BESCO, Robert Orin. THE MEASUREMENT AND PREDICTION OF SUCCESS IN GRADUATE SCHOOL.

Purdue University, Ph. D., 1960
Psychology, general

University Microfilms, Inc., Ann Arbor, Michigan

THE MEASUREMENT AND PREDICTION OF SUCCESS IN GRADUATE SCHOOL

A Thesis

Submitted to the Faculty

of

Purdue University

by

Robert Orin Besco

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

June 1960

16805

Grad. School
Form No. 9
Revised

PURDUE UNIVERSITY

Graduate School

This is to certify that the thesis prepared

By Robert Orin Besco

Entitled The Measurement and Prediction of Success in Graduate School

Complies with the University regulations and that it meets the accepted standards of the Graduate School with respect to originality and quality

For the degree of:

Doctor of Philosophy

Signed by the final examining committee:

Donald C King, chairman

Joseph Tiffin

N M Downie

Paul V. Johnson

Approved by the head of school or department:

May 25 1960 E. J. Asher

To the librarian:

This thesis is not to be regarded as confidential

Donald C King
Professor in charge of the thesis

ACKNOWLEDGMENTS

I wish to express my appreciation to my major professor, Dr. Donald C. King, for the latitude and freedom which he gave to me in conducting this study. Our relationship, both academic and personal, has been very satisfying to me.

I would also like to thank the members of my committee, Dr. Joseph Tiffin, Dr. N. M. Downie, and Dr. P. V. Johnson, for the atmosphere of kindness and understanding which has been present throughout my graduate school career as well as for the excellent professional guidance I have received from them.

Special appreciation must be expressed to Dr. V. H. Denenberg for the development of the paired comparison matrix scoring procedure. His advice and insight on the scaling problem in this study was invaluable.

Thanks must also be expressed to Dean C. H. Lawshe, who first conceived of this study and obtained the approval and cooperation to conduct it in the University. The friendship and guidance of Dean Lawshe throughout my first year of graduate school will never be forgotten.

The financial assistance of the Purdue Research Foundation and the moral support and encouragement of Dean E. C.

Young of the Graduate School made this study possible.

To the Chairmen and faculty members and students of the cooperating departments of Agronomy, Chemistry, Civil Engineering, Industrial Engineering, Pharmacy, Psychology, and Sociology, I give my appreciation for the time and effort they contributed to provide the ratings and test scores employed in this study.

The debt of appreciation I owe to my wife and children, for the loss of many experiences which should have been shared by us, will never be repaid. I will never be able to completely express my love and appreciation for these sacrifices. The only meager repayment possible is for me to attempt to reassume the role of husband and father, a role which I have been sadly neglecting these past few years.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Besco, Robert Orin. Ph.D., Purdue University, June 1960. The Measurement and Prediction of Success in Graduate School. Major Professor: Donald C. King.

The purpose of this study was to investigate the relationship between various criteria of academic and research performance in graduate school and scores on the Aptitude Test of the Graduate Record Examination. 331 graduate students in the departments of Agronomy, Chemistry, Civil Engineering, Industrial Engineering, Pharmacy, Psychology, and Sociology were included in the investigation.

The criteria of success employed were grade point averages and faculty ratings. Paired comparison ratings of overall performance and graphic trait ratings on the following variables were collected:

1. Knowledge of subject matter.
2. Imagination and originality.
3. Independence in work.
4. Motivation to succeed.
5. Ability to design research.
6. Ability to conduct research.
7. Overall performance.
8. Readmission recommendation.

A new procedure for scoring paired comparison ratings was developed which involved fewer assumptions than previous scoring procedures. The modal response for each pair of ratees was used to determine the preferred ratee of the pair. Modal responses were summed to obtain the final rankings within each department.

The rating variables and grade point averages were intercorrelated and factor analyzed using the centroid method of factor analysis. Only one general factor of performance in graduate school was disclosed. Correlations were computed between test scores and overall performance ratings and the weighted factor scores to determine the validity of the test.

There were significant validity coefficients in every department except in one sub-section of the Psychology Department. The significant correlations ranged from .23 to .57.

It was concluded that graduate school performance as measured by these methods was unidimensional and that the Aptitude Test of the Graduate Record Examination would be a useful test to predict the general, overall factor of graduate school performance.

INTRODUCTION

This study was conducted to evaluate the degree to which performance of graduate students on the Graduate Record Examination Aptitude Test can be used to predict success in graduate school. Sponsored by the Purdue Research Foundation and the Graduate School of Purdue University, the primary purpose of the study was to investigate the relationship between various criteria of academic and research performance in graduate school and scores on the Aptitude Test of the Graduate Record Examination for graduate students enrolled at Purdue University.

The problem of selection of graduate school applicants has become increasingly acute in recent years, and if predictions of expansion in college and graduate school enrollments are accurate, the selection problems in the future will be next to impossible to solve using present selection methods (Parkhurst, 1955; Thompson, 1954). The problem to be solved is one of selecting from all the applicants the established number of students, in terms of departmental capacity, who will have the greatest probability of successfully completing graduate work. Most graduate departments have a limited number of students whom they can accept and are faced with selecting a relatively low percentage of the applicants. The selection ratio, i.e.

$$\frac{\text{number of students admitted}}{\text{number of applications received}},$$

varies from one graduate department to another, but the highest selection ratio of the departments involved in this study was one student accepted for every three applications received. These ratios varied all the way down to one admission out of ten applications. Statements appearing in graduate catalogs of other institutions imply that the selection ratio can be as low as one out of twenty and even lower.

It is anticipated that these selection ratios will become tighter in the future due to two factors of change in the population and one factor of permanence in our society. The first dynamic factor is the general population "baby boom" of World War II. The peak of this boom is in high school now and will soon be expanding college enrollments far above the enrollments of the past few years, especially considering the fact that the students in the past few years came from a period during the depression with a relatively low birth rate (Elmo Roper & Associates, 1960).

The second change factor is the increased emphasis on advanced education. This emphasis was increasing at a steady but rapid rate until the technological and ideological races with Russia caused increased emphasis on higher education. Public interest in increasing the general educational level of our population was heightened by the drama of the sputnik and missile races.

The static factor in our society which will also contribute to the increasing pressure on graduate admissions committees is the everpresent fact that the financial appropriations to expand educational facilities are difficult to acquire. It is analogous to Mark Twain's famous statement about the weather, only in this situation it must be rephrased to read, "Everyone talks about improving education, but no one wants to pay for it."

It seems then to be only a question of time until graduate schools are receiving many more applications that meet or exceed their present admission requirements. Since the total number of students a department will be able to admit will remain relatively constant, the selection problem will become more acute. Admissions committees will then have to search for additional admissions criteria with which to select from the qualified applicants only those with the greatest potential to succeed, both in graduate school and in their professional and scientific careers.

PURPOSE

This study was conceived to try to anticipate the projected increased pressure on the graduate admissions procedure. An attempt was made to establish more scientific and objective indices of success in graduate school and at the same time determine whether the Graduate Record Examination Aptitude Test could be used to further select applicants, all of whom were considered qualified under present admissions standards.

The primary purpose was to validate a specific selection instrument, the Graduate Record Examination Aptitude Test. Establishing the degree of relationship between the performance of graduate students and their scores on the GRE Aptitude Test would give the respective departments some indication of the usefulness of the test in selection of future students.

It was also desired to develop and refine various performance evaluation procedures to evaluate the different aspects or dimensions of graduate performance. The major purpose of this research on performance measurement was to establish techniques to assess reliably all relevant factors of a student's academic and research performance. At the same time, a conceptually and logically defensible definition of relevant performance factors was desired as well as an operational definition of performance in terms of rating methods employed in this study.

BACKGROUND

Much research has been conducted and many reports published on the types of selection and performance evaluation problems to be investigated in this study. The problem of predicting and measuring academic success is a challenging and popular issue (Walton, 1956). It is difficult to find a copy of a popular magazine or a professional educational journal that does not devote some space to the problems of selecting and/or evaluating the performance of students.

This writer was faced with a monumental task in selecting publications to be cited as source and background material in this study. The number of references used in this study is low and the publications included in the bibliography were selected on a very arbitrary basis. The literature search was carried out more to obtain a wide sampling of the literature than to select the few most outstanding publications in this area. The areas searched for pertinent publications were (1) testing instruments used to predict graduate school success, (2) selection studies, and (3) criterion studies.

Testing Instruments

To insure that the most promising students will be admitted, it is necessary to be able to predict accurately, at the time of their application, the subsequent success of students in graduate school. To do this, we must find measures of student potential which are correlated with later success in graduate school. Among the most important characteristics which determine a student's performance in graduate school are:

(1) present level of knowledge concerning his area of intended concentration,

(2) motivation to be a success both in graduate study and later life,

(3) personal adjustment to society and to his immediate academic environment, and

(4) ability and aptitude to absorb, understand, and utilize the content and techniques of his field.

Previous academic experience will give some information concerning the present level of knowledge. However, undergraduate grade point averages and personal recommendations are difficult to equate and interpret, and some investigators have found these measures to be of limited value in the prediction of success in graduate school (See Patterson, 1958). At present, behavioral scientists are attempting to measure motivation and personal adjustment, but results to date have been inconclusive and do not point to integrated principles or reliable measures in these areas. It would seem, then, that the ability and aptitude factors of graduate success would be the most promising untapped source of information concerning subsequent success in graduate study.

There have been many discussions concerning the types of aptitude tests that are most appropriate for academic selection (Hountras, 1956; Lannholm, 1957). Test developers become economically and egoistically involved with their instruments. Claims and counter claims are presented for and against speeded tests, multiple choice items, essay exams, and on ad infinitum for all possible types and variations of tests.

This writer feels that there is very little in the way of theory or principle that will enable the a priori selection of a valid test. Tests must be situationally validated and revalidated as the situation changes. The test selected in this study was selected on the bases of administrative convenience and face validity as well as the fact that it had been demonstrated to have predictive value in as many, if not more, instances than any other instrument. A bibliography of the early validations of this test is presented by Lannholm (1951).

