of the six variables entered that contribute most to an estimate of the criterion variable. Evidence is again present for the validity of the MLAT battery in accounting for a foreign language criterion variance.

The disposition of Major Hypothesis I is as follows: the hypothesis of no significant reduction in the errors of estimation at each of six steps of the step-wise regression routine is rejected at steps one, two and four while it is accepted at steps three, five and six.

TABLE VIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Foreign Language (Spanish II)

N = 51

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Ref	20.43	5.08	.071	.68	.67	41.31**
2	MLAT Total	97.75	17.10	.010	.73	.63	8.04**
3	ITED Composite	19.53	4.81	-.039	.75	.61	3.12
4	MLAT Wds in Sent	18.12	5.00	.038	.76	.61	1.24
5	ITED Bkd NS	19.33	5.41	.037	.76	.61	.61
6	ITED Rdg SS	16.73	6.16	.038	.77	.61	1.34
			constant	.083			

Source of Predictor Variables: Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test

**p<.01
df=1,(N-1)-K

TABLE IX

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Foreign Language (German I)

N = 41

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Corr Espr	15.15	4.36	.108	.66	.66	30.02**
2	MLAT Prd Assoc	13.10	4.76	.056	.72	.62	6.62*
3	ITED G Vocab	17.32	3.06	.010	.73	.61	1.60
4	ITED Bkg SS	15.93	4.66	-.063	.77	.59	4.27*
5	MLAT Spell Clu	9.34	5.19	-.063	.79	.57	2.87
6	MLAT Total	82.80	21.07	.014	.80	.57	1.98
			constant	-.081			

Source of Predictor Variables: Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test

*p<.05
**p<.01
df=1,(N-1)-K

Because there is a 7% difference in the amount of variance explained between steps two and four as well as a small but steady decrease in the standard error of estimation, an efficient prediction equation would include the first four variables entered. The third variable in this case would have to be included despite its lack of significance because the regression equation as computed at the fourth step would include it.

<u>English</u>

<u>Basic Composition:</u> The regression equation and related statistics for this Level 1 English course are tabulated in Table X. Although the multiple R is increased from .47 to .64 with the addition of the six variables to the regression equation, only the first two variables entered, Use of Sources of Information and Quantitative Thinking, contribute significantly to the reduction of the error sums of squares; together they account for 28% of the criterion variance. Notably absent from the six listed variables of the ITED battery, a variable one would expect to find, is Correctness of Expression. Two variables tapping interpretive reading skills, Reading Natural Science Materials and Reading Literary Materials, are included but their predictive contribution is nonsignificant.

The null hypothesis of no significant reduction in the error sums of squares at each of six steps of the step-wise regression routine due to the variable entering is rejected at steps one and two and accepted at each of the last four steps. Therefore, while the amount of the criterion variance explained increases by 13% from steps two to step six, the decrease in the standard error of estimate is small (from .61 to .57). Such a small decrease supports the conclusion that an efficient prediction equation for estimating the basic composition

TABLE X

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

English (Level 1, Basic Composition)

N = 67

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Ref	11.01	4.72	.056	.47	.63	18.18**
2	ITED Q Thkg	9.42	4.49	.048	.53	.61	5.93*
3	ITED Rdg NS	10.85	4.69	-.067	.57	.60	3.39
4	ITED Rdg Lit	11.96	4.12	.046	.60	.59	3.18
5	ITED Bkd NS	13.16	4.89	-.049	.63	.58	3.71
6	ITED Compos	11.48	3.83	.057	.64	.57	1.51
		constant		1.824			

Source of Predictor Variables: Iowa Tests of Educational Development

*p<.05
**p<.01
df=1,(N-1)-K

criterion variable would include the first two variables entered.

Composition: In Table XI are presented the variables entered in the regression equation at the sixth step for estimating the criterion variable in this second level course in English. The predictive validity of the ITED subtest most relevant to English, Correctness and Appropriateness of Expression, is demonstrated. This variable, along with two others, Use of Sources of Information and Ability to do Quantitative Thinking (also the first two predictors in the basic composition Level 1 equation), are the only variables among the six entered that significantly account for the criterion variance (55% is accounted for).

The disposition of Major Hypothesis I is as follows: the null hypothesis of no significant reduction in the error sums of squares at each of the six steps is rejected at steps one, two, and three while it is accepted at each of the last three steps. Therefore, an optimally efficient preduction equation for estimating level two English GPA, using the ITED battery as the data source, would include the first three variables entered.

Composition (Honors): The variables included at the sixth step for predicting criterion variables of this Level 3 English course are listed in Table XII. The ITED subtest, Correctness and Appropriateness of Expression, entered first by the step-wise procedure, alone accounts for 34% of the criterion variance which is substantial evidence of its predictive validity for this level of the English course. The second variable entered, Use of Sources of Information, although not statistically significant in its reduction of errors of estimation, makes its third appearance as a predictor variable in the English course. The third variable entered, Understanding of Basic Social Concepts, makes

TABLE XI

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

English (Level 2, Composition)

N = 177

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Ref	17.21	5.37	.051	.67	.68	146.02**
2	ITED Corr Expr	15.35	3.97	.072	.73	.64	26.66**
3	ITED Q Thkg	14.28	5.67	.028	.74	.63	7.21**
4	ITED G Vocab	17.80	3.58	.028	.74	.62	1.57
5	ITED Bkd NS	17.15	5.06	-.022	.75	.62	1.86
6	ITED Rdg NS	15.74	5.31	.015	.75	.62	1.23
			constant	.400			

Source of Predictor Variables: Iowa Tests of Educational Development

**p<.01
df=1,(N-1)-K

TABLE XII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

English (Level 3, Composition Honors)

N = 71

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Corr Expr	20.94	3.29	.065	.58	.46	34.61**
2	ITED Ref	24.24	3.64	.043	.60	.45	3.31
3	ITED Bkd SS	22.55	3.98	-.038	.64	.44	4.49*
4	ITED Rdg Lit	22.04	3.75	.039	.66	.43	3.31
5	ITED G Vocab	23.25	3.44	-.035	.67	.43	2.36
6	ITED Rdg SS	21.87	4.49	.019	.68	.42	1.22
			constant	1.887			

Source of Predictor Variables: Iowa Tests of Educational Development

*p<.05
**p<.01
df=1,(N-1)-K

its first appearance as a predictor variable in English and in this equation its accounting of the criterion variance is statistically significant.

The null hypothesis of no significant reduction in the error sums of squares at each of the six steps is rejected at steps one and three while it is accepted at steps two, four, five, and six. Therefore, while the amount of standard error of estimation decreases very little from steps one to three (.46 to .44), the amount of criterion variance explained increases by 7%, leading to the conclusion that an efficient prediction equation would include the first three variables entered.

Mathematics

General Mathematics: Predicting the criterion variable in this level of mathemetics with the ITED battery as the data source was of limited success as can be observed by an inspection of the multiple R values listed in Table XIII. The first two variables entered by the step-wise procedure, Correctness and Appropriateness of Expression and the Ability to Do Quantitative Thinking, together account for only 18% of the criterion variance. These two variables are also the only ones of the six entered that significantly contribute to the reduction of the error sums of squares. When the standard error of estimation at step two, .97, is compared with the criterion variable's standard deviation (Table XVII, page 102) of 1.1, it is readily apparent that very little could be expected by employing this regression equation. A report of the application of this equation to a new sample is presented in Table XVIII, page 104.

The disposition of Major Hypothesis I is that it is rejected at

TABLE XIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Mathematics (Level 1, General Math)

N = 137

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Corr Expr	10.93	3.71	.063	.35	.99	18.84**
2	ITED Q Thkg	8.66	3.54	.061	.42	.97	8.96**
3	ITED Ref	10.42	4.67	.032	.43	.96	1.70
4	ITED Bkd SS	10.93	3.97	-.028	.44	.97	.89
5	ITED Rdg SS	10.07	4.03	-.019	.44	.97	.24
6	ITED Composite	10.78	3.51	.031	.44	.97	.33
			constant	1.426			

Source of Predictor Variables: Iowa Tests of Educational Development

**p<.01
df=1,(N-1)-K

steps one and two and is accepted at each of the last four steps. While the first two variables make a statistically significant contribution to a reduction in the error sums of squares, nevertheless, due to the very limited predictive validity of the variables entered, no conclusion is reached here about an optimum set of predictor variables drawn from the ITED data source for estimating the criterion variance.

Algebra: A brighter predictive picture emerges for this Level 2 mathematics course (Table XIV). Of the six variables entered in the regression equation, five significantly contribute to the reduction in the errors of estimation and together explain 49% of the criterion variance. Demonstration of the predictive validity of the ITED subtest, Ability to Do Quantitative Thinking, is effectively offered by its inclusion at step one, by the 38% of the criterion variance it explains and by its highly significant status as evidenced by the F value. There is an interesting variety in the kinds of skills represented by the five statistically significant variables entered in the equation. The appearance at step two of the subtest, Use of Sources of Information, suggests that it is either a considerably versatile predictor variable or that the algebra instructors were rather unusual in their teaching methododology or that this test assesses something akin to general academic ability.

The disposition of Major Hypothesis I is that it is rejected at steps one through five and is accepted at step six. It is therefore concluded that for predicting algebra, using the ITED as a data source, an optimum prediction equation would include the first five variables. It should be noted, however, that this conclusion must be considered with reference to the slight increase in the multiple R and the

TABLE XIV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Mathematics (Level 2, Algebra)

N = 318

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Q Thkg	14.85	4.95	.084	.62	.75	200.58**
2	ITED Ref	17.44	5.20	.045	.67	.72	36.56**
3	ITED Corr Expr	15.74	3.88	.048	.68	.71	8.22**
4	ITED G Vocab	17.88	3.68	-.048	.69	.70	9.74**
5	ITED Rdg NS	16.18	5.29	.025	.70	.69	4.09*
6	ITED Rdg Lit	16.16	4.66	-.011	.70	.69	.69
			constant	.962			

Source of Predictor Variables: Iowa Tests of Educational Development

*p<.05
**p<.01
df=1,(N-1)-K

correspondingly slight decrease in the standard error of estimate from step three to step five. The slight gain in accounting for the criterion variance by including five variables as against three in a prediction equation would depend on whether the cost of adding the fourth and fifth variables was worth it.

Algebra (Honors): Success in the estimation of the algebra (honors) criterion variable by means of predictor variables drawn from the ITED battery falls between that of general mathematics and algebra, as can be noted in Table XV. The first two variables entered, The Ability to Do Quantitative Thinking and Correctness and Appropriateness of Expression are the only ones that contribute, at a significant level, to a reduction in the error sums of squares. Together, they explain 27% of the criterion variance. It will be recalled that these two variables also appeared, significantly, in the Level 1 and Level 2 mathematics equations.

The disposition of Major Hypothesis I is that it is rejected at the first two steps and accepted at each of the last four. While the multiple R steadily increases from .52 to .59 over steps two to six (an 8% increase in the amount of criterion variance explained), the standard error of estimate decreases very little, offering additional support for the conclusion that an optimally efficient prediction equation would be comprised of the regression equation computed at step two which includes the first two variables entered.

Summary of Major Hypothesis I, Combined Validation Sample

To facilitate a review of results just presented, Table XVI has been prepared. In this table, the variables selected by the stepwise routine for each instructional level that significantly

TABLE XV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Combined Group Validation Sample

Mathematics (Level 3, Algebra Honors)

N = 70

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	F
1	ITED Q Thkg	22.96	4.73	.070	.47	.61	19.20**
2	ITED Corr Expr	20.40	3.59	.070	.52	.60	4.83*
3	ITED Rdg Lit	21.07	3.85	-.060	.55	.59	2.46
4	ITED Rdg SS	21.00	4.78	.448	.56	.58	1.79
5	ITED Ref	24.70	2.78	-.047	.58	.58	1.90
6	ITED Bkd NS	22.96	3.56	-.029	.59	.58	1.47
			constant	3.341			

Source of Predictor Variables: Iowa Tests of Educational Development

*p<.05
**p<.01
df=1,(N-1)-K

TABLE XVI

STATISTICALLY SIGNIFICANT VARIABLES INCLUDED IN THE STEP-WISE REGRESSION EQUATION FOR EACH INSTRUCTIONAL LEVEL OF THE SUBJECT AREAS

Combined Group Validation Sample

Subject Area & Level	Data Source	Significant Predictor Variables at Steps: 1	2	3	4	5	6
Social Stu World Geog	ITED	Compos	Ref	Bkd NS	Q Thkg		
West Civil	ITED	Compos	Ref		CorrExp		
For Lang French I	ITED MLAT Co-opRdg	MLAT Total	ITED Compos	MLAT Wds in Sent			
French II	ITED MLAT Co-opRdg JrHi GPA	LanGPA	MLAT Total	Eng GPA			
Spanish I	ITED MLAT Co-opRdg	MLAT Total	ITED Corr Exp	Co-opRdg Compre			
Spanish II	ITED MLAT Co-opRdg	ITED Ref	MLAT Total				
German I	ITED MLAT Co-opRdg	ITED CorrExp	MLAT Prd Assoc		ITED Bkg SS		
English Basic Comp	ITED	Ref	Q Thkg				
Comp	ITED	Ref	CorrEx	Q Thkg			
Comp (Hon)	ITED	CorrExp		Bkd SS			

TABLE XVI (Continued)

Subject Area & Level	Data Source	Significant Predictor Variables at Steps:					
		1	2	3	4	5	6
Mathematics							
Gen Math	ITED	CorrExp	Q Thkg				
Algebra	ITED	Q Thkg	Ref	CorrExp	G Voc	RdgNS	
Alg(Hon)	ITED	Q Thkg	CorrEx				

contribute to a reduction in error sums of squares are displayed. This finding is in keeping with Nunnally's statement (1967, p. 162) that:

> . . .beyond (the third variable entered), adding additional tests produces only small increases in the multiple correlation.

Immediately apparent in this table is the fact that no variables at the sixth step are significant, only one at the fifth, and three at the fourth step. It is possible in this table to observe how many times, across instructional levels, certain variables were entered, as for example ITED Use of Sources of Information (Ref) which was selected six times. One generalization that emerges from this summary table is that by the third step, based on the data sources recorded and for the population involved, most of the significant estimated criterion variance had been accounted for.

Major Hypothesis II

This hypothesis states that there are no significant differences within achievement levels in the multiple R's between the validation and check samples of English and mathematics (the only two subject areas, in the combined sample, in which the regression equations were evaluated by applying them to a new--check--sample). The statistical procedure appropriate to testing this hypothesis is the $\underline{z}$ test for the significance of differences between R's derived from independent samples (Wert et al., 1954, p. 296).

The means and standard deviations of the criterion variables, the multiple correlation coefficients and the standard error of estimate of the predicted GPA are presented for validation and check samples in Table XVII. Also included in this table is an R_c value which is the validation sample multiple R corrected for bias (Guilford, 1965, p. 401). As Guilford points out:

> The multiple R represents the maximum correlation between a dependent variable and a weighted combination of independent variables. . .The multiple R is. . .an inflated value. It is a biased estimate of the multiple R in the population. If we were to apply the same regression weights in a new sample and to correlate predicted $\hat{X}$ values with obtained X values, we should probably find that the (check sample multiple R) would be smaller than (the validation sample multiple R). It is desirable, therefore, to find some means of estimating a parameter R which gives a more realistic picture of the general situation. A common way of 'shrinking' R to a more probable population value is (by computing the R_c value).

When the check sample multiple R's (computed as per procedures presented by Wert et al., 1954, p. 240) are reviewed (Table XVII) and compared with the corresponding validation sample multiple R's as well as with the R_c values, the general conclusion is that the combined sample regression equations within achievement levels of English and mathematics subject areas held up very well. Supportive evidence can be observed in the tabled standard errors of estimate (computed as per a procedure presented by Guilford, 1965, p. 373) which are of approximately the same magnitude, proportionally, with respect to the check sample criterion variable standard deviation as are the validation sample standard errors with respect to their corresponding criterion standard deviations. Another way of interpreting the standard error of estimate is that the closer its value approaches that of the criterion standard deviation, the less accurate is the estimated GPA. A simple subtraction of the check sample standard error values from the standard deviation values indicates that the most accurate estimate of the criterion variable is obtained within the composition and algebra achievement levels. The size of the samples of these two levels probably contributes to this finding.

TABLE XVII

MEANS AND STANDARD DEVIATIONS OF CRITERION VARIABLES, MULTIPLE CORRELATION COEFFICIENTS AND STANDARD ERROR OF THE PREDICTED FIRST SEMESTER NINTH GRADE ENGLISH AND MATHEMATICS GPA

Combined Group Validation and Check Samples

Subject Area & Level	Sample Size		Criterion Variable (First Semester GPA)				Multiple Correlation Coefficient			Standard Error of Estimate of Predicted GPA	
			Mean Sample		Std Dev Sample		Sample			Sample	
	Validation	Check	Val.	Check	Val.	Check	Val. R	Val. Rc	Check R	Val.	Check
ENGLISH											
Basic Comp	67	71	2.7	2.4	.71	.83	.64	.57	.63	.57	.65
Composition	177	191	3.1	3.0	.92	.96	.75	.74	.70	.62	.69
Comp (Honors)	70	82	3.9	3.8	.55	.61	.68	.62	.67	.42	.45
MATHEMATICS											
Gen Math	137	113	2.8	2.8	1.1	.90	.44	.36	.49	.97	.78
Algebra	318	444	3.1	2.8	.96	.99	.70	.69	.62	.69	.78
Algebra (Hon)	70	45	4.3	4.4	.68	.64	.59	.50	.54	.58	.54

Source of Predictor Variables: Iowa Tests of Educational Development

To evaluate Major Hyothesis II, the z test of the difference between validation and check sample multiple R's was computed where the differences in these R's was large enough to warrant the test. In this test, a standard normal deviate, z, is computed by transforming each multiple R into its equivalent Z value, then calculating the standard error of the difference between the two Z's. This value is referred to the normal table where z at the .025 probability level equals 1.96. If the observed z is smaller than 1.96, the null hypothesis is accepted.