The Aptitude Test of the Graduate Record Examination is a high level intelligence test designed specifically to measure reliable differences in mental aptitude at the extreme upper range of intellectual ability (See Schulz and Angoff, 1953; Buros, 1959; Educational Testing Service, 1959). This test is in wide use for both institutional and student evaluation (Harvey, 1957).

The main competitor of the Graduate Record Examination in the past few years has been the Miller Analogies Test (Buros, 1959). The Graduate Record Examination was employed in this study instead of the Miller Analogies partially because of administrative convenience but mostly because it is a more comprehensive test.

The Graduate Record Examination yields separate scores for verbal and quantitative abilities. These scores are standard scores based on a large normative sample and make comparisons possible between students tested in all parts of the country and in different years. Although these standard scores have had one major scale revision (Schulz and Angoff, 1953), the test publisher anticipates that the present scoring formulas will remain constant so that interpretation and comparisons of scores over long periods of time will be possible.

Although it is difficult if not impossible to obtain information on item statistics and item content of the tests, this security of items plus the standardization of the testing sessions are the main advantages of this test. Some tests have suffered from the lack of security of the items and from non-standard administration of the tests. The publishers of this test take more than the necessary and required precautions to insure that all examinees have the same testing conditions and the same lack of prior knowledge concerning the items on the examination.

Selection Studies

Two main approaches or experimental designs have been used in validating selection tests (Tiffin & McCormick, 1958, pp. 82-87; Lawshe, 1948, pp. 13-19). The first and most desirable design is to test all applicants and disregard the test scores for selection and/or performance evaluation (Manning & DuBois, 1958; Capps & DeCosta, 1957; Patterson, 1958; and Fisher, 1954). After a suitable period of time has elapsed to allow the selected applicants to demonstrate reliably their level of performance on the task involved, test scores are correlated with a measure of task performance to determine the validity of the test to predict performance "on the job." This is commonly called "follow up validity" or "predictive validity" according to American Psychological Association recommendations (APA, 1954).

The concept underlying the second procedure is called "concurrent validity." The procedure involves the administration of the test to individuals already "on the job" or enrolled in school (King and Besco, 1960; Robinson, 1958; and Hountras, 1956). Then the correlation is computed between the criterion of performance and test scores to determine the validity of the test.

The main difference in the two methods concerns the differences in test performance between an applicant and the individual already established in his position. For purposes of this investigation, the main problem is that of motivation. The applicant is almost certainly going to put forth all his effort to obtain the best possible score on the test while the admitted student may perceive the test as an unnecessary and extremely difficult task and consequently may not put forth all his efforts to obtain the highest possible test score. It can also be hypothesized that the highly motivated student may be anxious and nervous and so tense that his performance will be lowered. It has yet to be demonstrated that the motivation level of the examinee will seriously affect performance on this particular test except at the extreme ranges of motivation. Obviously, if a student is so unconcerned that he does not even attempt an item, he will lower his score, or if he is so anxious and concerned that his intellectual functions will be impaired, his performance will be lowered. It is hoped that a graduate student is proud of his performance on any task and is able to control his emotions to the extent that motivation would not be a serious question in a testing situation such as this one.

The validities obtained in academic selection studies have been low, particularly at the graduate level. Validities have typically been in the range from .20 to .40 with the higher validities being obtained by large batteries of tests which consume many hours and even days of testing time. It has been very rare to find a single test with a predictive validity over .35. These low validities have been caused by two factors: (1) the inadequacy of selection tests to measure all facets of academic potential and (2) the restriction of range in ability and performance factors for graduate students due to natural selection.

As mentioned earlier, there are many factors which determine the degree of success for graduate students. Motivation, interest, present level of knowledge, and mental aptitude are a few of the factors. To build a test or a battery of tests to measure even these few is nearly impossible. The changes in requirements from department to department would preclude the construction of tests which would have the breadth to cover all facets of graduate school potential and still have the depth to satisfy the needs in all departments.

The restricted range of ability among graduate students is caused primarily by natural selection or "survival of the fittest." Comparing the graduate student population with the general population reveals that graduate students cluster in the extreme upper ranges of any ability, motivation, or knowledge factor. By the time a student enters graduate school he has passed a number of academic performance hurdles which have eliminated the lower performing students.

Criterion Studies

One of the most crucial problems in a selection study is the development of acceptable performance criteria. Typically, studies in academic selection have used only grade point averages as criteria of performance (Capps & DeCosta, 1957; Durnall, 1954; Fisher, 1954; and Robinson, 1958). Other authors (Patterson, 1958; Webb, 1951; and Webb, 1956) have pointed out the limitations of grade point averages as a criterion, particularly in graduate school. As a criterion, the most common limitation of grade point averages is the decided restriction of range from the lowest to highest averages. Many graduate schools require a grade of B or higher to receive credit for a course. In addition to the restriction of range, grading practices and absolute grading standards vary widely between instructors and between departments. A student can theoretically "raise" his grade point average by avoiding courses from low grading instructors or avoiding minors in low grading departments.

There are many problems of performance criteria in any situation, but the wide variance in educational philosophy and goals of education make the definition of an exact criterion in graduate school extremely difficult. Thorndike (1949, pp. 119-159) describes the general problems of criterion development and Harmon et al (1954) discuss the problem of criteria in scientific work in academic, in research, and in applied settings. Brogden and Taylor (1950) provided a classification scheme for conceptualizing the various types of errors that can curtail the effectiveness of a criterion measure. They classify criterion bias or error into four types (p. 161):

1. Criterion deficiency--omission of pertinent elements from the criterion.

2. Criterion contamination--introducing extraneous elements into the criterion.

3. Criterion scale unit bias--inequality of scale units in the criterion.

4. Criterion distortion--improper weighting in combining criterion elements.

Brogden and Taylor (1950, p. 162) set up three steps to be followed when selecting or constructing criteria and criterion measures.

1. Careful analysis of the total situation in which the criterion behavior occurs for the purpose of isolating all sub-criterion variables and obtaining preliminary estimates of their relative importance--determination of what is to be measured.

2. The construction of procedures and/or scales for the measurement of these elements--determination of how each element is to be measured.

3. Development of a procedure for combining these elements into the desired single composite--determination of the relative importance of each element to over-all efficiency.

In this study the criterion problem will be approached using the three steps outlined by Brogden and Taylor. In the determination of what is to be measured, it can be hypothesized that there are two facets of performance in graduate school that contribute to overall performance. These two are: (1) academic performance and (2) research performance. It should be pointed out that adequate personal adjustment and acceptance by the faculty are also critical factors in graduate school success. In keeping with previous statements concerning the inadequacy of present measures of personal adjustment and personality, these measures of "personality" will be used only to study their reliability and their relationship with other more directly observable performance variables.

The two most popular measurement techniques which have been used to collect judgments or evaluations of graduate students are the graphic rating scale technique and the rank order technique. The graphic rating scale typically requires the rater to check the point along a continuum which represents the rater's judgment of the ratee. As Guilford (1954) points out, these scales are simple to construct and score and are very acceptable to raters. However, they suffer from some common rater errors such as the errors of halo, leniency, and central tendency. Although statistical correction can be made for these errors, it is impossible to tell whether the obtained distribution of ratings suffers from these errors or whether the ratings are truly representative of the ratee's performance.

The ranking methods overcome some of these three errors but are time consuming when the number of ratees is over twenty. Also they are frustrating to the rater and actually statistically misleading when the ratees are very uniform in their performance. The method of paired comparisons is a technique for making the judgment task simpler for the rater. This technique will not artificially spread the scores assigned to ratees and since the judgments are made in pairs, the judgment task is reduced to a number of "better-than-or-worse-than" comparisons. A further advantage of paired comparisons is that Kendall (1955) has developed a coefficient to determine the consistency with which the rater has made his judgments based on the number of circular triads in the rater's judgments.

Another problem in criterion development is encountered when the various criterion measures are combined to obtain one composite score for each individual. Ghiselli (1956) and Ryans (1954) have discussed the problems of combining subcriteria into a composite criterion. Ghiselli implies that, in a multidimensional criterion, high performance on one factor should compensate for low performance on another. The logic of Grimsley (1949) and Dvorak (1956) on the combining of predictor scores can also be applied to the combining of sub-criteria into a composite criterion. Grimsley and Dvorak feel that combining measures by statistically determined weights is not feasible in all cases. For example, it may not be possible for a graduate student to be successful if he is failing in every course he is taking, regardless of the brilliance of his research performances. Grimsley and Dvorak view separate scores more as establishing successive hurdles or minimum acceptable standards on each variable. This technique has been labeled the multiple cut-off method.

Guilford and Michael (1948) and Ryans (1954) employed a technique for combining measures on the basis of their loadings derived from factor analysis. This method might provide a compromise between the statistical weighting and multiple cut-off methods, in that the subfactor scores would be combined according to weights but the main factors could remain relatively independent.

However, Ryans found that simple unitary weightings were almost as reliable statistically as the more complex weighting procedures. Ryans concluded that complex weighting schemes may not be worth the extra effort involved to obtain them.

PROCEDURE

The procedures used in this study follow the basic outline of the design described earlier as the "concurrent validity" design. Most of the students were tested after they had been admitted to graduate school.

This study consists of a number of smaller studies carried out separately in various departments throughout Purdue University. Departments participated in this study on a voluntary basis although the study had the complete backing of the University and the Graduate School. Because of the autonomy of the study within each department, there were some minor variations in procedure between the departments. Most of these were due to problems of scheduling and established departmental policies and did not affect the basic design or hypotheses to be tested.

Sample

The names of the departments participating and the number of students involved in each department appear in Table 1. As can be seen, the size of the sample in each department varied from sixteen in Industrial Engineering to 82 in Chemistry. The sampling fraction, i.e.

$$\frac{\text{the number of students in sample}}{\text{the total number of students in department}},$$

Table 1

Departments and Number of Students Included in This Study

Department	Number of Students
Agronomy	42
Chemistry	82
Civil Engineering	26
Industrial Engineering	16
Pharmacy	44
Psychology	
Clinical	40
Experimental	20
Industrial	37
Sociology	24
Total	331

varied from a low of 52% in Chemistry to a high of 100% in Industrial Psychology. The total number of students in the study was 331.

All of the students in all samples had completed at least 2½ semesters of graduate study at the time they were rated by staff members. The grade point averages were based on at least three semesters of graduate study and at least 27 semester hours of graduate course work.

Students from non-English speaking countries were excluded from the study. The test performance of students from these countries was significantly lower than the test performance of English speaking natives, especially on the verbal section of the test. The norms for Purdue students whose native tongue is other than English appear in Table 21 in Appendix A. A comparison of these norms with the Purdue-wide norms in Table 22 of Appendix A, based on only English speaking students, points out the markedly higher scores of the English speaking natives. It is felt that this score difference is largely due to the language handicap of foreign students and that these test scores are not representative of the mental capacities of foreign students. For this reason it was decided before criterion data had been collected to eliminate foreign students from this study.

The elimination of foreign students accounted for a large proportion of the losses from 100% samples in the departments. Table 2 contains a breakdown of the reasons students were dropped from the study before any data analysis had taken place. The three reasons for dropping students were as follows: (1) Incomplete criterion data. In all cases this was caused by a lack of knowledge of some or all phases of a student's performance by anyone in the department. (2) Lack of test scores. Since the students were not required, but only positively requested, to take the test, a few students did not cooperate. (3) The language handicap. As mentioned above, students whose native language was not English were eliminated due to the severe language handicap, resulting in poorer test performance.

Test Administration

The test scores were collected under conditions which varied between departments. These testing conditions and procedures for each department appear below.

Psychology

Students in the Psychology Departments were required to submit test scores with their applications. Because of this requirement, only five students in the Psychology Departments took the tests under an administration for this study. These five had all been admitted to graduate school prior to September 1958 and they took the test in September 1959. The fifteen students who did not take the test were doing research in absentia. Industrial Psychology had completed

Table 2

Reasons for Dropping Students from Study

Department	Total Number of Students Enrolled	Total Number Dropped	Reason for Drop: Foreign Student	Reason for Drop: Incomplete Criterion	Reason for Drop: No Test Scores
Agronomy	62	20	11	9	0
Chemistry	156	74	20	42	12
Civil Engineering	45	19	16	1	2
Industrial Engineering	26	10	7	0	3
Pharmacy	54	10	2	6	2
Psychology					
Clinical	52	12	1	1	10
Experimental	29	9	1	3	5
Industrial	37	0	0	0	0
Sociology	43	19	2	7	10
Totals	504	173	59	69	44

a previous validation of test scores and the test scores were used in student selection. Although scores were available for most of the other Psychology students at the time of application, no definite score limits or weights had been established in the Clinical and Experimental sections of the Psychology Department.

Chemistry

The tests were administered to entering Chemistry students for the entering classes of 1957 and 1958. The tests were taken as part of a three day battery of tests during orientation week. Scores for the entering class of 1957 had been distributed to the Chemistry staff before this study was initiated, but the scores of the entering class of 1958 were withheld from the Chemistry staff until criterion data were collected.

Sociology

The Sociology Department urged that all their applicants submit scores as part of the application procedure, but they had not established minimum scores or weights to evaluate an applicant's test performance. Since this test was not positively required of all applicants to the Sociology Department, 10 of the 23 students in this study were tested in May of 1959. Scores of the students tested were withheld until criterion data had been collected.

Agronomy, Industrial Engineering, Pharmacy, and Civil Engineering

Students in the Agronomy, Pharmacy, and Industrial Engineering Departments were tested in May of 1959 and the students in the Civil Engineering Department were all tested in September of 1959. The test scores were withheld from all four of these departments until criterion data had been collected.

All of the testing sessions followed the very rigid and specific instructions from the Educational Testing Service. During the reading of the instructions to the examinees, they were told that these scores would not be used to evaluate their performance in graduate school. However, they were told that a copy of the test scores would be sent to their department heads and to their major professors. They were also told that transcripts of these scores would be available from the Educational Testing Service for submission to other graduate schools or various Fellowship programs which require test scores as part of the application procedure.

Criterion Development

Graduate student performance was perceived as being very uniform and very high in all departments. Seldom was any student's performance acknowledged to be poor. The measurement problem consisted of developing rating methods, scales, and scoring procedures that would yield a reliable scale to provide at least a rank order of student performance. Although it was desired to obtain an interval scale of measurement, the requirements for this study would have been satisfied if a reliable rank order of students within each department could be obtained.

Performance Measurement

The two main rating methods used were the graphic rating scale and the paired comparison technique. The graphic rating scale appears in Appendix B. The traits were selected to try to measure the two performance factors hypothesized earlier, academic performance and research performance. The knowledge trait ratings and official university grade point averages were to be the central measures of the academic factor. The two traits concerning research performance were designed to load heavily on the research factor.

The scales concerned with imagination, motivation, and independence were included to study the reliability of ratings on these traits and to study the correlation between ratings on these traits with other ratings and with grade point averages.

The overall performance scale was included to study the "halo effects" in the ratings and also to provide a base from which to evaluate the degree of independence of the other performance ratings.

The readmissions scale was designed to provide the raters with a rating scale which was more concrete and tangible than some of the more nebulous traits described in the above paragraphs. It was felt that the readmission criterion was easily understandable and that staff members already conceptualized student performance in terms of readmission.

In addition to the graphic trait rating scales and grade point averages, paired comparison ratings on an overall performance trait were used. Each pair was presented on an individual IBM card as illustrated in Appendix B. This was to provide another overall performance "anchor" to evaluate the other scales and to provide one overall criterion of performance in the event that the graphic trait ratings would not be adequate performance measures.

The general hypothesis in the measurement phase of the study was that performance in graduate school is due primarily to two factors, academic performance and research performance.

The specific hypotheses are listed below:

1. Two factors, academic and research, will emerge and account for nearly all the variance in these performance measures.

2. The overall performance measures will be heavily loaded on the academic factor in the social science departments.

3. The overall performance measures will be heavily loaded on the research factors in the physical and natural science departments.

Criterion Collection

After the cooperation of the individual departments had been obtained and the department head had notified his staff members of the study, an introductory dittoed form letter was sent to each faculty member in each department. This letter also contained a roster of the graduate students to be included in the study. The professors were asked to check the names of every student whom he could evaluate in terms of performance in graduate school, in class work, and in research. An example of this letter appears in Appendix B.

After these rosters were received from the faculty, rating forms and paired comparison IBM cards were prepared for each professor. This rating packet contained forms for only the students whom the professor had indicated he could evaluate. The rating forms and a form letter containing instructions were then sent to each professor.

All correspondence was done through the offices of the respective department heads. The letters and forms were distributed and collected by the department heads. This was done primarily to facilitate the distribution and collection of materials, but also to insure that each faculty member would be aware that his department head was sponsoring the study and was interested in the results obtained. Of the 132 professors who were contacted to participate in evaluating students, 127 of them cooperated completely. The other five either failed to respond to the initial inquiry or failed to return the rating forms.

As a control measure, the number of months a student had been enrolled in graduate school was obtained. If there was a significant relationship between the various criterion measures and length of time in graduate study, this fact would indicate that long term graduate students had been able to survive in the graduate program due to better performance. If the relationship were significant, length of time in graduate school would be partialled out of the other criterion measures by employing partial correlation techniques.

Criterion Analysis

Scoring. The trait ratings were scored on an a priori basis by assigning a value of one to the lowest rating on the scale, two to the next, on up to six for the best possible rated performance. The readmission rating was scored in the same manner except the values assigned ranged from one to five for the possible readmission decisions. For all the ratings, a high rating score was indicative of high performance. The arithmetic means of the ratings received by each student were used as the performance measures for each student on the respective performance trait.

Grade point averages were collected from the files of the University Registrar. These averages were based on at least three semesters of graduate work and at least twenty-seven semester hours of course credit. This is three semester hours of course credit above the minimum course requirements for the Master of Science Degree at Purdue. However, not all of the students in the study had been granted the M.S. degree at the time the criteria were collected.

The paired comparison ratings of overall performance were scored by two procedures. The first was the Personnel Comparison System described by Lawshe et al (1949). In this system, a standard or "Z" score is assigned each student based on the proportion of preferences from each rater. These standard scores are averaged to obtain one composite "Z" score for each student. This score will be hereafter referred to as the "Paired Comparison Z Score."

The above procedure involves the assumption of equal means and variances of student performance within each rater. Due to the wide variations in the number of students rated by individual professors, it was felt that these assumptions might be violated in this study.

To avoid the above assumptions, an a priori decision was made to develop another scaling procedure to score judgments of the raters in this study. The situation was one in which a number of students were to be ranked on a common scale, but where all raters were not able to evaluate all the students. An approximate conception of the number of missing cells and disproportionality of the cells in the judgment matrix can be obtained from inspection of Table 3. It can be seen that most students were rated by only two or three faculty members.

Given an incomplete matrix such as this, the problem is to find some method of combining all the judgments of the different raters so that all students in a department can be placed on a common scale. If there were only a few missing cells, then the procedure of fitting missing values (Cochran & Cox, 157) is feasible. However, this procedure is not applicable when there are a large number of empty cells, as in this study.

In the procedure used in this study the judgments of the raters were in the form of paired comparison ratings. The modal judgment of all raters who evaluated a particular ratee was entered into a "ratee by ratee" matrix. This matrix was then used to determine the rank order among the ratees as well as the internal consistency of the judgments.

This method assumes that the failure of a particular rater to know a ratee well enough to evaluate him was due to random factors. It also assumes that the raters were from the same population, so that their judgments could be averaged. No assumptions were made concerning the form of the rater or ratee populations. The procedure was as follows:

1. An "n x n" matrix of the ratees was laid out as in

Table 3

Number of Students Rated by Indicated Number of Faculty Members

Department	Total Number of Raters in Department	Number of Faculty Ratings per Student						
		7 or more	6	5	4	3	2	1
Agronomy	25	0	4	2	7	11	13	5
Chemistry	26	0	0	0	1	4	14	63
Civil Engineering	25	0	3	4	2	9	6	3
Industrial Engineering	10	0	0	3	1	7	5	0
Pharmacy	14	0	0	1	0	12	21	10
Psychology								
Clinical	20*	9	7	9	8	5	2	0
Experimental	20*	4	4	2	4	4	2	0
Industrial	20*	9	3	5	2	9	9	0
Sociology	8	0	0	0	6	9	7	2

*The same 20 faculty members rated students in the three sections of the Psychology Department.