The disposition of Major Hypothesis II (Table XVIII), as a result of the application of the z test, is that it was accepted for all cases in which it was evaluated. Therefore, it is concluded that the regression equations developed on the combined group validation sample and then evaluated on a new (check) sample have functioned effectively. It is further concluded that the few departures from the normality assumption recorded in Table II (page 70) for certain of the ITED variables did not adversely affect the effectiveness of the regression equations when applied to new samples.

Junior High Group Sample

Prior to a presentation of the results of the regression analyses for the junior high group sample, a report is given of the test of the assumptions underlying the use of the correlation/regression model for this sub sample. As noted in Chapter III, page 58, the following assumptions will be tested for the ITBS and Lorge data of the junior high group sample:

TABLE XVIII

SUMMARY OF THE z TEST OF THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN VALIDATION AND CHECK SAMPLE MULTIPLE CORRELATION COEFFICIENTS IN SELECTED SUBJECT AREAS

Combined Group

Subject Area and Level	Multiple Correlation Coefficient		z	p	Disposition of Hypothesis
	Validation Sample	Check Sample			
English Composition	.75	.70	1.00	$> .05$	Accepted
Mathematics					
General Math	.44	.49	.50	$> .05$	Accepted
Algebra	.70	.62	1.94	$> .05$	Accepted

$\alpha .05 = 1.96$

Source of Predictor Variables: Iowa Tests of Educational Development

1. that the independent variables are normally distributed.
2. that homoscedasticity of the criterion variable (GPA) prevails.
3. that a linearity of the regression line obtains.

The first of these assumptions, normality of distribution, will be tested for all of the ITBS and Lorge subtest score distributions of the total junior high group validation sample. The other two assumptions, homoscedasticity and linearity of regression, will be tested for the distribution of each of the six predictor variables appearing in the sixth step of the step-wise regression routines for the three achievement levels within the junior high group English and mathematics subject areas. A report of the results of the test of these assumptions will be found in what follows as well as on page 125.

Normality of distribution assumption: The null hypothesis under test is: the observed distribution of ITBS and Lorge subtest score frequencies do not depart significantly from the theoretical normal distribution.

A test of this assumption is accomplished by applying the chi squared test of goodness-of-fit (Guilford, 1965, pages 243-247). Summarized in Table XIX are the results of the application of the chi squared goodness-of-fit test to all eighteen ITBS and Lorge subtest score distributions.

It can be noted that for only four of the eighteen subtests (ITBS Vocabulary, L_4 Usage, A_2 Arithmetic Problems and Lorge Non-Verbal) is the null hypothesis rejected, indicating that the normal distribution assumption is largely upheld. The inference is drawn that fourteen out of eighteen ITBS and Lorge test score distributions comprising the total

TABLE XIX

SUMMARY OF THE RESULTS OF THE GOODNESS-OF-FIT CHI SQUARED TEST OF THE NORMALITY ASSUMPTION AS APPLIED TO THE DISTRIBUTION OF ITBS AND LORGE-THORNDIKE SCORES USED IN THE MULTIPLE REGRESSION AND MULTIPLE DISCRIMINANT ANALYSES

Junior High Group Validation Sample

Variables	Observed Chi Sq.	Tabled Chi Sq.	df	p.05*	PMAX*	Disposition of Hypothesis
ITBS						
V Vocab.	31.68	15.51	8	<		Rejected
R Rdg.	1.27	14.07	7		.98	Accepted
L_1 Sp.	4.90	14.07	7		.50	Accepted
L_2 Cap.	8.00	14.07	7		.30	Accepted
L_3 Punc.	12.24	15.51	8		.10	Accepted
L_4 Usage	14.48	14.07	7	<		Rejected
L_t Lan. Total	3.68	14.07	7		.80	Accepted
W_1 Map Rdg.	9.72	14.07	7		.20	Accepted
W_2 Rdg. Grs. & Tables	6.27	14.07	7		.50	Accepted
W_3 Kn. & Use Ref. Mat.	7.51	15.51	8		.30	Accepted
W_t Work Study Skills Total	7.23	14.07	7		.30	Accepted
A_1 Arith. Con.	3.24	14.07	7		.80	Accepted
A_2 Arith. Prob.	17.00	14.07	7	<		Rejected
A_t Arith. Total	5.77	14.07	7		.50	Accepted
C Composite	3.64	14.07	7		.80	Accepted
Lorge Thorndike						
V Verbal	17.15	19.68	11		.10	Accepted
Non-Verbal	21.27	18.68	11	<		Rejected
Total	18.14	18.31	10		.05	Accepted

*$\propto$ set at .05 as a minimum level for accepting/rejecting the null hypothesis. However, to permit the observed chi squared values to indicate by what proportion of chances in 100 samples the observed chi squared value could be obtained, the p max value is also recorded.

junior high sample data source are normally distributed. It is to be noted, however, that while some of the ITBS and Lorge test score distributions comprising the levels within each subject area may not be normally distributed, it may be that for some of the levels (e.g., English and mathematics, levels 3) the relatively small N's and the negative skew (an excess of high scores) as well as a truncation of the distribution may well contribute to assymetry of the distributions within levels. What effect any possible non-normal distribution of the independent variables within levels may have on the generalizability of the regression equations (i.e., the stability of the estimated criterion variables) will be evaluated in this study by application of the regression equations to new (check) samples.

Predicting Achievement Within Levels of Four Subject Areas

The results of this application of the regression equations yielded by the step-wise routine described in Chapter III, page 55, are reported first for social studies followed by foreign languages, English and mathematics. For each subject area, the results are reported first for the predictor variables derived from the ITBS and Lorge data sources, followed immediately by the results for the instances when data sources additional to the ITBS and Lorge were employed. The disposition of Major Hypothesis I will be given for each instructional level at the end of the presentation and discussion of the results of the variables entered in the regression equation by the step-wise procedure.

Social Studies

World Geography: A summary is given in Table XX of the ITBS and Lorge variables that were entered at each step of the step-wise regression routine for predicting first semester ninth grade GPA in

TABLE XX

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTEIN PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Social Studies (Level 1, World Geography)

N = 124

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS Vocab	44.98	8.96	.051	.61	.74	.61	65.88**
2	ITBS W_3	45.03	8.04	.036	.63	.73	.55	5.60*
3	ITBS Rdg	44.89	8.39	-.020	.64	.73	.49	1.51
4	Lorge N.V.	116.27	13.85	.008	.65	.72	.41	1.80
		constant		-1.001				

Source of Predictor Variables: Iowa Test of Basic Skills
Lorge Thorndike Intelligence Test

*p<.05
**p<.01
df=1,(N-1)-K

world geography.

The variable listed at step 1, ITBS Vocabulary, was entered into the step-wise regression equation first, since, of all the variables in the data source, it contributed most to the reduction of the error sums of squares (residuals). It is observed that the step-wise regression routine was terminated at step 4 since the computer program criterion set for inclusion in the regression equation was not met by any of the remaining variables in the data source.

When examining the significance of the F values, it can be seen that, for predicting Level 1 GPA, as accurate an estimate would be yielded by employing the predictor variables at step two, Vocabulary and Knowledge and Use of Reference Material as would be yielded by all four variables. The disposition of Major Hypothesis I, which states that there is no significant reduction in the error sums of squares due to the variable entering, is rejected at the first two steps and accepted over the last two.

Notably absent from the predictor variables in Table XX are the ITBS Map Reading and Reading Graphs and Tables subtests. These are variables one would expect to appear among the variables in the equation. A characteristic common to the variables that were included is a heavy concentration of verbal ability (e.g., vocabulary), an interesting finding in view of the fact that this is the level of the social studies course into which the students thought to be of lesser aptitude for learning were registered. A finding such as this can be pertinent to a staff as it considers what it seems to take to be successful, GPA-wise, in a particular instructional level of a subject area.

<u>Western Civilization</u>: Turning now to Table XXI in which similar

TABLE XXI

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Social Studies (Level 2, Western Civilization)

N = 136

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS Composite	55.76	7.95	.064	.67	.77	.67	109.44**
2	ITBS L_3	54.10	9.01	-.027	.69	.76	.27	5.06*
3	ITBS A_2	54.07	9.51	-.043	.70	.75	.35	4.15*
4	ITBS W_t	55.17	8.62	.027	.71	.74	.64	3.22
5	ITBS A_t	54.43	9.19	.038	.72	.73	.50	3.63
6	ITBS L_2	54.57	8.85	.223	.73	.72	.46	4.49*
		constant		-1.249				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*p<.05
**p<.01
df=1,(N-1)-K

data relevant to the Level 2 section of social studies, western civilization, is recorded, it is immediately apparent that a greater variety of variables are included among the six best predictors; two of the variables are from the Language subtest, two are arithmetical in content, one is a total of the Work Study Skills area and one is a composite of all the ITBS subtests. The Composite variable, which entered first with the highest F value, represents a kind of general achievement factor.

Although different variables were entered in the regression equations for each social studies instructional level, an approximate comparison can be made of their means. It is apparent that the groups are different. Within each group, achievement status, on these variables at least, appears to be quite even.

It may be noted in Table XXI and the tables to follow, that the zero order correlation coefficient, r with criterion (reported when available), does not necessarily decrease in relation to the F value. This observation points up the fact that a predictor variable's correlation with the criterion variable is but one factor in determing its inclusion; its correlation with the other predictor variables is also a factor. Thus, if one simply inspects a correlation matrix and selects the variables with the highest r's in decreasing order, one might be led to choose variables which would not necessarily contribute to the most effective reduction of the errors of estimation.

The disposition of Major Hypothesis I is that, as a function of the variables entered, it is rejected at steps one, two, three, and six while it is accepted at steps four and five. Since the multiple R's do increase and since the standard errors of estimate do decrease to some extent, a prediction equation might include all six variables entered. Application of the above regression equation to check sample data was not made because the Level 2 grouping in social studies no longer prevailed during the check sample year.

Foreign Languages

In the junior high sample, regression equations were computed for two levels of two foreign languages: French I, French II, Spanish I, and Spanish II. Although called levels, they are not levels as defined for the other three subject areas under study, since, as pointed out in Chapter III, METHOD, page 46 , students enrolled in the second level of French or Spanish were those who had taken the first level (year) while in the junior high.

In predicting GPA within the two levels of French and Spanish, attention was focused on the contribution variables derived from different data sources might make to the prediction of the criterion variable when they were present in the data pool in different combinations. Of particular interest was the contribution variables in the MLAT data source might make to the reduction of the errors of prediction. Following the presentation of the results in which the ITBS and Lorge were the data source will be the results produced by additional data sources. Application to check sample data of the regression equations derived from the ITBS and Lorge data source will then be displayed.

French I: In Table XXII, the multiple R's and beta weights along with related statistics, are tabulated for the predictor variables derived from the ITBS and Lorge data sources. It is interesting to observe that three of the six variables included in the regression equation are heavily concentrated in verbal abilities (ITBS Vocabulary,

TABLE XXII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (French I)

N = 27

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weight	Multiple R	S.E. of Estimate	F
1	ITBS Vocab	47.48	8.92	-.116	.66	.72	18.91**
2	ITBS L_1	50.33	7.43	.056	.70	.70	2.69
3	ITBS L_4	47.07	8.41	-.039	.74	.67	2.87
4	Lorge Verbal	117.22	12.64	-.033	.77	.65	2.57
5	Lorge N.V.	118.15	11.07	.028	.79	.63	2.13
6	ITBS Rdg	47.04	9.40	-.036	.81	.63	1.43
		constant		-1.684			

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

**p<.01
df=1,(N-1)-K

Lorge Verbal IQ and ITBS Reading). This is a finding that replicates the results reported by other investigators (Pimsleur, et al., 1962: see Chapter II, REVIEW, page 25). Of the variables entered in the step-wise regression routine, only the first variable, ITBS Vocabulary, contributed significantly to the reduction in the errors of prediction.

The disposition of Major Hypothesis I is that it is rejected at step one and accepted at each of the remaining five steps. Although the proportion of criterion variance explained between steps one and six increased by 20% while the standard error of estimate decreases substantially (from .72 to .63), nevertheless, the additional contribution in the reduction of the error sums of squares made by the five variables beyond the first step is not statistically significant. This suggests that using the first variable, Vocabulary, in predicting GPA would be as effective as using all six. Such a finding is suggestive only, because of limited sample size. The sample size also placed constraints on the use of additional sources of data.

French II: The results of the regression analysis for this level are presented in Table XXIII. A greater variety in the kinds of variables comprising the sixth step of the equation is apparent. Again, as in French I, the first variable entered was the only one of the six that contributed significantly to the reduction of the error sums of squares. Of some interest is a comparison in the predictor variable means of French I and II, the larger mean values being associated with French II. This finding suggests the possibility that the students comprising the French I group may have received suggestions as seventh graders from a variety of sources to defer taking French until they were in high school; the assumption may have been that such deferment was desirable

TABLE XXIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA
WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (French II)

N = 52

Step	Predictor Variable	$\bar{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS L_t	57.21	8.85	.073	.53	.78	.53	19.95**
2	ITBS Vocab	59.60	9.15	.015	.55	.78	.27	1.39
3	Lorge N.V.	129.16	11.89	.015	.56	.78	.38	.92
4	ITBS W_3	58.13	8.11	-.016	.57	.78	.26	.80
5	ITBS L_2	56.46	8.50	-.014	.58	.78	.34	.50
6	ITBS L_3	56.73	9.17	-.016	.58	.79	.42	.39
		constant		-.793				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

**$p < .01$
df=1,(N-1)-K

because the students might not succeed in the French I course as given in the junior high school.

Inspection of the F values indicates that Major Hypothesis I is again rejected at step one only while it is accepted at each of the last five steps. The increment in the proportion of criterion variance explained between steps one and six, 6%, is less than was the case for French I; the standard error of estimate reflects no decrease. Thus, it is readily concluded that an optimally efficient prediction equation using the ITBS and Lorge batteries as the data source would include the first variable entered, ITBS Language Total.

When the predictor variables for French II were derived from data sources additional to ITBS and Lorge (ITED, Co-op Reading, MLAT and junior high subject GPA's), the resulting regression equation and associated statistics are as reported in Table XXIV. It is apparent from the multiple R's and the standard errors of estimate that, with the additional data sources, more of the criterion variance is accounted for. Not surprisingly, the variable contributing most to the reduction of the errors of estimation is the junior high French I junior high GPA which, by itself, accounts for nearly half of the criterion variance ($R^2 = .45$). A heavy concentration of verbal abilities is represented by four of the six variables. While not as significant in its reduction of the errors of estimation, the presence of the MLAT Total score demonstrates its contribution to the prediction of the criterion variable. It is noted that the Lorge Verbal IQ score appears again, although its contribution is not significant, statistically.

Inspection of the F values indicates that Major Hypothesis I is

TABLE XXIV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (French II)

N = 52

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	Jr Hi Lang GPA	3.29	1.27	.509	.61	.73	.61	20.91**
2	MLAT Total	104.36	16.92	.011	.68	.68	.58	8.23**
3	Jr Hi Eng GPA	4.33	.88	-.304	.72	.65	.25	5.67*
4	Co-op Rdg Vocab	59.06	6.81	.071	.75	.64	.43	3.62
5	ITBS Vocab	59.60	9.15	-.026	.76	.62	.27	3.27
6	Lorge Verbal	129.85	9.46	-.017	.77	.62	.20	1.68
		Constant		1.570				

Source of Predictor Variables: Iowa Tests of Basic Skills
Iowa Tests of Educational Development
Lorge Thorndike Intelligence Test
Modern Language Aptitude Test
Co-operative Reading Test
Junior High Subject GPA

*p<.05
**p<.01
df=1,(N-1)-K

rejected at each of the first three steps and is accepted at each of the last three steps. Therefore, an optimally efficient prediction equation would include the first three variables entered. Caution in generalizing beyond the present population is particularly necessary because the multiple R is probably quite inflated due to the large number of variables in the data pool in relation to the sample size. The multiple R at step three, corrected for bias (Guilford, 1965, p. 401), reduces to zero. Such statistical shrinkage may be too severe as will be indicated later (p. 149) where the ratio of the number of variables to sample size (Level 1 English and mathematics) led to a zero value for the multiple R but, when the regression equation was applied to a new sample, the check sample multiple R held up well. The equation reported here was not evaluated on a check sample owing to the fact that the MLAT and Co-op Reading test batteries were discontinued during the second year of the study.

Spanish I: The independent variables and related statistics derived from the ITBS and Lorge data sources for predicting Spanish I GPA are listed in Table XXV. The predictor variables entered in this equation tap skills in three areas: language, mathematics and ability to define words. The appearance of the two arithmetic subtests in the regression equation is similar to Dunn's finding (1959) cited earlier in Chapter II, REVIEW, page 38 based on a college sample.

The null hypothesis of no significant reduction in the error sums of squares due to the variable entering is rejected at steps one, two, and four and is accepted at steps three, five, and six. An efficient prediction equation would therefore include the first four variables entered.

TABLE XXV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (Spanish I)

N = 62

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS L_t	46.61	8.23	.098	.57	.93	.57	28.87**
2	ITBS L_3	46.47	7.94	-.090	.64	.87	.31	9.19**
3	ITBS A_1	46.84	7.92	.039	.67	.86	.25	3.35
4	ITBS A_2	48.97	8.34	-.042	.70	.82	.10	5.37*
5	ITBS L_1	47.98	9.57	.039	.72	.81	.56	2.55
6	ITBS Vocab	47.66	8.66	.027	.73	.80	.49	2.31
		constant		-0.559				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*$p<.05$
**$p<.01$
df=1,(N-1)-K

In Table XXVI are reported the results when additional variables are included in the data pool. A noticeable increase in the multiple R's can be observed. At step one, in which the junior high English GPA entered the regression equation, 46% of the criterion variance is already accounted for. The early appearance of the English GPA predictor variable supports the finding by Hascall (1959) cited in Chapter II, REVIEW, page 26.

Variety of content in the variables entered in the regression equation is again observable. An arithmetic subtest appears again. The MLAT variable, Side A (total of Parts I and II, Number Learning and Phonetic Script, respectively), enters at the fifth step but its contribution is not statistically significant.

The disposition of Major Hypothesis I is that it is rejected at five of the six steps; it is accepted at the fifth step. Nearly as much of the criterion variance is explained by the first four variables as by all six--the increment in the proportion of the criterion variance explained is only 6%; therefore, since the standard error of estimate decreases very little with the addition of the last two variables, a user may wish to limit the prediction equation to the first four variables entered, especially if cost is a factor.

Spanish II: As can be noted in Table XXVII, in predicting GPA in the second year of Spanish, three variables are statistically significant in the contribution they make to the reduction of errors of estimation: the two arithmetic subtests and the Language total, all from the ITBS. The other three variables, Maps (W_1), Reading, and Vocabulary, represent, essentially, two content areas, reading and map reading skills. Thus, among all six variables entered, half of them (Language total, Reading, Vocabulary) are measuring skills that, on the

TABLE XXVI

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreigh Language (Spanish I)

N = 62

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	Jr Hi Eng GPA	3.17	.82	.627	.68	.83	.68	51.71**
2	ITBS L_1	47.98	9.57	.044	.76	.74	.56	15.81**
3	ITED Rdg NS	12.73	4.82	.063	.79	.71	.54	6.29*
4	ITBS A_2	48.97	8.34	-.021	.81	.68	.10	6.51*
5	MLAT Side A,To'l	9.11	4.69	.042	.82	.67	.42	3.43
6	Co-op Rdg Compr	48.77	7.90	-.026	.84	.65	.20	4.46*
		constant		-.320				

Source of Predictor Variables: Iowa Tests of Basic Skills
Iowa Tests of Educational Development
Lorge Thorndike Intelligence Test
Modern Language Aptitude Test
Co-operative Reading Test
Junior High Subject GPA

*p<.05
**p<.01
df=1,(N-1)-K

TABLE XXVII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (Spanish II)

N = 44

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	*r* with Criterion	F
1	ITBS A_1	55.89	9.21	.066	.65	.68	.65	31.16**
2	ITBS L_t	56.48	7.16	.033	.70	.65	.56	5.20*
3	ITBS A_2	56.61	9.30	-.026	.75	.61	.19	5.91*
4	ITBS Rdg	54.18	8.42	.015	.76	.60	.56	2.68
5	ITBS W_1	55.59	8.53	-.025	.77	.60	.44	1.10
6	ITBS Vocab	54.95	7.97	.020	.78	.60	.49	1.06
		constant		-.799				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*p<.05
**p<.01
df=1,(N-1)-K

surface at least, seem to have a relationship to the skills involved in a foreign language study.

The disposition of Major Hypothesis I is as follows: it is rejected at each of the first three steps and is accepted at each of the last three. Therefore, an optimally efficient prediction equation would include the first three variables entered. Supportive evidence for this statement can also be found by an inspection of the standard errors of estimate at steps four through six--no further decrease occurs.

Interestingly, when the variables are drawn from the six data sources listed at the bottom of Table XXVIII, two variables that contribute most to the reduction of the error sums of squares are the ITBS Arithmetic Concepts subtest and the junior high mathematics GPA. Together they account for 52% of the criterion variance. The next three variables entered, ITED Correctness of Expression, ITBS Arithmetic Problem Solving and Junior high English GPA, all represent skill areas that have frequently been among the predictor variables in the other foreign languages.

It is again of interest to observe for the Spanish I and II groups the differences in the mean values of the same predictor variables. Again, as was observed in the case of French I and French II (Tables XXII, XXIII, pp.113, 115), the larger mean values are associated with the second year course. The possible explanation offered for this difference for French I and II is probably relevant here.

The disposition of Major Hypothesis I is as follows: it is rejected at each of the six steps. No inference is made here concerning an optimally efficient prediction equation since the ratio of sample size to number of variables in the data source is such as to lead

TABLE XXVIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Foreign Language (Level II, Spanish)

N = 44

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS A_1	55.89	9.21	.044	.65	.68	.65	31.16**
2	Jr Hi Math GPA	3.68	1.05	.496	.72	.63	.62	7.40**
3	ITED Corr Expr	17.52	3.47	.066	.75	.60	.56	4.86*
4	ITBS A_2	56.61	9.30	-.030	.79	.57	.19	5.49*
5	Jr Hi Eng GPA	3.91	1.01	-.314	.81	.55	.34	4.40*
6	Co-op Rdg Speed	54.89	8.33	.022	.83	.53	.34	4.38*
		constant		.076				

Source of Predictor Variables: Iowa Tests of Basic Skills
Iowa Tests of Educational Development
Modern Language Aptitude Test
Co-operative Reading Test
Junior High Subject GPA
Lorge Thorndike Intelligence Test

*p<.05
**p<.01
df=1,(N-1)-K

to a zero value for the multiple R corrected for bias. The results are suggestive of what might be expected were a larger sample size available.

<u>Tests of Two Additional Regression Model Assumptions, Homoscedasticity and Linearity</u>

Earlier in this study (page 105) the normality of distribution assumption underlying the use of multiple correlation/regression model was tested for the ITBS and Lorge test score distributions for the total junior high group sample. A review of the results of that test (Table XIX) indicates that of the eighteen distributions, fourteen were found to be normally distributed, based on the chi squared goodness-of-fit test. It was pointed out in that section that a test of two other assumptions underlying the use of the multiple correlation/regression model would be limited to the English and mathematics subsamples. These assumptions are presented at this point before examining the regression equations within achievement levels for the English and mathematics subject areas.

1. The null hypothesis under test is: the array of the criterion scores associated with each of the six predictor variables within each level of two subject areas does not depart significantly from the mean of the array.

The rest of this assumption is accomplished by applying the Welch-Nayer L_1 test, a procedure described by Johnson (1949, pp. 240-246). In Tables XXIX and XXX are presented summaries of the results of the application of the Welch-Nayer L_1 test of homoscedasticity to each of the eighteen criterion arrays for the English and mathematics subject

TABLE XXIX

SUMMARY OF THE RESULTS OF THE APPLICATION OF THE WELCH-NAYER L_1 TEST OF THE ASSUMPTION OF HOMOSCEDASTICITY OF THE GPA CRITERION VARIABLES ABOUT THE REGRESSION LINE IN THE MULTIPLE REGRESSION ANALYSIS

Junior High Group Validation Sample, English

Distribution	Calculated L_1	f_s	K	Table Value L_1	Disposition of H_o
Basic Comp 1					
Lorge Total	1.35	1.8	10	.374	Accepted
W_2	.99	3.0	9	.527	Accepted
L_2	1.12	2.5	8	.520	Accepted
Reading	1.30	3.4	9	.527	Accepted
L_1	1.35	4.8	10	.696	Accepted
W_1	1.01	1.9	9	.367	Accepted
Composition 2					
W_3	1.089	18.8	9	.912	Accepted
A_1	.860	7.9	8	.795	Accepted
L_2	.982	2.8	11	.540	Accepted
L_4	.922	4.2	15	.652	Accepted
A_2	1.19	6.9	9	.772	Accepted
L_1	.924	3.9	10	.631	Accepted
Composition (Honors)					
L_t	1.37	4.2	10	.631	Accepted
L_3	1.10	2.7	10	.534	Accepted
L_1	1.25	2.6	9	.527	Accepted
L_4	1.54	2.7	9	.527	Accepted
W_1	1.44	2.4	9	.367	Accepted
A_1	1.74	3.7	10	.631	Accepted

TABLE XXX

SUMMARY OF THE RESULTS OF THE APPLICATION OF THE WELCH-NAYER L_1 TEST OF THE ASSUMPTION OF HOMOSCEDASTICITY OF THE GPA CRITERION VARIABLES ABOUT THE REGRESSION LINE IN THE MULTIPLE REGRESSION ANALYSIS

Junior High Group Validation Sample, Mathematics

Distribution	Calculated L_1	f_s	K	Table Value L_1	Disposition of H_o
General Math 1					
Reading	1.07	4.2	10	.631	Accepted
W_2	.99	3.7	11	.636	Accepted
Lorge Total	.82	3.9	9	.626	Accepted
A_1	.71	3.8	8	.620	Accepted
L_t	.80	3.0	8	.520	Accepted
W_1	.83	3.8	9	.626	Accepted
Algebra 2					
A_t	.93	9.6	7	.828	Accepted
L_2	.97	8.6	12	.818	Accepted
A_2	.94	5.1	10	.696	Accepted
Lorge N.V.	1.02	9.7	11	.840	Accepted
W_1	.96	3.6	9	.626	Accepted
W_2	.97	4.6	9	.691	Accepted
Algebra (Honors)					
W_t	1.39	4.2	9	.626	Accepted
W_2	1.05	2.7	8	.520	Accepted
L_t	1.60	4.1	11	.635	Accepted
Lorge N.V.	1.09	4.8	9	.691	Accepted
W_1	1.01	3.5	11	.635	Accepted

areas, respectively. Shown in these tables are the calculated L_1 values. The listed values for the harmonic mean, f_s, and the number of arrays, K, are used as degrees of freedom for entrance to Nayer's tables (Johnson, 1949, p. 366) to locate the critical L_1 values used for accepting or rejecting the null hypothesis. In Nayer's tables the null hypothesis is accepted when the calculated L_1 value is larger than the tabled value. It can be noted in Tables XXIX, XXX that the null hypothesis is accepted in each application. It is concluded that the assumption of homoscedasticity of the different arrays is fully met for these criterion distributions under study.

2. The null hypothesis under test is: the mean of the arrays of criterion scores does not depart significantly from the regression line.

The statistical procedure for testing this assumption is the analysis of variance for the linearity of regression. This procedure is outlined in detail by Johnson (1949, pp. 240-246). The test was applied to each of the six arrays of predictor and criterion variables that comprised the sixth step of the step-wise regression equation. A summary of the application of the analysis of variance test to the thirty-six distributions of criterion variables is given in Table XXXI. Tabulated in this table are the observed or calculated F values, the degrees of freedom for entry into the F table, the probability level (α.05) for rejection and the disposition of the hypothesis.

The null hypothesis is sustained for all thirty-six criterion variable distributions but one, algebra, Level 2, where the mean of the criterion variable, GPA, departs significantly from the predictor variable, ITBS Arithmetical Concepts (A_1), regression line. These findings indicate that the linearity of regression assumption is fulfilled

TABLE XXXI

SUMMARY OF THE APPLICATION OF THE ANALYSIS OF VARIANCE FOR LINEARITY OF REGRESSION OF THE ITBS AND LORGE VARIABLES COMPRISING THE MULTIPLE REGRESSION EQUATIONS FOR JUNIOR HIGH GROUP VALIDATION SAMPLE, ENGLISH AND MATHEMATICS SUBJECT AREAS

Distribution	Observed F	df N_1	df N_2	Tabled F	Disposition of Hypothesis
Basic Comp. 1					
Lorge Total	1.49	8	15	2.64	Accepted
W_2	.34	7	16	2.66	Accepted
L_2	1.83	6	16	2.74	Accepted
Reading	.88	7	16	2.66	Accepted
L_1	.80	8	15	2.64	Accepted
W_1	.72	7	16	2.66	Accepted
Composition 2					
W_3	1.96	7	84	2.12	Accepted
A_1	1.23	6	84	2.21	Accepted
L_2	1.29	9	82	1.99	Accepted
L_4	1.71	12	79	1.88	Accepted
A_2	1.20	7	84	2.12	Accepted
L_1	1.25	8	82	2.05	Accepted
Comp. (Honors) 3					
L_t	2.33	8	25	2.34	Accepted
L_3	.65	8	25	2.34	Accepted
L_1	.53	7	26	2.39	Accepted
L_4	.30	7	26	2.39	Accepted
W_1	.74	7	26	2.39	Accepted
A_1	.04	8	25	2.34	Accepted
General Math 1					
Reading	.67	8	43	2.60	Accepted
W_2	1.05	9	42	2.11	Accepted
Lorge Total	.29	7	44	2.23	Accepted
A_1	1.12	6	45	2.30	Accepted
L_t	1.23	6	45	2.30	Accepted
W_1	.91	7	44	2.23	Accepted
Algebra 2					
A_t	5.31**	5	164	2.27	Rejected
L_2	.46	10	159	1.89	Accepted
A_2	.69	8	161	2.00	Accepted
Lorge N.V.	1.46	9	160	1.94	Accepted
W_1	1.50	7	162	2.07	Accepted
W_2	1.36	7	162	2.06	Accepted
Algebra (Honors) 3					
W_t	1.27	7	30	2.34	Accepted
W_2	.41	6	31	2.42	Accepted
L_t	1.77	9	28	2.24	Accepted
A_1	1.16	6	31	2.42	Accepted
Lorge N.V.	.08	7	30	2.34	Accepted
W_1	.59	9	28	2.22	Accepted

**$p < .01$

in thirty-five out of the thirty-six applications of the analysis of variance test, a finding which demonstrates that this assumption has been substantially met.

Basic Composition: In Table XXXII it can be observed that the best predictor, and the only one of the six entered into the regression equation for Level I English that contributes at a statistically significant level to a reduction in the errors of estimation is the Total score from the Lorge Thorndike Intelligence test. Thirty-seven percent of the criterion variance is explained by this variable.

The disposition of Major Hypothesis I is as follows: it is rejected at step one and is accepted at steps two through six. No inference is made concerning an optimally efficient regression equation because the equation is based on such a limited sample size. However, stability of the equation as applied to a new sample is reported later (see p. 149). The finding reported there offers support for the suggestion that with a larger sample it is quite likely that the results reported here will be replicated.

Composition: Three subtests from the ITBS battery, Knowledge and Use of Reference Materials (W_3), Capitalization (L_2) and Usage (L_4) are listed in Table XXXIII in the order of their inclusion in the step-wise regression equation. Together, at a statistically significant level, they explain 44% of the criterion variance. The appearnce at step one of the W_3 subtest is analagous to the counterpart of this variable in the ITED battery, Use of Sources of Information, which was also entered at step one in the estimate of the combined sample Level 2 composition GPA (Table XI). It can be noted that after step three, although the multiple R increases slightly, no further decrease in the standard

TABLE XXXII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

English (Level 1, Basic Composition)

N = 25

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	Lorge Total	111.00	11.38	.045	.61	.56	.61	13.99**
2	ITBS W_2	43.92	7.49	-.045	.67	.54	.23	2.55
3	ITBS L_2	43.00	6.44	.035	.70	.52	.51	2.21
4	ITBS L_1	45.80	8.40	.026	.73	.51	.39	1.87
5	ITBS Rdg	42.08	7.66	-.026	.75	.51	.34	1.08
6	ITBS W_1	42.92	7.49	.009	.76	.52	.21	0.29
		constant		-2.303				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*$p<.05$
**$p<.01$
df=1,(N-1)-K

TABLE XXXIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

English (Level 2, Composition)

N = 93

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS W_3	51.71	7.83	.038	.60	.70	.59	53.38**
2	ITBS L_2	51.92	8.63	.021	.64	.68	.46	8.40**
3	ITBS L_4	50.95	8.82	.013	.66	.66	.35	4.62*
4	ITBS A_1	50.90	7.91	.029	.68	.66	.50	3.26
5	ITBS A_2	51.71	8.12	-.015	.68	.66	.34	1.39
6	ITBS L_1	51.37	8.39	.011	.69	.66	.36	1.24
		constant		-1.741				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*$p < .05$
**$p < .01$
df=1,(N-1)-K

error of estimate occurs.

The disposition of the Major Hypothesis I is as follows: it is rejected at each of the first three steps and is accepted at each of the last three. Therefore, it is concluded that an optimally efficient prediction equation would include the first three variables entered.

Composition (Honors): Evidence for the predictive validity of the ITBS Language Total variable for this level of English can be observed (with caution because of sample size) by the data presented in Table XXXIV. This variable alone, at a statistically significant level, accounts for 32% of the criterion variance. A subtest within the ITBS Language area, Punctuation (L_3), which was the second variable included in the regression equation, is the only other variable among the six entered that significantly contributes to a reduction in the errors of estimation.

The disposition of Major Hypothesis I is as follows: it is rejected at steps one and two and is accepted at steps three through six. It is therefore suggested that an optimum prediction equation would include the first two variables entered. If the sample size were larger and the results were the same as reported here, a conclusion could be reached relative to an optimum prediction equation for estimating the criterion variable.

When another set of predictor variables, the first semester eighth grade GPA's within four subject areas (English, mathematics, science, social studies), were added to the data source along with the ITBS and Lorge data for estimating the instructional level criterion variables, none of the junior high GPA's were included in the regression equation by the step-wise procedure. Therefore, it is concluded that for the

TABLE XXXIV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

English (Level 3, Composition Honors)

N = 35

Step	Predictor Variable	$\bar{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS L_t	62.26	7.88	.119	.57	.40	.57	15.87**
2	ITBS L_3	60.69	8.40	-.049	.63	.38	.32	4.18*
3	ITBS W_1	60.66	8.68	-.022	.67	.37	-.03	2.61
4	ITBS L_4	61.49	7.47	-.022	.68	.37	.34	1.05
5	ITBS L_1	59.89	7.83	-.028	.71	.37	.51	1.77
6	ITBS A_1	63.86	8.43	.011	.72	.37	.13	1.04
		constant		3.308				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*p<.05
**p<.01
df=1,(N-1)-K

instructional levels within English, previous grades made no contribution to a reduction in the errors of estimating the criterion variables.

Mathematics

General Mathematics: In Table XXXV are displayed a tabulation of the six variables entered in the regression equation by the step-wise procedure. It is clearly evident that the estimation of the criterion variable through use of predictor variables derived from the ITBS and Lorge data sources is of limited success. The first variable entered, Reading Graphs and Tables (W_2), the only variable of the six entered which contributes significantly to a reduction of the error sums of squares (only 8% of the variance is explained, however), bears a relationship, by test title, to presumed aspects of the content of this course. The very limited success reported here in estimating the criterion variable with the ITBS and Lorge as data sources is similar to the limited success reported earlier (Table XIII, p.93) for the combined general mathematics instructional level in which the ITED battery served as the predictor variable data source.