Table 4. Table 4 is an example, for purposes of illustration, from another study (Denenberg & Besco, 1960).

2. Taking the paired comparison judgments of each rater separately, a plus (+) was entered in the appropriate cell each time the ratee on the horizontal axis was preferred over the ratee on the vertical axis. A minus (-) was entered whenever the ratee on the horizontal axis was less preferred than the ratee on the vertical axis. This was done both above and below the diagonal. In Table 4 it can be seen that eight staff members compared students 1 and 2, with five of them judging 2 superior to 1.

3. After the judgments of all raters have been entered, a plus value was assigned to each cell that had the greater proportion of plus signs and a minus value to each cell that had a greater proportion of minus signs. If there was an equal number of plus and minus signs in a cell, or if the cell was blank, that cell was assigned a zero value. An illustration of this step is in Table 5; the ratees have also been arranged in rank order in this table.

4. The number of positive cells were summed for each ratee. If two ratees had the same number of positive cells, the tie was broken by observing the proportion in the cell in which these two ratees were compared; the student with the greater proportion of favorable choices in that cell received the more favorable rank (See student pairs 6-11 and 3-8 in Tables 4 and 5). If that cell was equally split or

Table 4

Example:

Matrix of Judgments Made by 15 Raters on 11 Ratees Showing Number of Times Ratee on Horizontal Axis Is Preferred over Ratee on Vertical Axis (+) and Number of Times Ratee on Horizontal Axis Is Less Preferred Than Ratee on Vertical Axis (-)

	Ratees																					
	1		2		3		4		5		6		7		8		9		10		11	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
1			5	3	1	0	3	0	2	0	3	1	1	6	2	1	2	0	1	2	2	0
2	3	5			1	2	6	0	4	0	4	4	1	8	1	3	2	3	1	6	1	4
3	0	1	2	1			3	0	4	0	1	2	0	1	2	1	1	3	0	2	0	2
4	0	3	0	6	0	3			3	1	0	5	0	3	0	4	0	5	0	4	0	2
5	0	2	0	4	0	4	1	3			0	5	0	2	1	4	1	5	0	3	0	2
6	1	3	4	4	1	2	5	0	5	0			0	6	4	2	2	4	0	7	2	3
7	6	1	8	1	1	0	3	0	2	0	6	0			3	0	2	0	4	1	3	1
8	1	2	3	1	1	2	4	0	4	1	2	4	0	3			2	3	0	3	2	0
9	0	2	3	2	3	1	5	0	5	1	4	2	0	2	3	2			1	3	2	0
10	2	1	6	1	2	0	4	0	3	0	7	0	1	4	3	0	3	1			4	1
11	0	2	4	1	2	0	2	0	2	0	3	2	1	3	0	2	0	2	1	4		

Table 5

Matrix of Pooled Judgments Arranged in Rank Order

	Ratees										
	5	4	2	8	3	6	11	9	1	10	7
5		-	-	-	-	-	-	-	-	-	-
4	+		-	-	-	-	-	-	-	-	-
2	+	+		-	-	0	-	-	-	-	-
8	+	+	+		-	-	+	-	-	-	-
3	+	+	+	+		-	-	-	-	-	-
6	+	+	0	+	+		-	-	-	-	-
11	+	+	+	-	+	+		-	-	-	-
9	+	+	+	+	+	+	+		-	-	-
1	+	+	+	+	+	+	+	+		-	-
10	+	+	+	+	+	+	+	+	+		-
7	+	+	+	+	+	+	+	+	+	+	
Number of Preferences	10	9	7	6	6	5	5	3	2	1	0
Rank	1	2	3	4	5	6	7	8	9	10	11

blank, then the ratees were treated as being tied.

Reliability. In the paired comparison matrix score, Kendall's coefficient of consistence was used as a measure of internal consistency among the raters (Kendall, 1955; Edwards, 1957). The coefficient is reduced as the number of circular triads increases. Whenever ties occurred in the data, they were counted as one half in determining the coefficient of consistence and consequently reduced the value of the coefficient.

The reliabilities of the graphic trait ratings and the paired comparison "Z" score were estimated using an analysis of variance approach suggested by Ebel (1951). The coefficient used was the intraclass correlation which treats the differences among the mean ratings of the faculty members as contributing to error variance and consequently reducing the reliability.

Factor Analysis of Performance Ratings. The eight graphic trait rating scores and the two paired comparison rating scores and grade point averages were factor analyzed using the complete centroid method as suggested by Thurstone (1947) and by Fruchter (1954). Communalities were estimated by the multiple correlation squared between the variable and the other ten variables. Separate factor analyses of the eleven performance variables were computed in each of the nine departments.

The rating variables used were designed to provide factor loadings on academic, research, and overall performance. As Guilford (1952) suggested, more than one performance variable was used for each of the hypothesized performance factors. Thus the hypothesized factors should emerge with loadings on more than one variable. The hypothesized factors would then be common factors and would be more easily interpreted than if the factor was specific to one variable.

Compiled Criterion Scores. A composite score was obtained for each student on each factor of performance. The average rating for individual students on each variable was weighted by the factor loading for that variable in the respective departments. These weighted rating scores were summed to give each student one performance score for that factor.

Validity Determination

Validity of the two tests was determined by correlating the test scores with the various performance measures.

Departmental Validities

For each department, rank order correlation coefficients were computed between the two test scores and the paired comparison matrix score. Pearson product-moment correlations were computed in each department between the weighted factor criterion scores and the two test scores.

Validities were computed only within departments. No attempt was made to obtain a general expression of validity across all departments. This was not attempted primarily because of the wide differences in test scores between departments (King & Besco, 1960) and the wide differences in grade point averages and ratings between departments. To equate test scores, grades, and ratings across departments was a problem which would involve assumptions concerning absolute performance levels between departments. This author had no basis on which to make these assumptions.

Departmental Intercorrelations

Pearsonian correlations were computed between the fourteen variables in the study for each department. These fourteen variables were: the ten rating variables, grade point averages, months in graduate school, plus the two test scores.

RESULTS

Reliability of Ratings

The reliability of the ten rating variables for each department appears in Table 6. It can be seen that most of the reliabilities on the graphic trait ratings were very low. The highest reliabilities were found with measures of overall performance, especially when measured by one of the paired comparison techniques.

The reliabilities of the paired comparison matrix scored rating ranged from a low of .63 in Chemistry to a high of .92 in Sociology. This rating technique had consistently higher reliabilities across departments than any of the other rating procedures.

Factor Analysis of Ratings

The original hypotheses concerning the factor structure of graduate school performance were rejected by these data. In every department the only factor of performance to emerge was a general performance factor. The factor loadings on the one general factor for all the performance variables within each department appear in Table 7. All the variables were highly loaded on this general performance factor. Grade point averages were consistently the variable with the lowest factor loading.

Table 6

Interrater Reliability of Ratings[a]

Variable	Department								
						Psychology			
	Agronomy	Chemistry	Civil Engineering	Industrial Engineering	Pharmacy	Clinical	Experimental	Industrial	Sociology
1. Knowledge	28	77	44	12	46	50	58	46	73
2. Imagination	31	51	33	15	49	42	56	49	71
3. Independence	35	61	19	14	57	39	45	50	63
4. Motivation	36	29	14	04	57	34	53	38	56
5. Design Research	29	61	29	02	57	35	51	49	70
6. Conduct Research	28	72	29	02	51	38	44	49	70
7. Overall Performance	36	67	46	17	58	44	57	54	74
8. Readmission	32	63	53	55	65	37	57	55	78
9. Paired Comparison Z Score	50	44	27	24	50	45	61	59	77
10. Paired Comparison Matrix Score[b]	72	63	72	76	73	66	82	78	92

[a]All decimals omitted.

[b]Based on Kendall's coefficient of consistence.

Table 7

Factor Loadings in Each Department*

Variable	Department								
						Psychology			
	Agronomy	Chemistry	Civil Engineering	Industrial Engineering	Pharmacy	Clinical	Experimental	Industrial	Sociology
1. Knowledge	885	893	946	887	887	961	926	958	965
2. Imagination	905	851	968	899	893	876	856	941	972
3. Independence	941	895	944	861	905	946	959	956	964
4. Motivation	916	741	895	896	928	888	894	866	877
5. Design Research	936	903	934	853	920	959	970	967	981
6. Conduct Research	931	906	901	930	896	946	971	975	965
7. Overall Performance	981	953	969	962	969	968	990	977	983
8. Readmission	637	695	844	937	678	805	806	877	861
9. Paired Comparison Z Score	833	654	759	880	858	860	942	897	894
10. Paired Comparison Matrix Score	929	760	775	813	867	952	941	917	765
11. Grade Point Averages	455	728	649	837	570	600	713	739	851

*All decimals omitted.

The intercorrelations between all the variables appear in Tables 21 through 29 in Appendix C. The intercorrelations in this table are all Pearson product-moment correlation coefficients, and they appear below the main diagonals of the matrices. The communalities inserted for the factor analyses are on the main diagonal. The residuals remaining after the extraction of the first centroid factor are above the main diagonal. Inspection of these residuals reveals that almost all of them are below an absolute value of .10. Considering the fact that the variables were so highly intercorrelated, these small residuals indicate that almost all of the variance is accounted for by the first centroid factor.

Further evidence of the fact that there is only one common factor in these data appears in Table 8. It can be seen that the first factor accounts for roughly 90% of the common variance in all departments. The second factor was extracted for each department and the greatest amount of common variance accounted for was 1.4% in Experimental Psychology. The second factor was not statistically or practically significant in any of the departments. Since each centroid factor is smaller than the factor preceding it, no factors were extracted after the second factor.