No conclusions are tenable concerning an optimum set of predictor variables for estimating the criterion variable. However, despite the inadequate explanation of the criterion variance provided by the regression equation listed in Table XXXV the six-variable regression equation was applied to a check sample, the results of which are presented in Table XLIII, page 151.

Junior high GPA's in the four subject areas were added to the ITBS and Lorge data sources but none of the GPA's were entered at any of the six steps of the step-wise regression routine. Therefore, the conclusion is apparent--the junior high grades were even less effective

TABLE XXXV

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Mathematics (Level 1, General Math)

N = 53

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	R
1	ITBS W_2	43.08	7.61	.045	.29	.89	.29	4.82*
2	Lorge Total	109.31	8.19	.027	.33	.88	.27	1.36
3	ITBS Rdg	42.26	7.32	-.048	.40	.87	.06	2.76
4	ITBS L_t	42.77	7.03	.023	.43	.87	.19	1.42
5	ITBS W_1	43.53	8.13	-.020	.44	.87	.002	.67
6	ITBS A_1	40.87	6.30	.025	.46	.87	.18	1.16
		constant		-.025				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*$p < .05$
df=1,(N-1)-K

than the six variables listed in Table XXXV for contributing to an explanation of the criterion variance.

Algebra: More encouraging results of the estimation of criterion variance in junior high group first semester algebra, using variables derived from the ITBS and Lorge data sources, can be observed in Table XXXVI. The first four variables entered, Capitalization (L_2), Arithmetic Concepts (A_1), Problem Solving (A_2) and the Lorge Non-Verbal subtest, all significantly contribute to a reduction in the errors of estimation and together explain 35% of the criterion variance.

The disposition of Major Hypothesis I is as follows: it is rejected at steps one through four and is accepted at steps five and six. Therefore, it is concluded that an optimally efficient prediction equation would include the first four variables entered.

With the addition of junior high GPA's to the predictor variable data source, it can be observed in Table XXXVII that the best single predictor of Level 2 mathematics (algebra) is the junior high first semester mathematics GPA which alone accounts for 31% of the criterion variance; the five additional variables together account for 15% more (all six account for 46%). It is interesting to note the new order in which the variables are entered in Table XXXVII as well as to observe which variables now contribute at a statistically significant level to a reduction in the error sums of squares. With the junior high GPA's in the data source, only the ITBS L_2 subtest, of the four significant predictor variables in Table XXXVI, enters at a significant level in the second equation (it entered first in the previous regression equation). Thus, with ITBS W_1, there are now only three variables that make a significant contribution to an explanation of the criterion variance,

TABLE XXXVI

SUMMARY STATISTICS AT STEP SIX OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Mathematics (Level 2, Algebra)

N = 171

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS L_2	51.21	8.36	.032	.46	.80	.46	44.37**
2	ITBS A_1	51.33	6.69	.038	.55	.75	.44	23.58**
3	ITBS A_2	51.84	8.18	-.019	.57	.74	.09	4.51*
4	Lorge N.V.	123.68	10.54	.011	.59	.73	.36	5.26*
5	ITBS W_1	51.18	7.84	.011	.60	.73	.32	2.77
6	ITBS W_2	51.27	8.11	.009	.60	.73	.38	1.05
		constant		-1.717				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

*$p < .05$
**$p < .01$
df=1,(N-1)-K

TABLE XXXVII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

Mathematics (Level 2, Algebra)

N = 171

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	r with Criterion	F
1	Jr Hi Math GPA	3.30	.90	.373	.56	.74	.56	79.16**
2	ITBS L_2	51.70	8.73	.025	.63	.70	.46	23.70**
3	ITBS W_1	51.18	7.84	.015	.66	.68	.32	8.27**
4	ITBS A_1	51.33	6.69	.022	.67	.68	.44	3.89
5	ITBS A_2	51.84	8.18	-.013	.67	.67	.09	2.73
6	Lorge N.V.	123.68	10.54	.008	.68	.67	.36	2.17
		constant		-1.490				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test
Junior High Subject GPA

**p<.01
df=1,(N-1)-K

but these three make a greater contribution (44%) than the four variables of the first equation (35%).[1]

The disposition of Major Hypothesis I is that it is rejected at each of the first three steps and is accepted over the last three. Therefore, for this population, an optimally efficient prediction equation that included ITBS, Lorge and first semester eighth grade GPA's junior high in the data source would consist of the first three variables entered.

Algebra (Honors): The multiple R's, beta weights and related statistics for estimating the criterion variable in this Level 3 ninth grade mathematics course are presented in Table XXXVIII. Of the six variables entered in theequation, only the first two entered, Language Total and Arithmetic Total, contribute significantly to a reduction in the error sums of squares. These two explain 36% of the criterion variance. Further evidence for the predictive validity of the ITBS Arithmetic subtest is demonstrated by its statistically significant Accounting for a portion of the criterion variance. Although not statistically significant, the other four variables included by the step-wise procedure together explain 50% of the criterion variance. It is to be noted that, if the F values associated with the last four variables had been significant, the 15% increase in the amount of criterion variance explained could be regarded as a reliable increase.

The disposition of Major Hypothesis I is that it is rejected at steps one and two and is accepted at steps three through six. The tentative conclusion reached is that an optimally efficient prediction equation would include the first two variables entered.

As can be observed in Table XXXIX, when junior high GPA's are added

---

[1]See page 146 for a presentation of a test of the significance of the differences.

TABLE XXXVIII

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTING GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVEL

Junior High Group Validation Sample

Mathematics (Level 3, Algebra Honors)

N = 39

Step	Predictor Variable	$\bar{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	ITBS L_t	59.74	8.83	.033	.50	.48	.50	12.64**
2	ITBS A_t	64.64	6.61	.024	.60	.45	.42	5.66*
3	Lorge N.V.	136.18	8.44	.013	.63	.44	.47	2.19
4	ITBS W_1	63.13	8.65	-.003	.66	.44	.22	2.54
5	ITBS W_2	63.92	7.93	.042	.69	.42	.42	2.77
6	ITBS W_t	64.26	7.19	-.047	.71	.42	.33	1.92
		constant		-.040				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

---

*p<.05
**p<.01
df=1,(N-1)-K

TABLE XXXIX

SUMMARY STATISTICS OF THE STEP-WISE MULTIPLE REGRESSION ROUTINE PREDICTINF GPA WITHIN SUBJECT AREA ACHIEVEMENT LEVELS

Junior High Group Validation Sample

(Level 3, Algebra Honors)

N = 39

Step	Predictor Variable	$\overline{X}$	S.D.	Beta Weights	Multiple R	S.E. of Estimate	$\underline{r}$ with Criterion	F
1	Jr Hi Eng GPA	4.44	.68	.401	.70	.40	.70	36.21**
2	Jr Hi Math GPA	4.69	.52	.431	.81	.33	.65	18.06**
3	ITBS A_1	65.62	6.07	.020	.84	.31	.43	4.58*
4	ITBS W_t	64.26	7.19	-.025	.85	.31	.33	2.08
5	ITBS V	62.15	9.60	.008	.86	.31	.18	1.43
6	ITBS R	61.36	7.49	.014	.86	.30	.28	1.41
		constant		-.919				

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test
Junior High Subject GPA

*p<.05
**p<.01
df=1,(N-1)-K

to the data source, new variables are entered in the regression equation. Interestingly, the single best predictor of the criterion variable is the junior high English GPA which alone accounts for 49% of the criterion variance. Entered second by the step-wise procedure is the junior high mathematics GPA which explains 17% of the crition variance. The third variable to enter at a significant level is the ITBS A_1 subtest (Arithmetic Concepts).

The disposition of Major Hypothesis I is that it is rejected at steps one through three while it is accepted at steps four through six. Conclusions concerning an optimally efficient prediction equation are very tentative in view of the ratio of the number of variables in the data source to the sample size. It would appear that an efficient prediction equation for this population utilizing the listed data sources would be comprised of the first three variables entered.

Summary of Major Hypothesis I, Junior High Sample

A review of the results presented over four subject areas of the junior high group sub-sample is provided in Table XL. The variables selected by the step-wise regression procedure for each instructional level which signficantly account for a portion of the criterion variance are listed in this table. The following trends appear to emerge from the tabulated results:

1. When the social studies and foreign languages criterion variables are being estimated by predictor variables drawn from the ITBS and Lorge data sources, the significant variables tend to be those entered at one of the first four steps.

2. When other variables are added to the data pool for predicting foreign language GPA, all six variables selected by the step-wise

TABLE XL

SUMMARY OF STATISTICALLY SIGNIFICANT VARIABLES INCLUDED IN THE STEP-WISE REGRESSION EQUATION FOR EACH INSTRUCTIONAL LEVEL OF FOUR SUBJECT AREAS

Junior High Group Validation Sample

Subject Area & Level	Data Source	Significant Predictor Variables at Steps: 1	2	3	4	5	6
Social Stu							
World Geog	ITBS Lorge	ITBS Vocab	ITBS W_3				
West Civ	ITBS Lorge	ITBS Compos	ITBS L_3	ITBS A_2			ITBS L_2
For Lang							
French I	ITBS Lorge	ITBS Vocab					
French II	ITBS Lorge	ITBS L_t					
French II	ITBS Lorge ITED MLAT Co-opRdg JrHi GPA	Fr I GPA	MLAT Total	Eng GPA			
Spanish I	ITBS Lorge	ITBS L_t	ITBS L_3		ITBS A_2		
Spanish I	ITBS Lorge ITED MLAT Co-opRdg JrHi GPA	Eng GPA	ITBS L_1	ITED Rdg NS	ITBS A_2		Co-opRdg Compre
Spanish II	ITBS Lorge	ITBS A_1	ITBS L_t	ITBS A_2			

TABLE XL (Continued)

Subject Area & Level	Data Source	Significant Predictor Variables at Steps: 1	2	3	4	5	6
Spanish II	ITBS Lorge ITED MLAT Co-opRdg JrHi GPA	ITBS A_1	Math GPA	ITED CorrEx	ITBS A_2	Eng GPA	Co-opRdg Speed
English							
Basic Comp	ITBS Lorge*	Lorge Total					
Comp	ITBS Lorge*	ITBS W_3	ITBS L_2	ITBS L_4			
Comp(Hon)	ITBS Lorge*	ITBS L_t	ITBS L_3				
Mathematics							
Gen Math	ITBS Lorge*	ITBS W_2					
Algebra	ITBS Lorge	ITBS L_2	ITBS A_1	ITBS A_2	Lorge N.V.		
Algebra	ITBS Lorge JrHi GPA	Math GPA	ITBS L_2	ITBS W_1			
Alg(Hon)	ITBS Lorge	ITBS L_t	ITBS A_t				
Alg(Hon)	ITBS Lorge JrHi GPA	Eng GPA	Math GPA	ITBS A_1			

*When junior high subject GPA's were added to the data source, none were chosen by the step-wise routine.

procedure were significant in two out of three foreign language areas.

3. In the estimation of the criterion variable in the English and mathematics subject areas (Table XL continued, page 144), it is clearly evident that by the third or fourth steps, all of the significant variables have been entered.

4. An examination of the ITBS and Lorge Variables entered over all four subject areas indicates that, of all the significant variables entered, the ITBS subtest A_2 (Arithmetic Problem Solving) was the most frequently selected variable.

5. The ITBS subtest, Reading, is conspicuously absent from the listing of significant variables.

6. Variables from the Lorge-Thorndike battery enter significantly only twice (Total and Non-Verbal, once each).

7. The junior high English GPA entered as a significant variable in each foreign language equation when it was include in the data pool.

Minor Hypothesis I

A report has already been given of the contribution to a reduction in the errors of estimation in the criterion variable within instructional levels of selected subject areas made by junior high GPA's (pages 116 ff.). The question arises whether the reported increments in the amount of variance accounted for by the addition of the junior high grade point averages to the ITBS and Lorge data pool is significant. Minor Hypothesis I is concerned with this question. This hypothesis states that there are no significant differences within the junior high validation sample in the proportion of variance explained (R^2) in the dependent variable (GPA) within levels of selected subject areas as a function of the addition of non-test variables (junior high GPA's within subject areas) to the ITBS and Lorge data pool.

To evaluate this hypothesis, Hotelling's correlated samples F test (Wert et al., 1954, p. 299) for the significance of differences between the multiple R's was employed. A summary of the results of applying this test to the pairs of multiple R's is given in Table XLI. Disposition of Minor Hypothesis I is that it is rejected at Levels 2 and 3 mathematics (algebra and algebra honors, respectively).

It is therefore concluded that when junior high subject area GPA's were added to the ITBS and Lorge data pool and were selected by the step-wise regression routine, as in the case of Levels 2 and 3 mathematics, they accounted for a significantly greater amount of the criterion variance. This conclusion is limited by the fact that the junior high subject areas GPA variables were not selected by the step-wise regression routine in Level 1 mathematics nor in the three levels of English. However, it will be recalled (Tables XXIV, XXVI, XXVIII, pages 117, 121, 124) that, in Spanish I and II and French II, when junior high GPA's were added to the data source along with other variables (from MLAT, Co-op Reading, ITBS and Lorge), certain junior high GPA's (among other variables) were selected by the step-wise routine, forming equations which yielded larger multiple R's than those resulting when ITBS and Lorge batteries provided the only data sources. It has been stressed (see page 101) that, because of the large number of variables in the data pool (MLAT, Co-op Reading, junior high grades, ITBS, Lorge) in relation to the instructional level sample size, the increments in the multiple R's are biased upward. For this reason, significance tests were not computed, as well as for the reason that the equations were not

TABLE XLI

SUMMARY OF THE HOTELLING F TEST OF THE SIGNIFICANCE OF DIFFERENCES BETWEEN JUNIOR HIGH GROUP VALIDATION SAMPLE MULTIPLE R'S RESULTING FROM DIFFERENT SETS OF PREDICTOR VARIABLES

Subject Area and Level	Data Sources		df	F	$F_{.05}$	p	Disposition of H_o
	ITBS Lorge	ITBS Lorge JrHi Subject GPA					
	Multiple R						
Mathematics							
Algebra	.56	.68	1,169	7.04	6.81	<.05	Rejected
Algebra (Honors)	.71	.86	1,37	5.52	4.11	<.05	Rejected

evaluated on check samples because the MLAT and Co-op Reading batteries were dropped in the check sample year.

<u>Major Hypothesis II, Junior High Group Sub-Sample</u>

Before the results of the test of Major Hypothesis II are presented, a report is given in Tables XLII, XLIII of the validation and check sample means and standard deviations of the criterion variables, the multiple R's and the standard errors of estimation of the predicted GPA. Included also in these tables are the validation sample multiple R's corrected for bias (R_c; see page 101 for a discussion of the rationale behind this statistic). The R_c values as recorded in Tables XLII, XLIII, would, by themselves, suggest that there was little or no value in some of the validation sample multiple R's because of the zero to near zero R_c values as reported. However, inspection of the check sample multiple R's for all levels of all subject areas reveals that all of the prediction equations demonstrated stability when applied to new samples. A further check on the efficiency of these equations as applied to the check sample data can be noted in the standard errors of estimation (S.E.). When validation and check sample S.E.'s are subtracted from their respective criterion variable standard deviations, a comparison between the validation and check samples of the resulting differences reveals that the relative errors of estimation between samples are quite similar. The largest differences occur in Spanish I and II and general mathematics. It can be noted that larger standard deviations are associated with the Spanish I and II check sample criterion variables. In fact, a further examination of the Spanish I S.E. (.80) in comparison with the criterion variable S.D. (.89) suggests that, despite the sizeable multiple R (.73),

TABLE XLII

MEANS AND STANDARD DEVIATIONS OF CRITERION VARIABLES, MULTIPLE CORRELATION COEFFICIENTS AND STANDARD ERROR OF THE PREDICTED FIRST SEMESTER NINTH GRADE SOCIAL STUDIES AND FOREIGN LANGUAGE GPA

Junior High Validation and Check Samples

Subject Area & Level	Sample Size		Criterion Variable (First Semester GPA)				Multiple Correlation Coefficient			Standard Error of Estimate of Predicted GPA	
			Mean		Std Dev						
			Sample		Sample		Sample			Sample	
	Vali-dation	Check	Val.	Check	Val.	Check	Val.		Check	Val.	Check
							R	Rc	R		
SOCIAL STUDIES											
World Geog	124	58	3.0	2.6	.90	.72	.65	.58	.69	.72	.62
West Civ	136		3.3		1.03		.73	.67		.72	
FOREIGN LANG											
French I	27	31	2.5	3.0	.94	1.2	.81	.15	.65	.63	.95
French II	52	36	3.6	3.4	.91	1.1	.58	.10	.40	.79	.98
Spanish I	62	74	3.7	3.0	.89	1.1	.73	.59	.69	.80	.77
Spanish II	44	40	3.8	4.0	.89	1.3	.78	.60	.75	.60	.83

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

TABLE XLIII

MEANS AND STANDARD DEVIATIONS OF CRITERION VARIABLES AND THE MULTIPLE CORRELATION COEFFICIENTS AND STANDARD ERROR OF THE PREDICTED FIRST SEMESTER NINTH GRADE ENGLISH AND MATHEMATICS GPA

Junior High Validation and Check Samples

Subject Area & Level	Sample Size		Criterion Variable (First Semester GPA)				Multiple Correlation Coefficient			Standard Error of Estimate of Predicted GPA	
			Mean		Std Dev						
			Sample		Sample		Sample			Sample	
	Validation	Check	Val.	Check	Val.	Check	Val.		Check	Val.	Check
							R	Rc	R		
ENGLISH											
Basic Comp	25	25	2.7	2.7	.69	.82	.76	.00	.66	.52	.62
Comp	93	83	3.2	1.1	.88	.92	.69	.61	.74	.66	.67
Comp (Honors)	35	40	4.1	3.9	.48	.52	.72	.20	.72	.37	.41
MATHEMATICS											
Gen Math	53	40	3.0	3.0	.95	.80	.46	.00	.78	.87	.50
Algebra	171	193	3.3	3.2	.90	.97	.56	.54	.55	.73	.68
Algebra (Hon)	39	26	4.6	4.7	.55	.56	.71	.30	.72	.42	.39

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Test

the estimate of the Spanish I criterion variable was minimally successful. It will be recalled (page 135) that no conclusions were reached concerning an efficient prediction equation for general mathematics. However, it can be noted that the regression equation held up surprisingly well. Nevertheless, the validation sample multiple R cannot be ignored and before any conclusions are made concerning an effective prediction equation for this instructional level, application of the prediction equation to additional samples would be necessary.

Consideration is now given to an evaluation of Major Hypothesis II, which states that there are no significant differences within instructional levels in the multiple R's between the validation and check samples, foreign lanuages, English and mathematics. As noted on page 103 for the combined sample, the appropriate statistical procedure for testing this hypothesis is the $\underline{z}$ test for the significance of differences between multiple R's drawn from independent samples (Wert et al., 1954, page 296). The $\underline{z}$ test was computed where the differences between the validation and check sample multiple R's was large enough to warrant the test (Table XLIV). Only one significant value was found and that was between differences of the general mathematics validation and check sample multiple R's ($\underline{z} = 2.53$, $p < .05$). Thus, the disposition of Major Hypothesis II is that it was accepted for all comparisons made between the validation and check sample multiple R's except one, general mathematics, where it was rejected at the .05 probability level. It is therefore concluded that prediction equations developed within junior high group instructional levels of four subject areas demonstrated predictive stability when applied to new samples.

TABLE XLIV

SUMMARY OF THE $\underline{z}$ TEST OF THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN VALIDATION AND CHECK SAMPLE MULTIPLE CORRELATION COEFFICIENTS IN SELECTED SUBJECT AREAS OF THE JUNIOR HIGH GROUP

Subject Area and Level	Multiple Correlation Coefficient		$\underline{z}$	p	Disposition of Hypothesis
	Validation Sample	Check Sample			
Foreign Languages					
French I	.81	.65	.38	$> .05$	Accepted
English					
Basic Composition	.76	.66	.67	$> .05$	Accepted
Composition	.69	.74	.67	$> .05$	Accepted
Mathematics					
General Math	.46	.78	2.53	$< .05$	Rejected

$\underline{z}_{.05} = 1.96$; df = ∞

Source of Predictor Variables: Iowa Tests of Basic Skills
Lorge Thorndike Intelligence Tests

TABLE XLV

COMPARISON WITH THE A PRIORI PROPORTIONS, THE PROPORTIONS OF STUDENTS PREDICTED/CLASSIFIED WITHIN INSTRUCTIONAL LEVELS OF TWO SUBJECT AREAS BY MEANS OF MULTIPLE REGRESSION EQUATIONS

Junior High Group Check Sample

	English			
	Number of students assigned by:			
Level	Regression Eq.		a priori	
	N	Prop.	N	Prop.
Basic Composition Level 1	27	0.18	25	0.17
Composition Level 2	93	0.63	83	0.56
Composition (Honors) Level 3	28	0.19	40	0.27
Total	148		148	

Mathematics

General Math Level 1	44	0.17	40	0.15
Algebra Level 2	190	0.73	193	0.75
Algebra (Honors) Level 3	25	0.10	26	0.10
Total	259		259	

TABLE XLVI

SUMMARY OF THE RESULTS OF THE APPLICATION OF THE SINGLE CLASSIFICATION CHI SQUARED TEST OF THE SIGNIFICANCE OF DIFFERENCES BETWEEN THE PROPORTION OF STUDENTS PREDICTED/CLASSIFIED OVER ALL IN TWO SUBJECT AREAS BY MEANS OF MULTIPLE REGRESSION EQUATIONS AND THE PROPORTION EXPECTED A PRIORI

Junior High Group Check Sample

Subject Areas	Observed Chi Squared	p	Disposition of Hypothesis
English	4.96	$> .05$	Accepted
Mathematics	.48	$> .05$	Accepted

Chi squared, .05, 2df = 5.99

PHASE TWO OF THE STUDY PROBLEM

The attention of the study now focuses on phase two of the study's problem: to assess, against various criteria (a priori proportions, hits/misses and chance expectations), the effectiveness of the multiple correlation/regression equations in predicting/classifying students within instructional levels of English and mathematics. This phase of the problem will be defined by Major and Minor Hypotheses III.

<u>Major Hypothesis III</u>

Major Hypothesis III states that there are no significant differences in the proportion of students predicted/classified within levels of English and mathematics, by means of multiple regression equations, and the proportion expected a priori. The Level 2 regression equations, developed within the junior high group validation sample subject areas, were then applied (see Chapter III, pages 57-58 for procedure) to a junior high group check sample which yielded proportions that were compared with the actual numbers (a priori) of students comprising the groups at the end of the first semester, ninth grade.

A summary is presented in Table XLV of the results of applying the Level 2 regression equations to the appropriate check sample predictor test scores of students in the English and mathematics subject areas. It is evident from these results that the proportions of students predicted/classified within levels of English and mathematics closely approximate the a priori proportions. To evaluate statistically the degree to which these proportions do approximate the a priori proportions, the single classification chi squared test was applied. A summary is presented in Table XLVI of the results. The results indicate that Major Hypothesis III, which states that no significant differences exist between proportions, is accepted in each instance ($p>.05$). This finding demonstrates the effectiveness of the application of the multiple regression equations in predicting/classifying students proportional to a priori expectations.

Minor Hypothesis III

The hypothesis under test is that:

The over-all proportion of students in the junior high group check sample correctly predicted/classified in each of two subject areas, English and mathematics, by means of multiple regression equations does not differ significantly from the proportion expected based upon the operation of chance. The rationale for using chance as a criterion with which to compare the hits/misses tallies was presented in Chapter I, page 9 .

An illustration of the procedure for determining the total number of students expected to be placed correctly in each subject area is presented on page 198 of the Appendix. The proportions of correct predictions by means of multiple regression equations compared with the proportion expected on the basis of chance are summarized and presented in Table XLVII.

The phrase, correctly classify, has been defined in Chapter I, page 15. However, it would be well to repeat the definition here: it means that the student has been classified by the statistical procedure (i.e., by each method) in the a priori group in which he is registered and is also succeeding in the tasks set for that level, i.e., his first semester GPA is 1.5 or higher (F = 1; A = 5). Thus, to take an illustration, if a student's a priori group membership was in Level 2

English but his first semester GPA was a grade of F and the regression equation classified him in Level 1 while the adjusted discriminant equation (see pp. 163ff. for presentation of the results of prediction/ classification by discriminant equations) classified him in Level 2, a 'hit' tally was recorded for the regression and a 'miss' for the adjusted equation.

As another example, if a student's a priori membership was Level 2 of mathematics where his GPA was an A (5.0) and the regression equation predicted/classified him as being most like the students whose a priori membership was Level 3, this classification would be tallied as a 'miss'.

It is evident in the case of English, when the proportion of total 'hits' expected by chance is 51% (75/148) the number of 'hits' achieved by the regression procedure is 71% (110/148). For mathematics, the over-all proportion of 'hits' expected by chance is 59% (152.6/259) while the number of 'hits' actually achieved by the regression procedure is 76% (197/259). It is clearly evident that, for both subject areas, chance expectation of 'hits' are exceeded by the number of 'hits' achieved by the regression procedure. An evaluation of whether the proportion of obtained differences of actual 'hits' over chance expectations are significant is discussed next. In Table XLVIII is presented a summary of the application of the single classification chi squared test to these data. In each instance, the null hypothesis is rejected ($p<.001$) indicating that the over-all proportions in the English and mathematics subject areas which were correctly predicted by means of multiple regression equations exceed the numbers correctly placed by a random procedure well beyond chance expectations.

TABLE XLVII

SUMMARY OF THE COMPARISON WITH TOTAL CHANCE EXPECTATIONS OF THE TOTAL NUMBER OF STUDENTS PREDICTED/CLASSIFIED (HITS) IN TWO SUBJECT AREAS BY THE MULTIPLE REGRESSION EQUATION

Junior High Group Check Sample

Levels of Achievement	Actual Number (a priori)	Number of Students Correctly Predicted by Multiple Regression Equation		Total Expected by Chance	
		Hits	Misses	Hits	Misses
English					
Basic Comp Level 1	25	19	6		
Composition Level 2	83	70	13		
Comp. (Honors) Level 3	40	21	19		
Total	148	110	38	75.097	72.903
Mathematics					
General Math Level 1	40	22	18		
Algebra Level 2	193	158	35		
Algebra (Honors) Level 3	26	17	9		
Total	259	197	62	152.606	106.374

TABLE XLVIII

SUMMARY OF THE RESULTS OF THE SINGLE CLASSIFICATION CHI SQUARED TEST FOR DEPARTURE OF CORRECT PREDICTION/CLASSIFICATIONS BY THE MULTIPLE REGRESSION EQUATIONS FROM THOSE EXPECTED BY CHANCE

Junior High Group Check Sample

Subject	Observed Chi Squared	p	Disposition of Hypothesis
English	32.93	< .001	Rejected
Mathematics	30.40	< .001	Rejected

Chi squared, .001, 1df = 10.83.

PHASE THREE OF THE PROBLEM

At this point, the investigation of the study problem will focus on phase three: to assess, against various criteria (a priori proportions, hits/misses, and chance expectations) the effectiveness of two versions of the multiple discriminant model in predicting/classifying students within instructional levels of English and mathematics. This phase of the problem will be defined by Major and Minor Hypotheses IV.

Before the results of the application of the multiple discriminant analysis procedure are given, the assumptions underlying the use of the discriminant model will be first tested. These assumptions are:

1. That the variables which are entered as input to the discriminant procedure are normally distributed.

2. That homogeneity of within-group variance prevails.

Normality of Distribution Assumption

The null hypothesis under test is that the observed distribution of test score frequencies do not depart signficantly from the theoretical normal distribtuion.

To test this assumption, the chi squared test of goodness-of-fit was applied to the test score distributions of the variables used in the multiple discriminant equations. The results of the chi squared test as applied are reported in Table XLIX. It will be noted that, of the ten test score distributions in the total junior high group subsample (L_t and A_1) were used in both the English and mathematics discriminant equations), eight yield chi squared values which lead to an acceptance of the null hypothesis. Only two, the Lorge Non-verbal and ITBS L_4, used in the mathematics and English subject areas respectively,

TABLE XLIX

SUMMARY OF THE RESULTS OF THE GOODNESS-OF-FIT CHI SQUARED TEST OF THE NORMALITY ASSUMPTION AS APPLIED TO THE DISTRIBUTIONS OF ITBS AND LORGE–THORNDIKE SCORES USED IN THE MULTIPLE DISCRIMINANT ANALYSIS

JUNIOR HIGH GROUP VALIDATION SAMPLE

Variables	Observed Chi Squared	Tabled Chi Squared (.05)	df	p		Disposition of Hypothesis
ITBS						
L_1 Spelling	4.90	14.07	7	> .50	< .70	Accepted
L_4 Usage	14.48	14.07	7	> .02	< .05	Rejected
L_t Language Total	3.68	14.07	7	> .80	< .90	Accepted
W_1 Map Reading	9.72	14.07	7	> .20	< .30	Accepted
W_2 Reading Graphs and Tables	6.27	14.07	7	> .50	< .70	Accepted
W_3 Knowledge and Use Ref. Mat.	7.51	15.51	8	> .30	< .50	Accepted
A_1 Arithmetic Con.	3.24	14.07	7	> .80	< .90	Accepted
A_t Arithmetic Tot.	5.77	14.07	7	> .50	< .70	Accepted
LORGE-THORNDIKE						
Non-Verbal	21.27	19.68	11	> .02	< .05	Rejected
Total	18.14	18.31	10	> .05	< .10	Accepted

were found to depart significantly ($p<.05$) from the critical chi squared values as tabled. It is therefore concluded that the normality assumption underlying the multiple discriminant model is essentially met by the total junior high group test score distributions which were the parent population for the variables included in the discriminant equations.

Homogeneity of Within-Group Variance Assumption

The null hypothesis under test is that there are no significant differences in the within-group variances across levels, within subject areas. Bartlett's test (Edwards, 1960, p. 127) is employed to test this assumption. A summary of the results of applying Bartlett's test is presented in Table L. These results indicate that the null hypothesis of equal within-groups variance is accepted in every instance except one, the Lorge Total. It is concluded therefore that the assumption of homogeneity of within-group variance is essentially met by the variables included in the discriminant equations.

Major Hypothesis IV

The two versions of the discriminant equations were developed on the junior high group validation sample (see Chapter III, METHOD, pages 60ff.) and were then applied to a check sample. The proportion of students classified within levels of the validation and check sample were then compared with the actual proportion (a priori) of students that constituted the groups at the end of the first semester ninth grade of the validation and check sample years. These data are presented in Tables LI and LII.

It is apparent from an inspection of the proportions recorded in this table that, in comparison with the a priori proportions, both the

TABLE L

SUMMARY OF THE RESULTS OF BARTLETT'S CHI SQUARED TEST OF THE HOMOGENEITY OF WITHIN-GROUP VARIANCE ASSUMPTION AS APPLIED TO THE DISTRIBUTION OF ITBS AND LORGE THORNDIKE SCORES USED IN THE MULTIPLE DISCRIMINANT EQUATIONS

Junior High Group Validation Sample

Subject area and Variables	Observed Chi Squared	p	Disposition of Hypothesis
English			
L_1 Spelling	0.17		Accepted
L_4 Usage	5.99		Accepted
L_t Lang. Total	2.41		Accepted
W_3 Kn. & Use of Ref. Mat.	1.24		Accepted
A_1 Arith. Con.	2.18		Accepted
Lorge Total	7.91	<	Rejected
Mathematics			
L_t Lang. Total	0.50		Accepted
W_1 Map Reading	0.48		Accepted
W_2 Rdg. Graphs & Tables	0.33		Accepted
A_1 Arith. Con.	0.67		Accepted
A_t Arith. Total	1.32		Accepted
Lorge Non-Verbal	3.79		Accepted

Chi Squared, .05, 2df = 5.99

TABLE LI

COMPARISON WITH THE *A PRIORI* PROPORTIONS, THE PROPORTIONS OF STUDENTS PREDICTED/CLASSIFIED WITHIN ACHIEVEMENT LEVELS OF TWO SUBJECT AREAS BY MEANS OF TWO VERSIONS OF THE MULTIPLE DISCRIMINANT EQUATIONS

Junior High Validation Sample

English

Instructional Levels	Number of students assigned by:					
	Multiple Discriminant				a Priori	
	Unadjusted		Adjusted			
	N	Prop.	N	Prop.	N	Prop.
Basic Composition Level 1	44	0.28	23	0.15	25	0.16
Composition Level 2	63	0.40	96	0.61	97	0.62
Composition (Honors) Level 3	50	0.32	38	0.24	35	0.22
Totals	157		157		157	

Mathematics

Instructional Levels	Unadjusted N	Unadjusted Prop.	Adjusted N	Adjusted Prop.	a Priori N	a Priori Prop.
General Math Level 1	74	0.28	59	0.22	53	0.20
Algebra Level 2	138	0.53	162	0.62	171	0.65
Algebra (Honors) Level 3	51	0.19	42	0.16	39	0.15
Totals	263		263		263	

TABLE LII

COMPARISON WITH THE A PRIORI PROPORTIONS, THE PROPORTIONS OF STUDENTS PREDICTED/CLASSIFIED WITHIN ACHIEVEMENT LEVELS OF TWO SUBJECT AREAS BY MEANS OF TWO VERSIONS OF THE MULTIPLE DISCRIMINANT EQUATIONS

Junior High Check Sample

English

Instructional Levels	Number of students assigned by:					
	Multiple Discriminant				a Priori	
	Unadjusted		Adjusted			
	N	Prop.	N	Prop.	N	Prop.
Basic Composition Level 1	55	0.37	46	0.31	25	0.17
Composition Level 2	64	0.43	76	0.51	83	0.56
Composition (Honors) Level 3	29	0.20	26	0.18	40	0.27
Totals	148		148		148	

Mathematics

Instructional Levels	Unadjusted N	Unadjusted Prop.	Adjusted N	Adjusted Prop.	a Priori N	a Priori Prop.
General Math Level 1	95	0.37	68	0.26	40	0.15
Algebra Level 2	130	0.50	168	0.65	193	0.75
Algebra (Honors) Level 3	34	0.13	23	0.09	26	0.10
Totals	259		259		259	

unadjusted and adjusted discriminant equations over-classified students in instructional Level 1 of both subject areas. The proportion classified by the unadjusted equation in Basic Composition and General Mathematics exceeded a two to one ratio over the a priori proportion while the proportion classified by the adjusted equation is approximately a three to two ratio over the a priori proportion for these same levels. Examination of the proportions classified in Level 3 by the two versions of the discriminant equations reveals that they tend to underclassify students in this level in relation to the a priori proportions--the smallest proportion being associated with the adjusted equation.

The question which arises is whether or not the discrepant proportions are statistically significant. This question is examined next as defined by Major Hypothesis IV which states that there are no significant differences in the proportion of students predicted/classified within levels of English and mathematics by means of the basic (or unadjusted) and the adjusted versions of the multiple discriminant equations and the proportions expected a priori. (Region for rejection: $\alpha = .05$).

Presented in Table LIII is a summary of the results of the application of the single classification chi squared test to the observed proportions within levels yielded by each version of the discriminant equation. The validation sample data is included for comparison purposes; the hypothesis relates specifically to the check sample.