The correlations between months in graduate school and the other performance variables were close to zero in most cases. Since these low correlations indicated that the

Table 8

Table of Variance Accounted for by Common Variance and by the First and Second Centroid Factors

	Per Cents			
Departments	Common Variance As A % of Total Variance	First Factor Variance As A % of Total Variance	First Factor Variance As A % of Common Variance	Second Factor Variance As A % of Common Variance
Agronomy	82.3	74.6	90.6	1.2
Chemistry	76.3	67.6	88.6	1.1
Civil Engineering	89.6	76.9	86.8	.5
Industrial Engineering	94.0	78.8	83.9	.6
Pharmacy	83.8	73.9	88.1	.4
Psychology				
Clinical	84.9	79.8	94.0	.8
Experimental	94.7	82.8	87.4	1.4
Industrial	91.5	84.3	92.0	.4
Sociology	91.3	84.4	92.5	.1

other performance variables were not highly related to tenure in graduate school, the number of months in graduate school was not included in the factor analysis.

Departmental Validity

Due to the unidimensional qualities of the graduate school performance measured in this study, it was felt that the criterion used should be an overall performance measure with high reliability. The paired comparison matrix score was selected as the one best criterion measure.

In the first two columns of Table 9 can be seen the rank order correlations between this paired comparison overall performance criterion and the test scores. One or both of the sections of the Aptitude Test had significant correlations with the overall criterion in all departments except Industrial Psychology. The lack of correlation in Industrial Psychology was due to the fact that students in this department were selected in part on the basis of these test scores. There was a decided restriction of range in the test scores. The standard deviation of test scores in Industrial Psychology was 70 and 60 for the Quantitative and Verbal sections, respectively. The standard deviations in the other departments ranged from 90 to 120 score points.

The significant validities ranged from .20 in Chemistry to a very high .56 in Sociology. In three departments, Chemistry, Experimental Psychology, and Sociology, both Quantitative and Verbal scores were significantly related to performance.

Table 9

Departmental Validities[a]

Criterion Variables	Paired Comparison Matrix Score#		Overall Performance Graphic Rating		Grade Point Average		Compiled Criterion Factor Score		
Test Score / Department	Quantitative	Verbal	Quantitative	Verbal	Quantitative	Verbal	Quantitative	Verbal	N
Agronomy	30*	05	28	-05	31*	-07	34*	05	42
Chemistry	27**	23*	26*	17	35**	21*	21*	20*	82
Civil Engineering	03	47**	23	21	34	13	08	33*	26
Industrial Engineering	-01	51*	30	14	-08	11	-17	29	16
Pharmacy	38**	14	41**	05	31*	27	36*	10	44
Psychology									
Clinical	00	32*	-03	36*	03	02	-03	24	40
Experimental	57**	47*	44*	37	17	03	47*	43*	20
Industrial	-09	09	-24	14	02	11	-21	16	37
Sociology	56**	37*	60**	51**	52**	48*	65**	31	24

#Rank order correlation between test scores and paired comparison matrix score, all other correlations in this table are Pearsonian.

*Significant at the .05 level.

**Significant at the .01 level.

[a]All decimals omitted.

Correlations between test scores and the overall performance graphic trait ratings, the grade point averages, and the compiled-criterion factor score are also in Table 9. Most of these validities are lower than the validities with the more reliable paired comparison rating. Even the compiled-criterion factor scores had lower average validities than the simple paired comparison criterion.

Multiple Correlations

The multiple correlations between graduate performance and both of the test scores were computed for the three departments in which both tests demonstrated significant validities. The multiple correlations were .32, .64, and .58 in Chemistry, Experimental Psychology, and Sociology respectively. None of these multiple correlations were significantly greater than the simple two-variable correlations.

DISCUSSION

Reliability of Ratings

The very low reliability of the trait ratings in this study was probably due to two factors. The first was that faculty members were observing the students in different contexts and phases of their graduate work. It would be logical that the performance of students would vary in different contexts. The second reason was the fact that differences between absolute rating level of the faculty members was treated as error and consequently lowered the reliabilities.

Some raters consistently assigned ratings higher or lower than the other raters. This fact was probably due to the response set of these "high" or "low" raters. This phenomenon is commonly described as some raters being "hard" or "tough," while other raters are "soft" or "easy." These differences in rater leniency were not statistically corrected in this study. It was desired to develop a rating procedure that would be immune to the effects of rater leniency tendencies. Even though it was quite obvious that the graphic trait ratings did not accomplish this objective, it is still almost a necessity that the rating procedure to be used in the graduate departments should eliminate this deficiency by other techniques than statistically equating scores. To equate scores between raters statistically, the assumption must be made that the "true" scores of the ratees being rated is known and that the level of ratings received is altered up or down to fit the "true" score level.

Quite obviously, it is impossible to determine if a rater is really a hard or soft rater. It could be that the students a rater is evaluating are all low or high performers. This undetermination of true ratings brings up unsoluble definitional and logical problems concerning true level of performance.

It is hoped that the paired comparison matrix score technique developed in this study will eliminate the problem concerned with differences between true scores of ratees and absolute performance standards of raters. The judgments made by raters are merely decisions as to which student of a pair is better in terms of the rated variable. The judgments of all the raters are pooled to determine the modal judgment concerning which ratee is better on a variable. The highest ranked ratee is the one who is preferred, on this modal judgment basis, over all the other ratees. The second highest rank is assigned to the ratee preferred over all others except the highest ratee and so on down to the last man who is not preferred over any of the other ratees.

Even though Kendall's coefficient of consistence provides an internal reliability estimate, it is felt that more work needs to be done with this technique to develop a statistic which will reflect the degree of agreement within cells as well as between cells. The technique as used assigns the same weight toward reliability for a cell with a six to four split as the weight assigned a cell in which all ten judgments are in agreement. With the development of this statistic, the paired comparison matrix score would be highly effective as a rating tool. The assumptions commonly made concerning the shape and form of the "true ratings" for the entire ratee populations as well as the shape and form of the "true ratings" within individual raters would no longer have to be made.

This technique could also be used to advantage in the peer rating context. Kubany (1957) has pointed to the danger in the peer rating situation of establishing absolute performance levels by peer nominations. Since the pair comparison technique is concerned with relative performance, this objection to peer ratings could be eliminated.

Buckner (1959) has raised an interesting point concerning reliability of ratings that is pertinent to this study. Buckner feels that high interrater agreement does not necessarily indicate "good" or "true" ratings. If raters observe the ratees in different situations or classes or phases of graduate work, as in this study, it would be expected that the performance would vary in the different contexts. If the ratings were accurate they would reflect these differences in performance and consequently the inter-rater agreement would be low. High inter-rater agreement in this type of situation might indicate that only the "halo effect" or at best a general performance factor was being evaluated.

Even though Buckner's rating context was similar to the context in this study, the low rater agreements in this study were probably due to rater "set" differences more than to the fact that different facets of performance was being evaluated. Especially since the factor analyses revealed only one large, general factor of performance, the hypothesis that the low interrater agreements were due to ratings of different facets of a student's performance must be rejected.

Factor Analysis

The correlations, on which the factor analyses were based, were between the average ratings of a student based on all the ratings that student received on each variable. Averaging ratings, in which there was little agreement between raters, could have had the effect of eliminating any variance in the ratings due to factors other than the general factor.

The number of students available in most of the departments was smaller than the recommended number of observations on which to base a factor analysis (Thurstone, 1947; Fruchter, 1954). The factor analyses carried out in this study, on the rating scores assigned to a student, were based on the average of several ratings or observations. Therefore, the factor analysis was actually based on more observations than the number of students in the department. However, since the primary purpose of the factor analysis was exploratory and the factors were to be used more as a means of gathering insight than to establish a theoretical framework, the factor analyses were conducted with the small number of cases.

The very high intercorrelations between all the rated performance variables suggested a strong "halo" effect between the traits measures in this study. The raters could not distinguish between the traits selected using the rating methods provided. This could be due to the fact that only one general performance factor exists in graduate school performance. This study did not prove that graduate school performance is in fact unidimensional. The results did clearly indicate that these raters using the rating techniques employed to evaluate these traits of graduate performance could not identify any factors except the general overall performance factor.

Putting the traits to be rated on separate rating forms and administering these forms on separate occasions could possibly eliminate some of the "halo" between ratings on these traits. More detailed instruction to the raters and some method of increasing rater motivation and increasing understanding of the rating purposes might also contribute to reduce the intercorrelations among the rated variables.

Validity

The validities obtained in this study were very high for validities in the graduate school context. Considering the crudeness of the criterion and the fact that most of the students took the test under "unmotivated" conditions, it is very probable that the validities are lower bound estimates of the "true" relationship between these variables.

An interesting fact to note in Table 9 is that in the two departments with the heaviest mathematical and numerical emphasis, Civil Engineering and Industrial Engineering, the quantitative scores had no relationship with performance. However, there was a definite, positive correlation between verbal scores and performance. This lack of relationship between quantitative scores and performance was probably caused by the restriction of range in quantitative score due to natural selection. The emphasis that these curricula place on mathematics courses has caused the low numerical ability students to be eliminated. The students who survive the undergraduate mathematics and engineering curricula and are admitted to graduate school are all highly qualified to handle quantitative and numerical problems.

The low relationship between quantitative ability and performance does not indicate that quantitative ability is irrelevant to success in these engineering departments. The students in these two departments had the necessary level of quantitative ability to survive in these curricula and this level of ability is so high that it precluded the possibility of obtaining a significant correlation between this narrow, upper range of test scores and the performance measures.

These two departments should establish minimum scores on the quantitative section of the test which would be necessary for a student to be considered for admission. This level of test score should be around the present 10th to 20th percentile for quantitative scores in the respective departments. Quantitative scores above this minimum level would not be effective in selecting the best possible students. Verbal scores or other admissions criteria would have to be used for this purpose. Quantitative scores are a good rejection device in these two departments but will not be of much help in making positive selection decisions.

The same phenomenon of selection caused the restricted range in test scores in Industrial Psychology. The restricted range here was not caused as much by natural selection as it was by the fact that test scores were used to make admission decisions. Students scoring low on the test were never admitted to graduate study.

Industrial Psychology had carried out a previous validity study and they were well satisfied with the contribution that the test was making to the selection procedure. The low validities indicated that, within the range of scores for this department, increased predictive efficiency for the test could not be realized.