It can be noted that for the unadjusted equation, the null hypothesis is rejected at all but one instructional level (Algebra Honors) of the validation sample and at all but two levels (Composition Honors

TABLE LIII

SUMMARY OF THE SINGLE CLASSIFICATION CHI SQUARED TEST OF SIGNIFICANCE OF DIFFERENCES BETWEEN PROPORTION OF STUDENTS CLASSIFIED BY MEANS OF DISCRIMINANT EQUATIONS AND THE EXPECTED A PRIORI PROPORTIONS, AT EACH INSTRUCTIONAL LEVEL

Junior High Group Validation and Check Samples

Subject Area and Level	Observed Chi Squared	
	Validation Sample	Check Sample
UNADJUSTED EQUATIONS		
English		
Basic Composition	14.44***	36.00***
Composition	11.92***	4.35*
Composition (Honors)	6.43*	3.00
Mathematics		
General Mathematics	8.32**	75.63***
Algebra	6.37*	20.56***
Algebra (Honors)	3.69	2.46
ADJUSTED EQUATIONS		
English		
Basic Composition	.16	17.60***
Composition	.01	.59
Composition (Honors)	.26	4.90*
Mathematics		
General Mathematics	.68	19.60***
Algebra	.47	3.23
Algebra (Honors)	.23	.35

α set at .05; df = 1 * <.05
Chi squared, .05 = 3.84; Chi squared, .01 = 6.64; ** <.01
Chi squared, .001 = 10.83 *** <.001

and Algebra Honors) in the check sample; it is accepted at Level 3 mathematics in the validation sample and at Level 3 mathematics and English in the check sample. A summation of the check sample chi squared values in both subject areas leads to an over-all rejection of the null hypothesis for the unadjusted equation (English: chi squared = 43.35, df = 2, $p<.001$; Mathematics: chi squared = 98.65, df = $p<.001$).

A review of the observed chi squared values for the adjusted equations listed in Table LIII indicates that, in the validation sample, the null hypothesis is accepted at all instructional levels in both subject areas while, in the check sample, it is rejected at Levels 1 and 3 in English and Level 1 in mathematics. The size of the check sample chi squared values at the levels in which the hypothesis is rejected (particularly at Levels 1) leads to an over-all rejection of the null hypothesis in each subject area when all three chi squared values are summed in each subject area (English: chi squared = 23.09, df = 2, $p<.001$; Mathematics: chi squared = 23.18, df = 2, $p<.001$).

The findings just reported are not unexpected in the case of the unadjusted equations since students were classified into homogeneous groups without regard to a priori proportions.

However, in the case of the adjusted equations, the findings are surprising as it is apparent that in the validation sample the adjusted equations were functioning effectively against a priori proportions. It is not clear why, when applied to check sample data, these equations did not function as well. It is evident that homogeneous grouping by means of the adjusted discriminant equations in this particular check sample would place a larger number of students in the lowest level of both subject areas.

Nevertheless, a priori proportions are only one criterion source for the discriminant equations and by themselves provide only part of the picture. The other element is the success factor as defined by observed achievement (i.e., grades) and this factor is evaluated in the next section.

Testing Minor Hypothesis IV

The minor hypothesis to be focused on here is concerned with whether or not, in the check sample, the two versions of the multiple discriminant equation correctly classify (group) students into three instructional levels of English and mathematics as well or better than a random procedure.

The procedure for arriving at the expected number of chance 'hits' differed for each version of the discriminant equation. For the unadjusted equation, the underlying assumption is that equal numbers (densities) occur in each of the groups; therefore, since there are three criterion groups, the number of 'hits' yielded by a random procedure would be one-third of the total number of students in a subject area.

In the case of the adjusted equations (as also for the regression equation) the procedure for determining the total number of students expected to be placed correctly in each subject area by chance was based on the a priori densities. Since the a priori densities within each instructional level in each subject area vary, the total number of 'hits' by chance was arrived at by weighted values (see Appendix, page 198 for an illustration of the computational procedures for arriving at these chance expectations).

A summary of the hits/misses tallies by levels and over-all levels in English and mathematics recorded for both versions of the discriminant equations together with the total numbers of hits/misses expected by chance in the check sample is presented in Table LIV. It is evident that in both subject areas for both versions of the discriminant equations, the over-all number of 'hits' yielded by these equations surpasses the total number of 'hits' expected if a random classification procedure were employed.

To determine whether the over-all hit/misses tallies for each discriminant equation in both subject areas depart significantly from the numbers expected by chance, the single classification chi squared test was applied, the results of which are summarized in Table LV. Examination of the probability levels listed indicates that the total number of correct predictions/classifications yielded by both versions of the multiple discriminant equation exceeds chance expectations at a highly significant level.

The next question of concern in this investigation is a comparison among actuarial methods (multiple regression and two versions of the multiple discriminant equations) of the over-all number of 'hits' achieved. This question is defined by Major Hypotheses V, VI and VII which are presented in the following section.

Following is PHASE FOUR of the problem: Comparison between Methods in Predicting/Classifying Students Within Levels.

Major Hypothesis V

This hypothesis states that there are no significant differences between the number of successful a priori check sample students and the number correctly predicted/classified in two subject areas by each statistical method.

TABLE LIV

SUMMARY OF THE COMPARISON WITH TOTAL CHANCE EXPECTATIONS OF THE TOTAL NUMBER OF STUDENTS CORRECTLY PREDICTED/CLASSIFIED (HITS) IN TWO SUBJECT AREAS BY TWO VERSIONS OF THE MULTIPLE DISCRIMINANT EQUATION

Junior High Group Check Sample

Levels of Achievement	Actual Number (a priori)	Number of Students Correctly Predicted by Unadjusted Discr. Eq.		Expected by Chance		Number of Students Correctly Predicted by Adjusted Discr. Eq.		Expected by Chance	
		Hits	Misses	Hits	Misses	Hits	Misses	Hits	Misses
English									
Basic Composition Level 1	25	25	0			25	0		
Composition Level 2	83	47	36			58	25		
Composition (Honors) Level 3	40	21	19			20	20		
Total	148	93	55	49.33	98.67	103	45	75.097	72.903
Mathematics									
General Math Level 1	40	35	5			27	13		
Algebra Level 2	193	125	68			149	44		
Algebra (Honors) Level 3	26	19	7			18	8		
Total	259	179	80	86.33	172.67	194	65	152.606	106.374

TABLE LV

SUMMARY OF THE RESULTS OF THE SINGLE CLASSIFICATION CHI SQUARED TEST FOR DEPARTURE OF CORRECT PREDICTIONS/CLASSIFICATIONS BY THE MULTIPLE REGRESSION AND BY TWO VERSIONS OF THE MULTIPLE DISCRIMINANT EQUATIONS FROM THOSE EXPECTED BY CHANCE

Junior High Group Check Sample

English

Type of Equation	Observed Chi Squared	p	Disposition of Hypothesis
Multiple Regression	32.93	$< .001$	Rejected
Multiple Discriminant Unadjusted	57.99	$< .001$	Rejected
Adjusted	21.05	$< .001$	Rejected

Mathematics

Type of Equation	Observed Chi Squared	p	Disposition of Hypothesis
Multiple Regression	30.40	$< .001$	Rejected
Multiple Discriminant Unadjusted	104.45	$< .001$	Rejected
Adjusted	26.91	$< .001$	Rejected

Chi Squared, .001, 1df = 10.83

The data from Table LVI, page 175, that are being compared are the total numbers of students correctly predicted by each statistical method and the total number of students succeeding in each a priori subject area group. Thus, the total a priori number succeeding in English, 146 (two students registered in Level 2 had first semester GPA's of F; two 'miss' tallies were therefore recorded for the a priori group), is compared with the 93 'hits' tallied for the unadjusted equation. For mathematics, the total a priori number succeeding, 251 (eight students, registered in Level 2 had GPA's of F; therefore eight 'misses' were tallied), is compared with the 179 'hits' tallied for the unadjusted equation. Differences between each pair of total 'hits' (a priori versus each method) are evaluated for statistical significance by the application of McNemar's chi squared test for correlated samples (Siegel, 1956, pp. 63-67).

It can be readily noted in Table LVII that, as might well be expected from the differences in the total number of 'hits' for each statistical method compared with a priori 'hits', significant departures are recorded for all comparisons. It would appear from the probability levels reported that, since the regression equation yields a 'hits' total for each subject that is the least significantly different of the three models ($p<.05$) from the a priori 'hits', it is the most effective of the three equations in approximating the a priori 'hits' criterion while the adjusted equation ranks next and the unadjusted equations are the least effective. The question is then posed: Are there significant differences between the methods themselves in the number of 'hits' achieved? Before this question is considered (under Hypothesis VII) attention is first given to a further question to be

TABLE LVI

SUMMARY OF THE COMPARISON OF THREE METHODS OF CORRECTLY PREDICTING/CLASSIFYING STUDENTS WITHIN THREE LEVELS OF TWO SUBJECT AREAS, ENGLISH AND MATHEMATICS

(Junior High Group Check Sample)

English

Level	Actual N (a priori)	Numbers and Proportion of Hits/Misses by: Multiple Discriminant, Unadjusted Eq. Hits	Prop	Misses	Prop	Multiple Discriminant, Adjusted Eq. Hits	Prop	Misses	Prop	Multiple Regression Eq. Hits	Prop	Misses	Prop
Basic Composition Level 1	25	25	.17	0	--	25	.17	0	--	19	.13	6	.04
Composition Level 2	83	47	.32	36	.24	58	.39	25	.17	70	.48	13	.08
Composition (Hon) Level 3	40	21	.14	19	.13	20	.14	20	.14	21	.14	19	.13
Sub Total	148	93	.63	55	.37	103	.70	45	.31	110	.75	38	.25
Total	148	148				148				148			

Mathematics

Level	Actual N (a priori)	Multiple Discriminant, Unadjusted Eq. Hits	Prop	Misses	Prop	Multiple Discriminant, Adjusted Eq. Hits	Prop	Misses	Prop	Multiple Regression Eq. Hits	Prop	Misses	Prop
General Math Level 1	40	35	.14	5	.02	27	.10	13	.05	22	.08	18	.07
Algebra Level 2	193	125	.48	68	.26	149	.58	44	.17	158	.61	35	.14
Algebra (Honors) Level 3	26	19	.10	7	.03	18	.07	8	.03	17	.07	9	.03
Sub Total	259	179	.69	80	.31	194	.75	65	.25	197	.76	62	.24
Total	259	259				259				259			

TABLE LVII

SUMMARY OF THE McNEMAR CHI SQUARED TEST OF THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE A PRIORI NUMBER OF STUDENTS SUCCEEDING AND THE OVER ALL NUMBER CORRECTLY PREDICTED/CLASSIFIED IN TWO SUBJECT AREAS BY EACH STATISTICAL METHOD

Junior High Group Check Sample

Subject Areas	Observed Chi Squared: Regression Equations	Observed Chi Squared: Unadjusted Discriminant Equation	Observed Chi Squared: Adjusted Discriminant Equation
English	4.79*	11.31***	7.08**
Mathematics	6.27*	11.72***	7.05**

$\alpha = .05$
Chi squared, .05 = 3.84; Chi squared, .01 = 6.64; Chi squared, .001 = 10.83; df = 1
* <.05
** <.01
*** <.001

TABLE LVIII

SUMMARY OF THE z TEST OF THE SIGNIFICANCE OF DIFFERENCES BETWEEN ENGLISH AND MATHEMATICS IN THE NUMBER OF STUDENTS CORRECTLY PREDICTED/CLASSIFIED BY THE MULTIPLE REGRESSION AND TWO VERSIONS OF THE MULTIPLE DISCRIMINANT METHODS

Junior High Group Check Sample

Statistical Method	z	Disposition of Hypothesis
Multiple Regression	1.24	Accepted
Multiple Discriminant Unadjusted Equations	1.16	Accepted
Adjusted Equations	1.29	Accepted

$\alpha = .05$; z .05 = 1.96; df = ∞

raised concerning whether each actuarial method is equally effective between subject areas in the prediction/classification task. This question is defined by Major Hypothesis VI, considered next.

Major Hypothesis VI

This hypothesis is concerned with the question whether each statistical method is more successful in correctly predicting/classifying students in English or in mathematics.

To test this hypothesis, the _z_ test of the significance of difference between uncorrelated proportions (Walker and Lev, 1953, p. 78) is applied to the over-all hits/misses data presented in Table LVI, page 17 . A summary of the results of the _z_ test is presented in Table LVIII. No significant differences were found. This is not a surprising result in view of the relatively small observed differences in the hits/misses proportion between subject areas for each method. The conclusion therefore is that each statistical method demonstrated the same degree of effectiveness in both English and mathematics in the over-all number of students correctly predicted/classified.

Attention now is given to the question on which this investigation is concluded, a comparison between the statistical methods of their prediction/classification effectiveness.

Testing Major Hypothesis VII

This hypothesis states that there are no signficant differences between statistical prediction procedures in the over-all number of students correctly predicted/classified.

The hit/misses ratio recorded in Table LVI provided the data for the test of this hypothesis. A summary of the results of the application of McNemar's chi squared test for correlated samples is presented in Table LIX. It is evident from the observed chi squared values, all

TABLE LIX

SUMMARY OF THE RESULTS OF THE APPLICATION OF THE McNEMAR CHI SQUARED TEST FOR THE SIGNIFICANCE OF DIFFERENCES BETWEEN METHODS OF CORRECTLY PREDICTING/ CLASSIFYING STUDENTS OVER ALL LEVELS OF TWO SUBJECT AREAS

Junior High Group Check Sample

Subject Area	Observed Chi Squared		
	Regression Eq. Versus Adjusted Eq.	Regression Eq. Versus Unadjusted Eq.	Adjusted Eq. Versus Unadjusted Eq.
English	0.21	1.42	0.51
Mathematics	0.02	0.86	0.60

α set at .05
Chi squared, .05 = 3.84, 1df

of which are smaller than the tabled values for rejection ($p > .05$), that the three statistical prediction methods are equally effective in the over-all number of correct classifications yielded. Thus, while it was found under Major Hypothesis V that, when the criterion against which each method was compared in correctly classifying students was the a priori success pattern, the prediction methods yielded significantly fewer correct classifications, the differences between methods are not significant.

CHAPTER V

SUMMARY

Overview

This investigation was designed to study, within the framework of a computer-based measurement system, the effectiveness of the multiple correlation/regression and two versions of the multiple discriminant models for predicting/classifying students within instructional levels of four first semester ninth grade subject areas, English, mathematics, social studies and foreign languages.

A validation sample of 544 students (designated the combined group) representing 82% of the total first semester ninth grade class of 1966 (year of high school graduation) of a suburban high school in the metropolitan Chicago area and a sub-sample of the combined group consisting of 272 students (designated the junior high group) comprised the population on which the statistical prediction/classification procedures were developed.

A check sample of 604 first semester ninth grade students (again referred to as the combined group) representing 91% of the total ninth grade class of 1967 (i.e., the new incoming ninth grade class) and a sub-sample of the combined group consisting of 259 students (again called the junior high group) were identified as the populations on which the statistical methods developed on the validation sample would be evaluated for predictive/classificatory stability.

Phase One of the Study Problem

The first phase of the study problem was to map out the predictive validity domain of a rather large number of test and certain non-test variables. With the domain of the predictive validity of the data pool thus developed, the other three phases of the study problem would then be explored. With this objective in view, the first task was to develop a set of prediction equations within each instructional level of four subject areas, using the step-wise regression routine. A variety of independent variables from a data pool (primarily the ITED battery for the combined sample and the ITBS and Lorge batteries for the junior high sample, with the addition of other test and non-test variables in selected instances) were entered as input to the step-wise multiple regression routine for estimating the criterion variable, end-of-first-semester-ninth-grade grade point average (GPA) within instructional levels of four subject areas: English, mathematics, social studies and foreign language. As each variable was selected by the step-wise regression routine it was tested by the analysis of multiple regression procedure for its statistical significance in reducing the error sums of squares, yielding information concerning its possible inclusion in an optimally effective prediction equation. In general, it was found that seldom were six variables needed for an efficient prediction equation for estimating the criterion variable. A more usual number was three; however, the recommended number of variables for estimating the criterion GPA varied with each instructional level of each subject area. Certain predictor variables were found to appear more frequently in the equations--for example, Correctness of Expression and Use of Sources of Information, subtests from the ITED battery. Certain variables appeared seldomly, if at all, as for instance the ITBS Reading subtest and the Lorge Verbal, Non-verbal and Total scores.

When five junior high subject area GPA's were added as independent variables to the data pool along with the variables from the ITBS and Lorge batteries for estimating the GPA's within the junior high group instructional levels of two subject areas, English and mathematics, no junior high GPA's were entered into the equations except in the cases of algebra and algebra (honors) where they then produced a significantly larger multiple correlation coefficient than was yielded by the equation which consisted of variables drawn only from ITBS and Lorge data pool.

Each validation sample prediction equation yielded a multiple correlation coefficient (the range for all equations: .44 to .81) which was then corrected for bias, a procedure taking into account the number of independent variables in the data pool in relation to sample size. It was found that the attenuated values displayed a shrinkage ranging from a change in the second decimal place to R values that diminished to zero. This finding emphasizes the importance of applying prediction equations to new samples to check out their stability. The value of this application to check samples was especially demonstrated, for example, in the case of the impressive-looking multiple R's yielded by predictor variables drawn from a pool of independent variables that nearly approached the sample size. When corrected for bias, these multiple R's were reduced to zero. In this investigation, when the regression equations were applied to new samples in selected subject areas, the resulting multiple R's did not differ significantly from the validation sample multiple R's (except at one instructional level, general mathematics, where the check sample multiple R was significantly larger than its validation sample counterpart). It was concluded that the predictive stability of the regression equations, when applied to new samples, was demonstrated.

Phase Two of the Study Problem

Phase two of this investigation was concerned with studying the effectiveness of the multiple correlation/regression model in predicting/ classifying students within instructional levels of two subject areas, English and mathematics. To accomplish this evaluation, validation and check samples of the junior high group were identified. In this phase of the study, the regression model was asked to function within the adjusted discriminant model's framework, that is, to classify the individuals according to the a priori groups. To accomplish this, an estimated GPA for each student in the check sample was computed by using the Level 2 regression equations. The mean and standard deviation of each estimated GPA distribution was determined; each standard deviation was then adjusted to the Level 1 and 3 a priori proportions. Students were then classified in Level 1 or 3 according to whether their estimated GPA was lower or higher than the adjusted standard deviation. The remaining students in each subject area were then placed in Level 2.

When the regression model's predictions were compared against a priori proportions, the regression equations yielded proportions that did not depart significantly from a priori proportions. When these predictions were evaluated against a correct predictions criteria, they performed significantly better than chance expectations at the .001 probability level.

Phase Three of the Study Problem

In phase three of the study, the prediction/classification capabilities of the multiple discriminant model were studied. The first task was to develop two sets of multiple discriminant equations, using the a priori validation groups as the criterion and a specified set of test scores as the independent variables. The purpose of this method was to separate students with similar test score characteristics into groups, thereby increasing the differences between groups. The equations thus derived (one set for each instructional level of English and mathematics, six in all) were applied to a check sample in two versions: 1) in their basic or unadjusted form and 2) in an adjusted form to conform to a priori proportions, according to a formula devised by Simpson (1957).

The evaluation proceeded similarly to that of the regression equations; first, a comparison was made with a priori check sample proportions. Here, both versions of the discriminant equations classified students within levels in proportions that were significantly different, over-all, from a priori proportions; however, the adjusted equations performed much better than the unadjusted, succeeding in three out of six levels, failing most conspicuously in the lower levels of both subject areas. When the correctness of the predictions were considered against chance, both versions performed significantly better than chance ($p < .001$).

Phase Four of the Study Problem

The fourth phase of the problem was a comparison between models (multiple regression and two versions of the multiple discriminant model) against various criteria. To accomplish this task, a hits/misses criterion was used, employing observed GPA's as the criterion. The actuarial methods achieved 'hits' which ranged from 63% (for the unadjusted discriminant equation) to 75% (for the regression equation) correct classifications in English and from 69% (unadjusted equation) to 76% (regression) in mathematics. The findings were: 1) An over-all comparison of hits/misses with a priori successes (i.e., the student was registered and succeeding in the level at the end of first semester, ninth grade) yielded significant differences. 2) No differences were found in the prediction/classification effectiveness of the three methods between the two subject areas, English and mathematics. 3) When the numbers of hits/misses were compared on a between-pairs-of-methods basis, the differences were found to be non-significant, demonstrating that the three models correctly predicted/classified students with equal effectiveness.

The evidence presented in phases two, three and four of this study problem, i.e, the parts concerned with the statistical prediction models' grouping abilities in check sample data, suggest the following conclusions:

1. The regression model, with associated adjustments, predicted/classified students within instructional levels in proportions which did not differ significantly from a priori expectancies.
2. In performance of the same task, the unadjusted and adjusted discriminant models did not predict/classify students in accordance with a priori proportions, i.e., their allocations departed significantly from the expected proportions. In the development of the equations on a validation sample the unadjusted version departed significantly from a priori

expectation in allocating students while the adjusted version did not.

3. When the task was that of correctly predicting/classifying students within a priori groups against chance expectancies, each method exceeded random assignment well beyond chance probabilities.
4. The total number of students correctly predicted/classified in the two junior high group subject areas, English and mathematics, by each statistical method departed significantly from the total a priori number succeeding in these groups.
5. No significant differences occurred in the total number of students each statistical method correctly predicted across two subject areas.
6. When each statistical prediction method was compared in turn with the other two against the criterion of the total number of students correctly predicted/classified, there were no statistically significant differences between them.

The evidence presented in this summation suggests that other factors for selecting one method over the other need to be considered. It would appear that the trend of the findings here would tend to favor the multiple regression model (e.g., the degree to which it approximated the a priori proportions). However, it is possible that some of the success associated with the regression model may be attributable to an artifact of the adjustment procedure employed in conjunction with this model (See Chapter III, page 57). Yet, the very simplicity of the 'adjustment' procedure is in line with the recommendation made by Laushe and Shuckler (1959) as a result of their investigation of a number of prediction methods (see Chapter II, page 36) that less sophisticated methods have much to recommend them because of their simplicity and because they may function as effectively as the more sophisticated ones.

There is, in this assessment against a priori grouping criteria, an implicit acceptance of the quasi-clinical method as the basis for comparison. Nevertheless, this is the only possible basis outside of an experimentally controlled situation; i.e., if one cannot experimentally assign a sample of students by actuarial methods while another sample is assigned in the quasi-clinical manner and still another sample in a rigorously defined clinical way, one must work with the groups as constituted. It would seem that the primary value of this study lies in the demonstration for other investigators and, through them, for educational administrators, that it would be safe to embark on an experimental comparison of these two methods (multiple correlation/regression and adjusted multiple discriminant) since they both functioned at considerably better than chance levels in correctly predicting/classifying students while at the same time they closely approximated (albeit to a lesser degree in the case of the adjusted discriminant) required proportions. Assurance of this kind is an absolute necessity, in the writer's opinion, before real 'flesh and blood' children are used as guinea pigs, as it were. In other words, before students' lives within the grouping process of the educational system are affected by a computer-based measurement system, there must be a validation-and-check-sample-based statistical exploration. This, in essence, is what the writer hopes to have accomplished--to have contributed to this exploration.

Implications for Research

The present study suggests, it seems, a number of avenues for further investigative effort. These might be:

1. To replicate the methodology guiding the present study but with a larger sample. The writer feels that Simpson's (1957) procedure for adjusting the discrminant equation merits further investigative effort.

2. To modify certain features of the methodology of this study. A beginning point here would be to consider designing one's own study by removing as many as possible of the limitations that were described in Chapter 1, pp. 11-14. Other entry points for design modification might be:

    a. Employ the step-wise discriminant routine (see Chapter III, page 60, footnote 4) as a way of selecting the variables.

    b. Consider different approaches to effecting comparability of the criterion variable, GPA, across instructional levels. Whatever the approach that is used, it should be objective and capable of being replicated.

3. To extend the present investigation into areas suggested by the methodology as described and/or by the ideas that may be gleaned from Chapter II, Review of the Literature. One area that might well be considered is an incorporation of Simpson's (1957) formula for adjusting the discriminant equation as an element in Tatsouka's joint probability model (1957; see also Rulon et al., 1967).

4. To apply the methodology of the present investigation by employing a 'True' or quasi-experimental design approach (Campbell and Stanley, 1963), and incorporate a quasi-clinical and/or clinical model as comparative element(s) in the design. Meehl's discussion (1954) is obviously relevant here as is also a more recent one by Sawyer (1966). Before embarking on such a strategy, however, one would do well to establish the validity of the statistical methodology and predictor variable domain in one's own situation. Intriguing lines of investigative effort would seem to lie in this area.

BIBLIOGRAPHY

Anderson, L. W. and Van Dyke, L. A. Secondary school administration. Boston: Houghton Mifflin, 1963.

Barnes, W. E. and Asher, J. W. Predicting students' success in first year algebra. Mathematics Teachers, 1962, 55; 651-54.

Bryan, J. G. A method for the exact determination of the characteristic equation and latent vectors of a matrix with application to the discriminant function for more than two groups. Unpublished doctoral dissertation, Harvard University, 1950.

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Carroll, J. B. The prediction of success in intensive language training. In R. Glaser (Ed.), Training research and education, Pittsburgh: University of Pittsburgh Press, 1962.

Cooley, W. W. A computer-measurement system for guidance. Harvard Education Review, 1964, 34, 559-572.

Crawford, D. A. The administrative organization of the curriculum in midwestern high schools. Unpublished doctoral dissertation, State University of Iowa, 1955.

Cronbach, L. J. Essentials of psychological testing (2nd Ed.) New York: Harper & Brothers, 1960.

Cronbach, L. J. and Gleser, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1957.

Diederich, P. The problem of grading essays. Princeton, N. J.: Educational Testing Service, 1957. (mimeographed).

Dinkel, R. E. Prognosis for studying algebra. Arithmetic Teacher, 1959, 6, 317-319.

Dixon, W. J. (Ed.) BMD Biomedical computer programs. (2nd Rev. Ed.) Los Angeles: Health Sciences Computing Facility, School of Medicine, University of California, 1964.

Dixon, W. J., and Massey, F. J. Introduction to statistical analysis, New York: McGraw-Hill, 1956.

Dunn, F. E. Two methods for predicting the selection of a college major. Journal of Counseling Psychology, 1959, 6, 15-27.

Edwards, A. L. Experimental design in psychological research. (2nd Ed.) New York: Holt, Rhinehart & Windston, 1960.

Ekstrom, R. Experimental studies of ability grouping. School Review, 1961, 69, 216-26.

Fahnle, C. E. The relative effectiveness of measures of grade school achievement in predicting scores on high school tests. Unpublished master's thesis, Iowa City: State University of Iowa, 1942.

Fisher, W. D. and Masia, B. B. In O. K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, N. J.; The Gryphon Press, 1965, 633-635.

Fleming, W. G. In O. K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press, 1965, 1084-1086.

Freeman, F. S. In O. K. Buros (Ed.), The fifth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press, 1959, 479-481.

French, J. W. The logic and assumptions under-lying differential testing. In Proceedings of the 1966 invitational conference on testing problems. Princeton, N. J.: Educational Testing Service, 1955, 40-48.

Fullmer, D. W., Benard, H. W. Counseling: content and process. Chicago: Science Research Associates, 1964.

Garrett, H. E. The discriminant function and its use in psychology. Psychometrika, 1943, 8, 65-79.

Gibbons, C. W. Contribution of selected variables in predicting academic achievement as measured by teacher grades. Unpublished doctoral dissertation, University of Maryland, 1962.

Guilford, J. P. Fundamental statistics in psychology and education. (4th Ed.) New York: McGraw-Hill, 1965.

Guilford, J. P., Hoepfner, R., and Peterson, H. Predicting achievement in ninth-grade mathematics from measures of intellectual aptitude. Educational and Psychological Measurement, 1965, 25, 659-682.

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

Hascall, E. O. Predicting success in high school foreign language. Unpublished doctoral dissertation, University of Michigan, 1959.

Helmstadter, G. C. An empirical comparison of methods for estimating profile similarity. _Educational and Psychological Measurement_, 17, 1957, 71-82.

Helmstadter, G. C. _Principles of psychological measurement_. New York: Appleton-Century-Crofts, 1964.

Herrick, V. E. In O. K. Buros (Ed.), _Fifth mental measurement yearbook_. Highland Park, N.J.: The Gryphon Press, 1959, 30-34.

Hoyt, C. J. and Johnson, M. C. Regression and correlation. _Review of Educational Research_, 1954, 24, 393-401.

Jacobs, J. N. Aptitude and achievement measures in predicting high school academic success. _Personnel and Guidance Journal_, 1959, 38, 334-341.

Johnson, P. O. _Statistical methods in research_. New York: Prentice-Hall, 1949.

Lavin, D. E. _The prediction of academic performance_. New York: Russell Sage Foundation, 1965.

Lawshe, C. H. and Schucker, R. E. The relative efficiency of four test weighting methods in multiple prediction. _Educational and Psychological Measurement_, 1959, 19, 103-114.

Layton, W. L. The relation of ninth grade test scores to twelfth grade scores and high school rank. _Journal of Applied Psychology_, 1954, 38, 10-11.

Leton, D. A. and Anderson, H. E. Discriminant analysis of achievement characteristics for multi-grade grouping of students. _The Journal of Experimental Education_, 1964, 32, 293-297.

Lindquist, E. F. and Hieronymus, A. N. _Manual for administrators, supervisors and counselors, Iowa Tests of Basic Skills_. Boston: Houghton Mifflin Co., 1964.

Lorge, I. and Thorndike, R. L. _Technical manual for administrators, directors of testing and research_. Boston: Houghton Mifflin Co., 1962.

Ludlow, E. D. A study of the relative efficiency of multiple regression analysis and multivariate discriminant analysis in predicting academic achievement success in college. Unpublished doctoral dissertation, University of Tulsa, 1962.

McLaughlin, K. P. The relation of performance on the Iowa tests of basic skills to subsequent school achievement and to persistence of school attendance. Unpublished doctoral dissertation, State University of Iowa, 1950.

Markwardt, F. C. Pattern analysis techniques in the prediction of college success. Dissertation Abstracts, 1960, 21, 2990.

Meehl, P. E. Clinical versus statistical prediction. Minneapolis: University of Minnesota Press, 1954.

Morgan, G. A. V. In O. K. Buros (Ed.), The fifth mental measurement yearbook. Highland Park, N. J.: The Gryphon Press, 1959, 34-36.

Morton, M. P. An experiment in the use of discriminant classification equations in grouping incoming tenth grade students at MacLain High School. Unpublished Manuscript, (c/o F. W. Simpson), University of Tulsa, 1961.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Owen, A. M. A comparative study of methods of predicting success in high school. Unpublished doctoral dissertation, University of Texas, 1956.

Pickrel, E. W. Classification theory and techniques. Educational and Psychological Measurement, 1958, 18, 37-46.

Pimsleur, P., Stockwell, R. P., and Comrey, A. L. Foreign language ability. Journal of Educational Psychology, 1962, 53, 15-26.

Rao, C. R. Advanced statistical methods in biometric research. New York: John Wiley & Sons, Inc., 1952.

Rulon, P. J., Tiedman, D. V., Tatsuoka, M. M., and Langmuir, C. R. Multivariate statistics for personnel classification. New York: John Wiley & Sons, Inc., 1967.

Sawyer, J. Measurement and prediction, clinical and statistical. Psychological Bulletin, 1966, 66, 178-200.

Scannell, D. P. Differential prediction of academic success from achievement test scores. Unpublished doctoral dissertation, State University of Iowa, 1958.

Schusler, M. M. Prediction of grades by computer for high school students: a cross-validation and experimental placement study. Dissertation Abstracts, 1964, 26, 1460.

Siegel, S. Nonparametric statistics. New York: McGraw-Hill, 1956.

Simpson, F. W. The applicability of multiple discriminant analysis to the interpretation of classification test scores at the University of Kansas. Unpublished doctoral dissertation, University of Kansas, 1957.

Tatsuoka, M. M. Joint-Probability of membership and success in a group: an index which combines the information from discriminant and regression analysis as applied to the guidance problem. Unpublished doctoral dissertation, Harvard University, 1955.

Tatsuoka, M. M. and Tiedman, D. W. Discriminant analysis. Review of Educational Research, 1954, 24, 402-420.

Thomas, R. M. The extent of ability grouping in English. Peabody Journal of Education, 1966, 43, 208-211.

Tiedman, D. W., Rulon, P. J., and Bryan, J. G. The multiple discriminant function--a symposium. Harvard Educational Review, 1951, 21, 71-95.

Travers, R. M. W. Significant research on the prediction of academic success. In W. T. Donahue, C. H. Coombs and R. M. W. Travers (Eds.), The measurement of student adjustment and achievement. Ann Arbor: The University of Michigan Press, 1949.

Traxler, A. E. (Ed.) Measurement and research in today's schools. Washington, D.C.: American Council on Education, 1961.

Van Dyke, L. A. and Sparks, J. N. Four-state survey of secondary school marking practices. Research digest, No. 2, Iowa City: Iowa Center for Research in School Administration, 1960.

Walker, H. M. and Lev, J. Statistical inference. New York: Holt, Rinehart & Winston, 1963.

Wellman, F. E. Differential prediction of high school achievement using single score and multiple factor tests of mental ability. Personnel and Guidance Journal, 1957, 36, 512-517.

Wert, J. E., Neidt, C. O. and Ahmann, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, 1954.

Wherry, R. J. Multiple bi-serial and multiple point bi-serial correlation. Psychometrika, 1947, 12, 189-195.

APPENDIX

TABLE LX

MEANS AND STANDARD DEVIATIONS OF THE PREDICTOR VARIABLES, DATA SOURCE FOR THE MULTIPLE REGRESSION AND MULTIPLE DISCRIMINANT ANALYSIS

Junior High Group Validation Sample

N = 272

No.	Predictor Variable		Mean	Standard Deviation
1	ITBS	Vocab	50.84	11.11
2	ITBS	Rdg	50.45	10.31
3	ITBS	L_1	50.06	10.37
4	ITBS	L_2	50.65	10.71
5	ITBS	L_3	50.46	10.35
6	ITBS	L_4	50.51	10.52
7	ITBS	L_t	50.12	10.30
8	ITBS	W_1	50.88	10.50
9	ITBS	W_2	50.99	10.61
10	ITBS	W_3	50.42	10.56
11	ITBS	W_t	50.54	10.45
12	ITBS	A_1	50.66	10.48
13	ITBS	A_2	51.12	10.34
14	ITBS	A_t	50.47	10.37
15	ITBS	Compos	50.50	10.39
16	Lorge	Verbal	121.32	10.86
17	Lorge	N.V.	123.67	10.54
18	Lorge	Total	122.75	9.02

ILLUSTRATION OF COMPUTATIONAL PROCEDURES FOR PREDICTING/CLASSIFYING STUDENTS INTO GROUPS USING REGRESSION EQUATIONS

1. Each Junior high English and mathematics subject area regression equation was applied to the six Level 2 predictor variables of all students within each subject area.

2. The means and standard deviations of the distributions of the estimated GPA were calculated for each subject area: (Level 2 English $\hat{Y}_2$ = 3.2, $\hat{\sigma}_1$ = 0.956; Level 2 Mathematics $\hat{Y}_2$ = 3.1, $\hat{\sigma}_2$ = 0.616).

3. The a priori proportions in Levels 1 and 3 in each subject area are: English Level 1 = .17, Level 3 = .27; Math Level 1 = .15, Level 3 = .10, based on N's of 148 and 259, respectively.