Tables 10 through 20 are Expectancy Charts which demonstrate the probability that an applying student, with a given test score, will exceed the present median performance level in the indicated department. For example, in Table 10 for Agronomy, a student scoring above 640 on the Quantitative section has a probability of .67 of exceeding the present median performance level in Agronomy. A student with a test score between 480-530 has a probability of .43 of exceeding the median performance level.

Each test score range in these tables was the range associated with 20% of the students in that department. For Agronomy, in Table 10, 20% of the students scored below 480 on the Quantitative section of the test, 20% scored between 480-530, and so on up to the top 20% who scored above 640.

The test score distributions on which these charts were based were theoretical distributions based on the mean and standard deviation of the scores obtained from the students in that department. This procedure suggested by Lawshe et al (1958) eliminates some of the minor incon-

Table 10

Expectancy Chart: Agronomy

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Agronomy

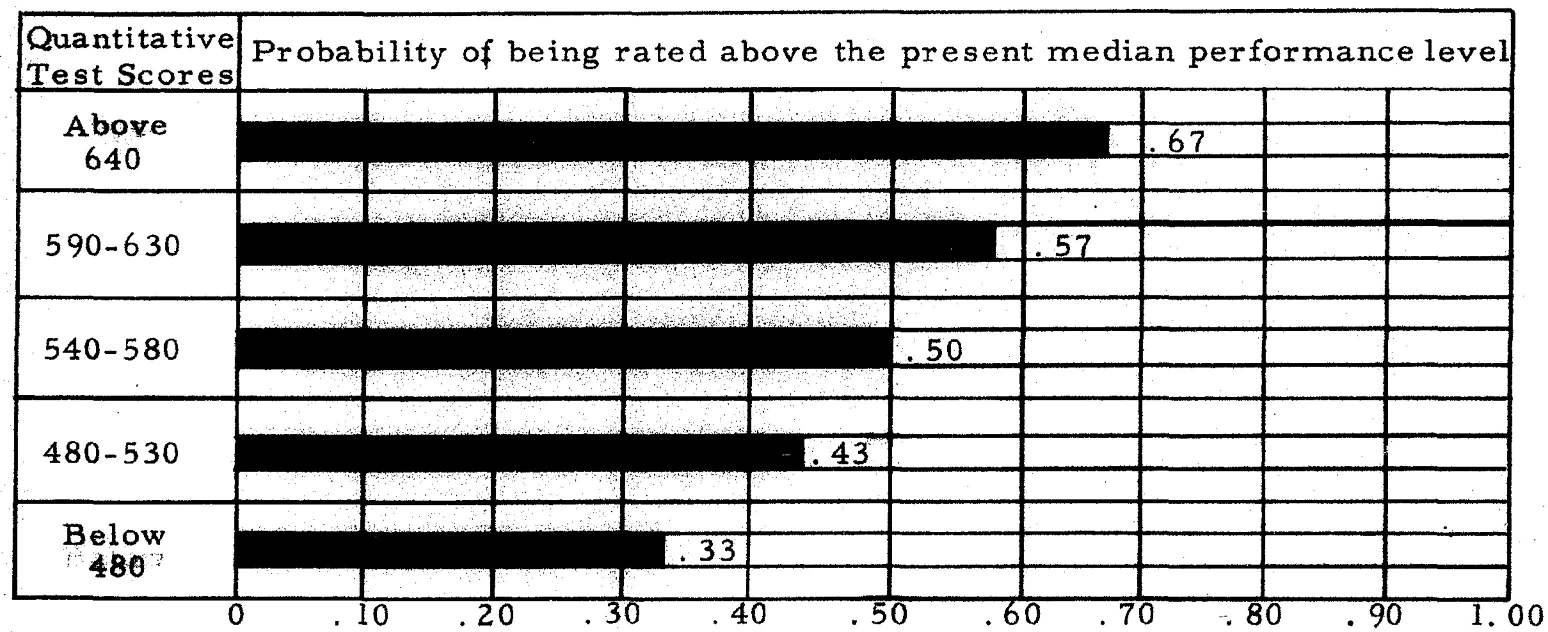


Table 11

Expectancy Chart: Chemistry

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Chemistry

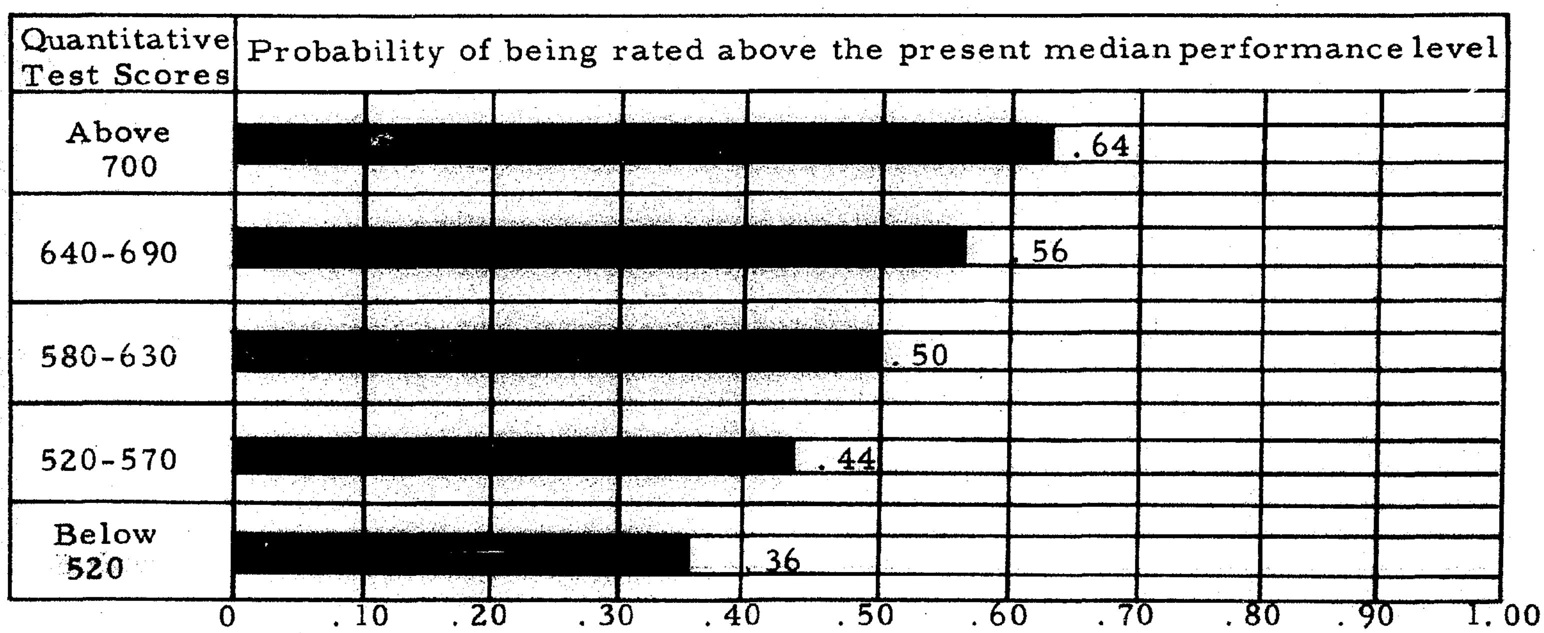


Table 12

Expectancy Chart: Chemistry

Probability that a Student with the Indicated Test Scores will be Rated

Above the Present Median Performance Level in Chemistry

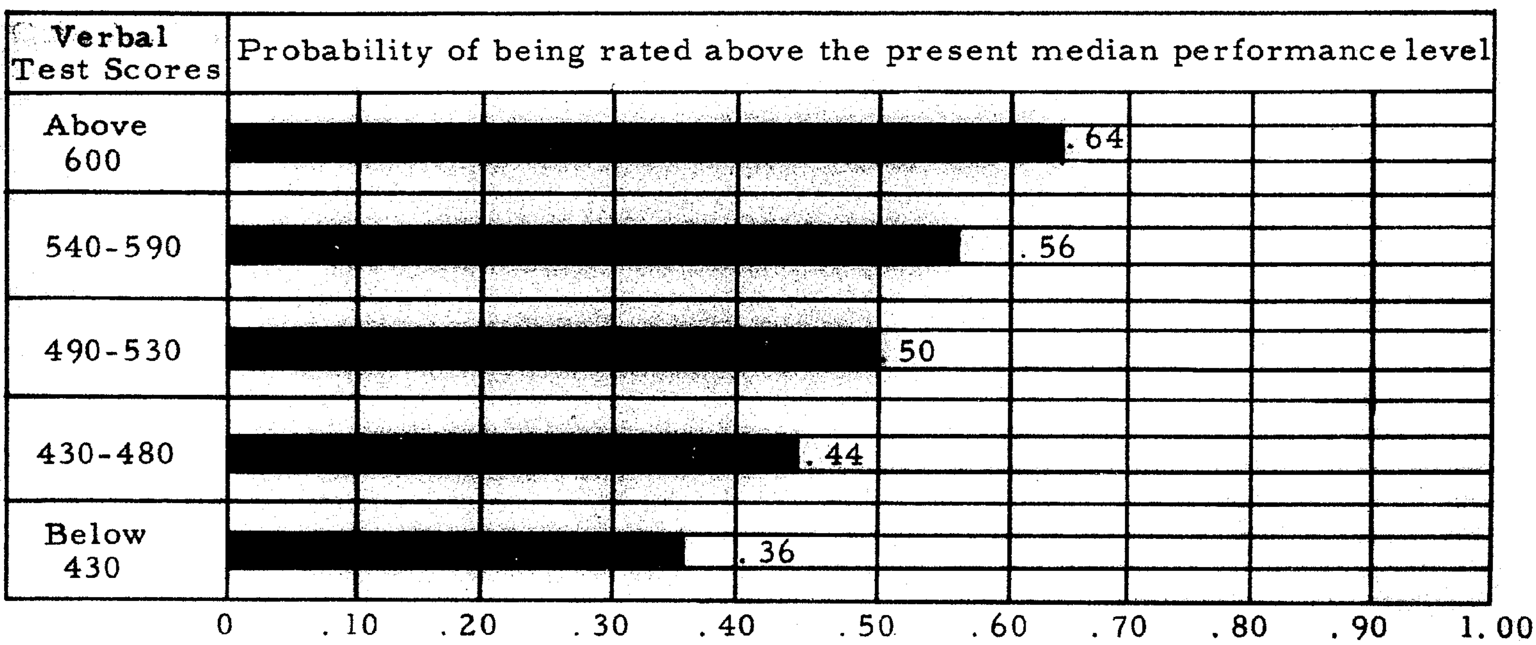


Table 13

Expectancy Chart: Civil Engineering

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Civil Engineering

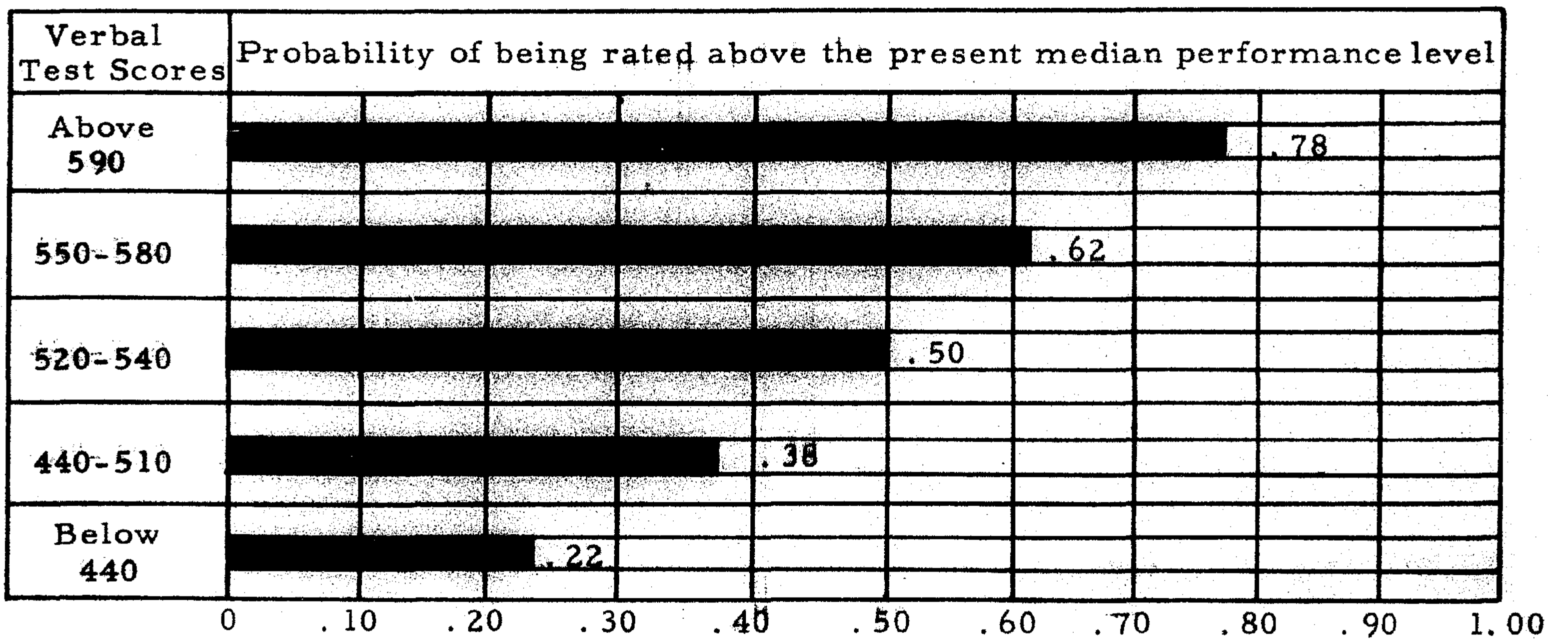


Table 14

Expectancy Chart: Industrial Engineering