4. The standard deviations of the estimated GPA, $\hat{\sigma}$, were adjusted as follows: using Level 1 English as an illustration:

<u>English</u>
<u>Level 1</u>

.1700 = a priori proportion
<u>.1587</u> = area beyond $-1\hat{\sigma}$
.0113 = adjustment needed in $-1\hat{\sigma}$

.956 = Level 2 $\hat{\sigma}$ ∴ (.0113)×(.956) = .944 = adjusted $\hat{\sigma}_2$

3.200 = mean of $\hat{Y}$ of Level 2
<u>-.944</u> = adj. $\hat{\sigma}$
2.256 or 2.3 ∴ students with $\hat{Y}_2 < 2.3$ were assigned to Level 1

<u>Level 3</u>, adjusted $\hat{\sigma}_2$ = .84 ∴ $\hat{Y}_2$ = 3.2 + .84 = 4.0.

Thus, students with an estimated GPA, $\hat{Y}$ of 4.1 or larger were assigned to Level 3. The remainder of the students with $\hat{Y}_2$ from 2.4 to 3.9 were assigned to Level 2.

<u>Mathematics</u>: for <u>Level 1</u>, adjusted $\hat{\sigma}_2$ = .621

∴ $\hat{Y}_2$ = 3.1 - .621 = 2.5; therefore, students with $\hat{Y}_2 < 2.5$ were assigned to Level 1

for <u>Level 3</u>, adjusted $\hat{\sigma}_2$ = .819
∴ $\hat{Y}_2$ = 3.1 + .819 = 3.9; therefore, students with $\hat{Y}_2 > 3.9$ were assigned to Level 3.
The remainder of the students with $\hat{Y}_2$ of 2.6 to 3.9 were assigned to Level 2.

ILLUSTRATION OF THE PROCEDURE FOR CALCULATING THE CHANCE EXPECTANCIES OF THE NUMBERS OF STUDENTS EXPECTED TO BE PLACED CORRECTLY BY CHANCE IN THE ENGLISH AND MATHEMATICS

Junior High Check Sample

The a priori densities, English subject area:

Level 1 = 25

Level 2 = 83 Total N = 148

Level 3 = 40

The formula used:

$$f_e = \frac{a^2}{N} + \frac{b^2}{N} + \frac{c^2}{N}$$

where: f_e = chance expectation

a, b, c, = actual densities

N = a + b + c

thus:

$$f_e = \frac{(25)^2}{148} + \frac{(83)^2}{148} + \frac{(40)^2}{148}$$

f_e = 75.097 (number of hits by chance)

148-75.097 = 72.903 (number of misses by chance)

TABLE LXI

MEANS AND STANDARD DEVIATIONS OF THE SIX VARIABLES COMPRISING THE MULTIPLE DISCRIMINANT EQUATIONS FOR PREDICTING/CLASSIFYING STUDENTS INTO ACHIEVEMENT LEVELS WITHIN TWO SUBJECT AREAS

Junior High Validation Sample

Subject Area and Level	N	Statistic	Predictor/Classification Variables					
			Iowa Tests of Basic Skills					Lorge Aptitude
			L_t	W_1	W_2	A_1	A_t	Non-Verbal
Mathematics								
General Math Level 1	53	Mean	42.77	43.53	43.08	40.87	41.30	109.96
		Std.Dev.	7.03	8.13	7.61	6.3	6.03	11.35
Algebra Level 2	171	Mean	51.20	51.18	51.27	51.33	51.22	123.68
		Std.Dev.	8.56	7.84	8.11	6.69	6.87	10.54
Algebra (Honors) Level 3	39	Mean	59.74	63.13	63.92	65.62	64.64	136.18
		Std.Dev.	8.83	8.65	7.94	6.07	6.61	8.44
			L_1	L_4	L_t	W_3	A_1	Lorge Total
English								
Basic Comp. Level 1	25	Mean	45.50	44.12	43.44	42.04	43.24	111.00
		Std.Dev.	8.40	7.46	5.90	9.34	8.83	11.38
Composition Level 2	97	Mean	51.28	51.07	51.22	51.87	50.93	122.49
		Std.Dev.	8.24	8.99	5.27	7.86	7.89	8.55
Comp. (Honors) Level 3	35	Mean	59.89	61.49	62.26	62.40	63.86	137.06
		Std. Dev.	7.84	7.47	7.88	8.31	8.43	6.76

TABLE LXII

THE BASIC DISCRIMINANT EQUATIONS FOR THE MULTIPLE DISCRIMINANT ANALYSIS

Junior High Validation Sample

English

Basic Composition Level 1 N = 25	Composition Level 2 N = 97	Composition (Honors) Level 3 N = 35
L_1 + .15453	L_1 + .11743	L_1 + .11442
L_4 + .33431	L_4 + .35386	L_4 + .38921
L_t - .04383	L_t + .03705	L_t + .03705
W_3 - .55885	W_3 - .50973	W_3 - .50973
A_1 - .29771	A_1 - .29729	A_1 - .29729
Lorge T + 1.83500	Lorge T - 1.93755	Lorge T + 2.04531
constant-93.62059	-110.87670	-137.05474

Mathematics

General Math Level 1 N = 53	Algebra Level 2 N=111	Algebra (Honors) Level 3 N = 39
L_t + .14547	L_t + .18437	L_t + .18865
W_1 + .17364	W_1 + .18760	W_1 + .24287
W_2 - .04838	W_2 - .04749	W_2 + .01582
A_1 + .07559	A_1 + .19756	A_1 + .37936
A_t + .27651	A_t + .34956	A_t + .45202
Lorge NV + .86102	Lorge NV + .92628	Lorge NV + .93806
constant -60.44291	-79.61181	-104.73488

SIMPSON'S (1957) FORMULA FOR ADJUSTING THE MULTIPLE DISCRIMINANT EQUATIONS FOR PREDICTING/CLASSIFYING STUDENTS INTO GROUPS

Simpson's formula for modifying the basic discriminant equation so as to take into account the unequal sizes of the groups (i.e., the a priori proportions) upon which the basic discriminant equations were developed involves, in essence, adding a constant, K, to the discriminant equation's constant. The formula is:

$K = I\ D$

where $I = \dfrac{.62}{1 + .45\ D} \left(\mathrm{Log}_e \pi_1 - \mathrm{Log}_e \pi_2\right)$

where the tems .62, 1, and .45 are constants and

D = the intergroup distance of the discriminant analysis

and

Log_e = the natural or Naperian logarithms of the a priori proportions of groups 1 and 2 respectively; in this study these are the proportion of students in Levels 1 and 2 within the English and mathematics subject areas of the validation sample. In the following illustration, π_1 = 0.159; π_2 = 0.618.

Continued

The inter-group distances, D_{1-2}, between Levels 1 and 2 of the English and mathematics areas of the junior high validation sample are calculated first. Use is made of the junior high English validation sample as an illustration:

For convenience in following the computational procedures, the means of the variables from Table LXI and the unadjusted basic discriminant equation for the junior high English subject area are extracted from Table LXII and reproduced here:

Means of the English subject area discriminant variables

Variable	Basic Composition Level 1 N = 25 $\pi_1 = 0.159$ $\bar{X}_1$	Composition Level 2 N = 97 $\pi_2 = 0.618$ $\bar{X}_2$	Composition (Honors) Level 3 N = 35 $\pi_3 = 0.223$ $\bar{X}_3$
L_1	45.50000	51.27835	59.88571
L_4	44.12000	51.07216	61.48571
L_t	43.44000	51.21649	62.25714
W_3	42.04000	51.86598	62.40000
A_1	43.24000	50.92784	63.85714
Lorge Total	111.00000	122.49485	137.05714

Unadjusted discriminant equations of English subject area

	Basic Composition Level 1	Composition Level 2	Composition (Honors) Level 3
L	0.15453	0.11743	0.11442
L	0.33431	0.35386	0.38921
L	-0.04383	0.03705	0.14892
W	-0.55885	-0.50973	-0.54810
A	-0.29771	-0.29729	-0.18899
Lorge Total	1.835	1.93755	2.04531

Unadjusted or Basic Characteristic Score (constant)

-93.62059	-110.87670	-137.05474

Continued

I. Steps in evaluating D_{1-2}

1. $(\bar{X}_{21}d_{11}) + (\bar{X}_{22}d_{12}) + \ldots + (\bar{X}_{26}d_{16}) = C'_2$
2. $C'_2 - C_2 = C''_2$
3. $C_1 - C''_2 = \frac{1}{2}D_1^2$
4. $C'_2 - C_1 = C''_1$
5. $C_2 - C''_1 = \frac{1}{2}D_2^2$
6. $D_{1-2} = \sqrt{\frac{1}{2}D_1^2 + \frac{1}{2}D_2^2}$

where:

C_1, C_2 = respectively, the unadjusted characteristic scores for Levels I and 2 discriminant equations

$\bar{X}_{21}$ = mean of the first variable (ie. L_1) in level 2

d_{11} = discriminant coefficient of the first variable (L_1) in level 1

C_2 = unadjusted characteristic score of the level 2 discriminant equation (110.87670)

C_1 = unadjusted characteristic score of the level 1 discriminant equation (93.62059)

Thus,

1. $C'_2 = [(51.27835)\times(0.15453)] + [(51.07216)\times(0.33431)] + [(51.21649)\times(-0.04383)] + [(51.86598)\times(-0.55885)] + [(50.92784)\times(-0.29771)] + [(122.49485)\times(1.8350)] = 203.38417$
2. $203.38417 - 110.8767 = 92.50747 = C''_2$
3. $93.62059 - 92.50747 = 1.11312 = \frac{1}{2}D_1^2$
4. $203.38417 - 93.62059 = 109.76358 = C''_1$
5. $110.8767 - 109.76358 = 1.11312 = \frac{1}{2}D_2^2$
6. $D = \sqrt{1.11312 + 1.11312} = \sqrt{2.22624} = 1.49205$

The inter-group distance, D_{2-3}, was calculated in the same manner but with the Level 3 $\bar{X}$'s substituted in discriminant equation for level 2 yielding a D_{2-3} value of 2.3694. Its K value = 0.67137. When the computational procedure discribed under II and III below was applied to the Level 2 unadjusted (Basic) discriminant equation, the adjusted characteristic score, C'_2, for Level 2 was computed to be: unadjusted characteristic score $C_2 + K = C'_2$ or $-110.87670+0.67137 = -110.20533$.

Continued

Similarly, the inter-group distance, D_{1-3}, was calculated by multiplying the Level 3 $\bar{X}$'s by the Level 1 discriminant coefficients which yielded a D_{1-3} value of 4.13066. Its K value =-0.22281. Its $C_3' = C_3 + K = -137.05474 + (-.0.22281) = -137.27755$.

II. Computational procedure for obtaining K, the constant by which the characteristic score of the discriminant equation for the Level 1 group (Basic Composition) is to be adjusted:

$$K = ID \qquad I = \frac{.62}{(1 - .45)\times(1.49205)} (\text{Log}_e\ 0.159 - \text{Log}_e\ 0.618)$$

$$= \frac{.62}{1.67141} (-1.83885 - 0.48127)$$

$$= \frac{.62}{1.67441} (-1.35758)$$

$$= -0.50559$$

$$K = (-0.50359)\times(1.49205)$$

$$= -0.75137$$

III. Computational procedure for applying the constant, K, (-0.75137) to the Level 1 (Basic Composition) equation:

Level I unadjusted discriminant equation = $0.15453\ L_1 + 0.33431\ L_4 - 0.04383\ L_t - 0.55885\ W_3 - .29771\ A_1 + 1.835$ Lorge T $- 93.62059$

1. $C_1 + K = C'_1$

2. $C'_1 - C'_2$ = final adjusted constant, $C_{1\ adj}$.

where C_1 = unadjusted characteristic score, Level 1 discriminate equation

K = as defined above

C'_2 = adjusted characteristic score, Level 2 discriminant equation

Continued

Thus,

1. $-93.62059 - 0.75137 = -94.37196 = C'_1$

2. $110.20533 - 94.37196 = 15.83337$ = Final adjusted constant, C_1 adj.

Therefore, the adjustment of the Level 1 discriminant equation was accomplished by replacing its characteristic score, -93.62059, with the adjusted value, 15.83337.

An illustration of the application of the computational procedures described above to the scores of three students, one from each level in English, is presented in Table LXIII.

The summary of the results of the computation and application of the adjustment factor to the basic unadjusted multiple discriminant equation is presented in Table LXIV.

TABLE LXIII

ILLUSTRATION OF THE COMPUTATIONS INVOLVED IN THE APPLICATION OF THE UNADJUSTED AND ADJUSTED MULTIPLE DISCRIMINANT EQUATIONS TO THE TEST SCORES OF THREE STUDENTS, ONE FROM EACH LEVEL

English Subject Area

Junior High Validation Sample

Level 1 Discriminant Equations	Student #19 Test Scores	Student #100 Test Scores	Student #191 Test Scores
L_1 0.15453***	61	59	44
L_4 0.33431	47	43	47
L_t -0.04383	45	53	48
W_3 -0.55885	51	57	51
A_1 -0.29771	34	49	69
Lorge T 1.83500	117	129	130
Σ of products	199.23806	211.44237	209.91471
Unadj. Charac. Score	-93.62059	-93.62059	-93.62059
Unadj. Discriminant Score	105.63277*	117.82178	116.29412
Adjusted Charac. Score	15.83337	15.83337	15.83337
Adjusted Discr. Score	215.07143	227,27574	225.74808
Level 2 Discriminant Equations			
L_1 0.11743	61	59	44
L_4 0.35386	47	43	47
L_t 0.03705	45	53	48
W_3 -0.50973	51	57	51
A_1 -0.29729	34	49	69
Lorge T 1.93755	117	129	130
Σ of products	216.05116	230.43013	228.94900
Unadj. Charac. Score	-110.87670	-110.87670	-110.87670
Unadj. Discriminant Score	105.17446	119.55343*	118.07230
Adjusted Charac. Score	----	----	----
Adjusted Discr. Score	216.05116**	230.43013**	228.94900**
Level 3 Discriminant Equations			
L_1 0.11442	61	59	44
L_4 0.38921	47	43	47
L_t 0.14892	45	53	48
W_3 -0.54810	51	57	51
A_1 -0.18899	34	49	69
Lorge T 2.04531	117	129	130
Σ of products	236.89640	254.72235	255.37240
Unadj. Charac. Score	-137.05474	-137.05474	-137.05474
Unadj. Discriminant Score	99.70666	117.66761	118.31766*
Adjusted Charac. Score	-27.07222	-27.07222	-27.07222
Adjusted Discr. Score	209.82418	227.65013	228.30018
Level assigned by	#19	#100	#191
Unadj. Discr. Equat.	1	2	3
Adjusted Discr. Equat.	2	2	2
Level in which Student Enr'ld	1	2	3

* - highest unadjusted discriminant score-student assigned to that level
** - highest adjusted discriminant score-student assigned to that level
***- Multiply the test scores of each student, 19, 100, 191 by each coefficient and sum the products; thus, for student #19, (0.15453)×(61)-. . .-(1.83500)×(117) = 199.23806

TABLE LXIV

SUMMARY OF THE RESULTS OF THE COMPUTATION AND APPLICATION OF THE ADJUSTMENT FACTOR TO THE BASIC MULTIPLE DISCRIMINANT EQUATION

Junior High Group Validation Sample

Subject Area and Level	Adjustment Factor K	Basic Discr. Characteristic Score C	Adjusted Discr. Characteristic Score C + K = C'	Final Adj. Discr. Characteristic Score $C'_{1,2,3} - C'_2 = C_{adj.}$
English				
Basic Comp. (Level 1)	-0.75137	C_1 = -93.62059	C'_1 = -94.37196	+110.20533 - 94.37196 15.83337
Composition (Level 2)	0.67137	C_2 = -110.87670	C'_2 = -110.20533	-110.20533 +110.20533 0
Composition (Honors) (Level 3)	-0.22281	C_3 = -137.05474	C'_3 = -137.27755	-137.27755 +110.20533 - 27.07222
Mathematics				
General Math (Level 1)	-0.72663	C_1 = -60.44291	C'_1 = -61.16594	+78.55975 -61.16594 17.38381
Algebra (Level 2)	1.05206	C_2 = -79.61181	C'_2 = -78.55975	-78.55975 +78.55975 0
Algebra (Honors) (Level 3)	0.27865	C_3 = -104.73488	C'_3 = -104.45623	-104.45623 + 78.55975 -25.89648

VITA

3

Richard A. Laliberte

Candidate for the Degree of

Doctor of Education

Thesis: MULTIVARIATE STATISTICAL PREDICTION/CLASSIFICATION OF STUDENTS WITHIN INSTRUCTIONAL LEVELS IN SELECTED NINTH GRADE SUBJECTS: A COMPARISON OF THE RELATIVE EFFECTIVENESS OF THE MULTIPLE REGRESSION AND DISCRIMINANT MODELS

Major Field: Educational Psychology

Biographical:

Personal Data: Born in Minneapolis, Minnesota, December 9, 1922, the son of A. H. and Marian B. Laliberte.

Education: Attended the elementary and secondary schools of Minneapolis, Minnesota; graduated from Marshall High School in 1940; received the Bachelor of Arts degree from St. John's University, Collegeville, Minnesota, with a major in English, in June, 1948; received the Master of Arts degree from the University of Minnesota with a major in Educational Psychology in August, 1951; completed requirements for the Doctor of Education degree in May, 1969.

Professional Experience: Taught in the North Branch, Minnesota Public Schools 1948-1949; research and teaching assistant, University of Minnesota College of Education 1951-1953; Director of Student Personnel Services, University of Minnesota Laboratory School 1953-1955; Counselor and Reading Coordinator, Roseville Public Schools, St. Paul, 1955-1959; Coordinator of Guidance, Testing and Research, Hinsdale, Illinois Public Schools, 1959-1965; Research Psychologist, Oakland, California Public Schools, 1965-1967; Director of Psychological Services, Educational Research and Development Council of Northeastern Minnesota, University of Minnesota, Duluth, 1967 to present.

Professional Organizations: American Psychological Association, American Educational Research Association, American Personnel and Guidance Association, National Council on Measurement in Education, National Education Association, Phi Delta Kappa.