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Industrial Engineering

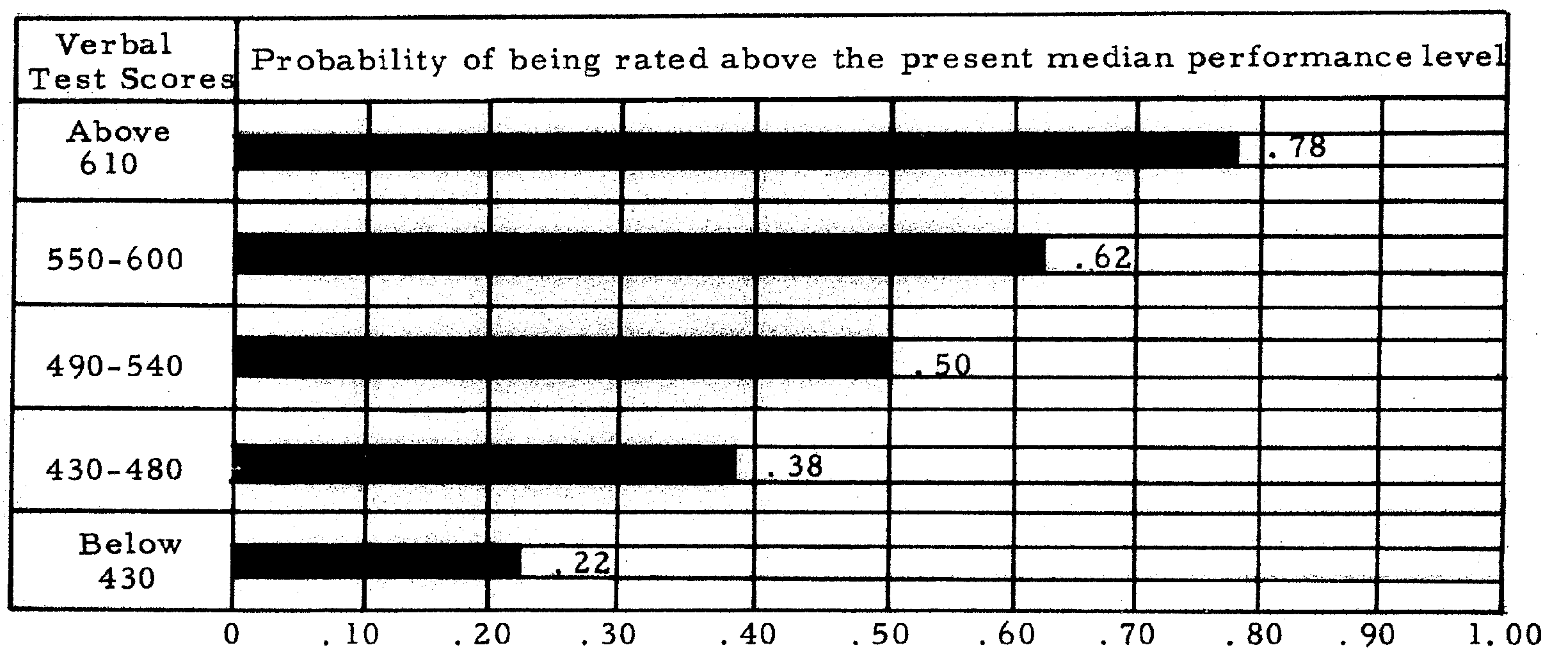


Table 15

Expectancy Chart: Pharmacy

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Pharmacy

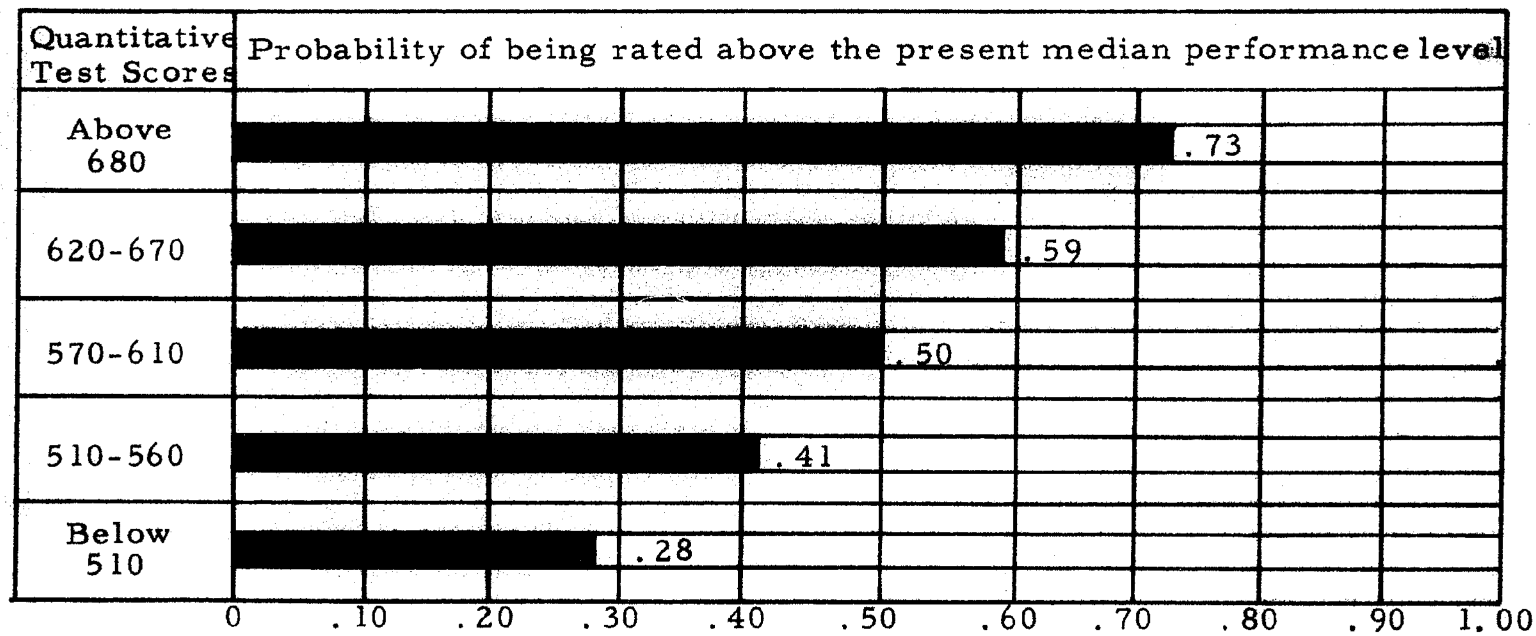


Table 16

Expectancy Chart: Clinical Psychology

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Clinical Psychology

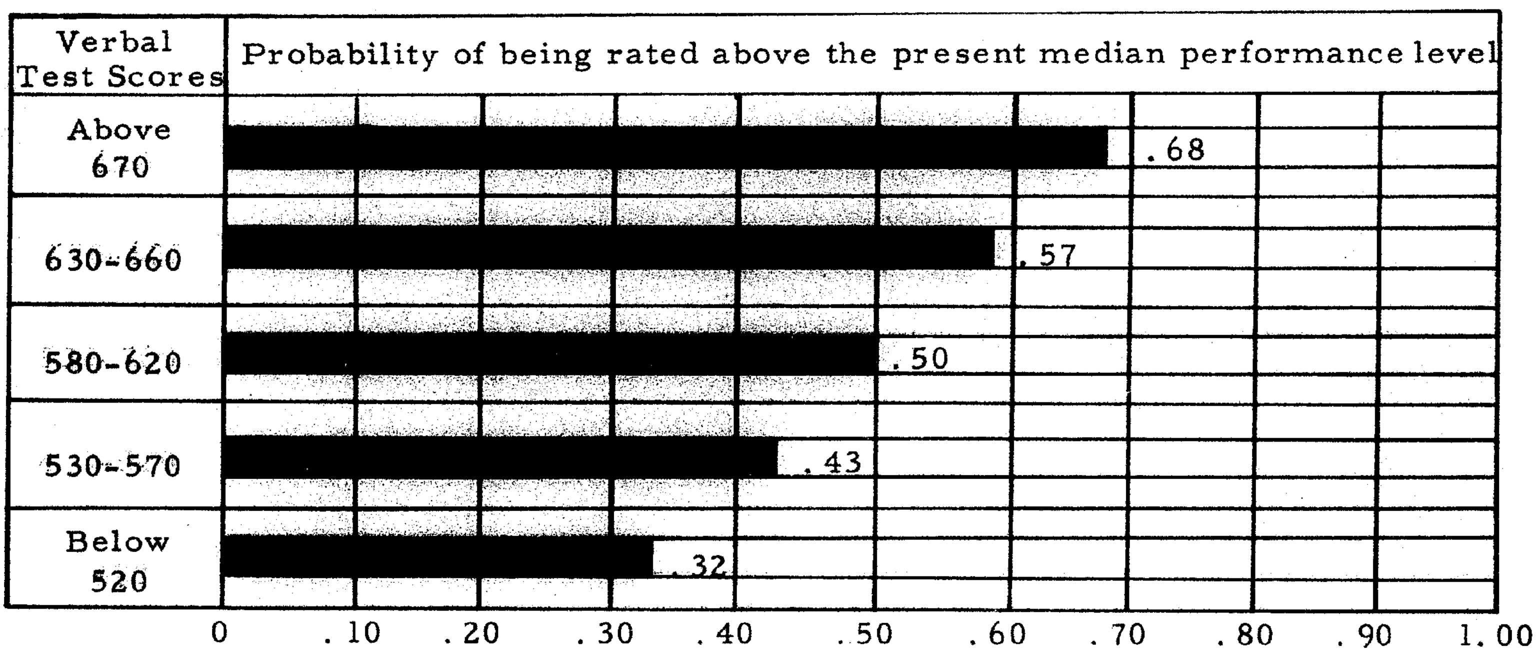


Table 17

Expectancy Chart: Experimental Psychology

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Experimental Psychology

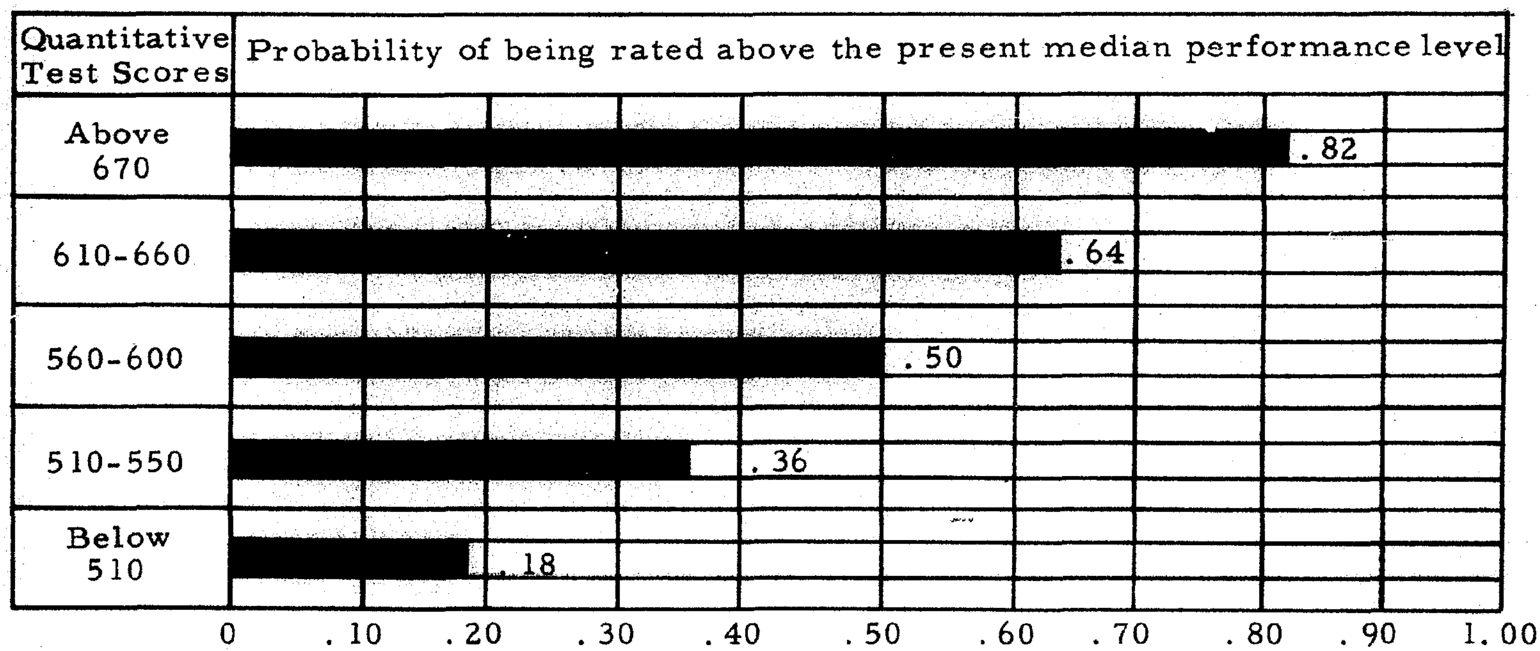


Table 18

Expectancy Chart: Experimental Psychology

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Experimental Psychology

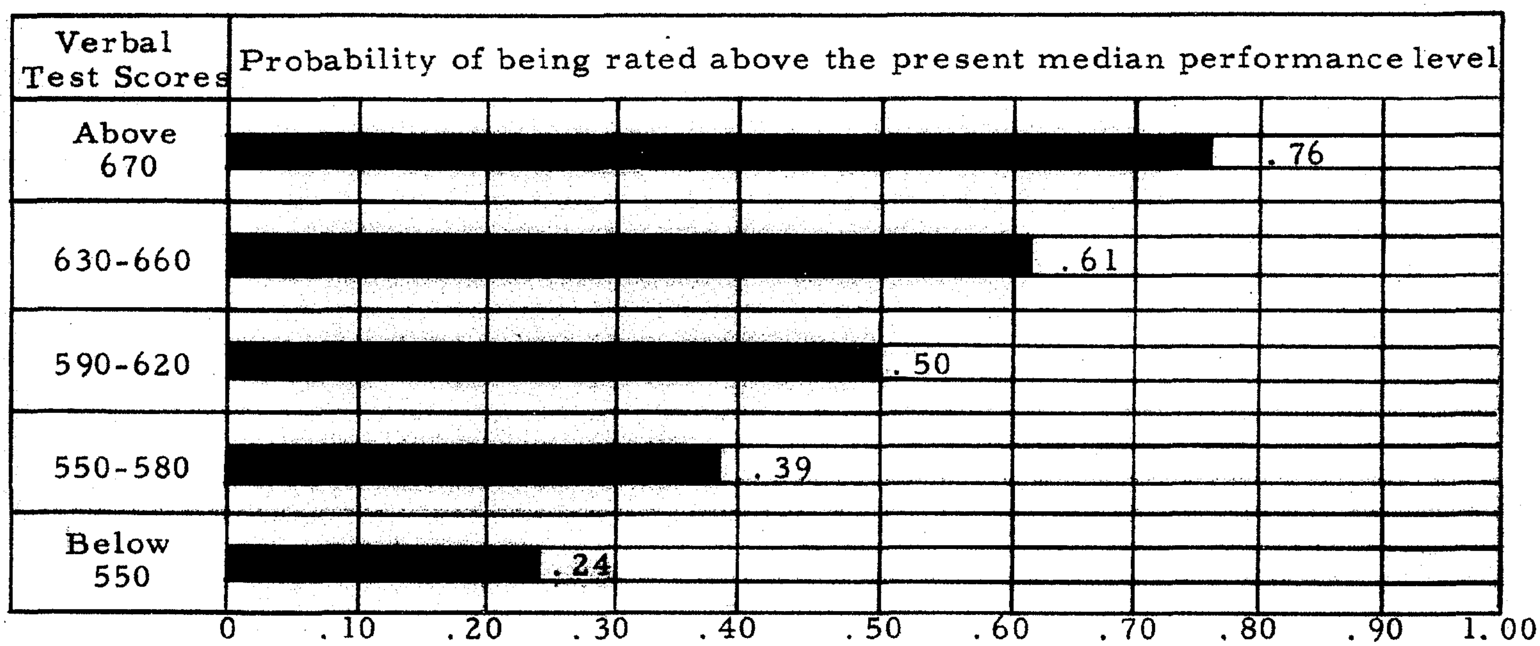


Table 19

Expectancy Chart: Sociology

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Sociology

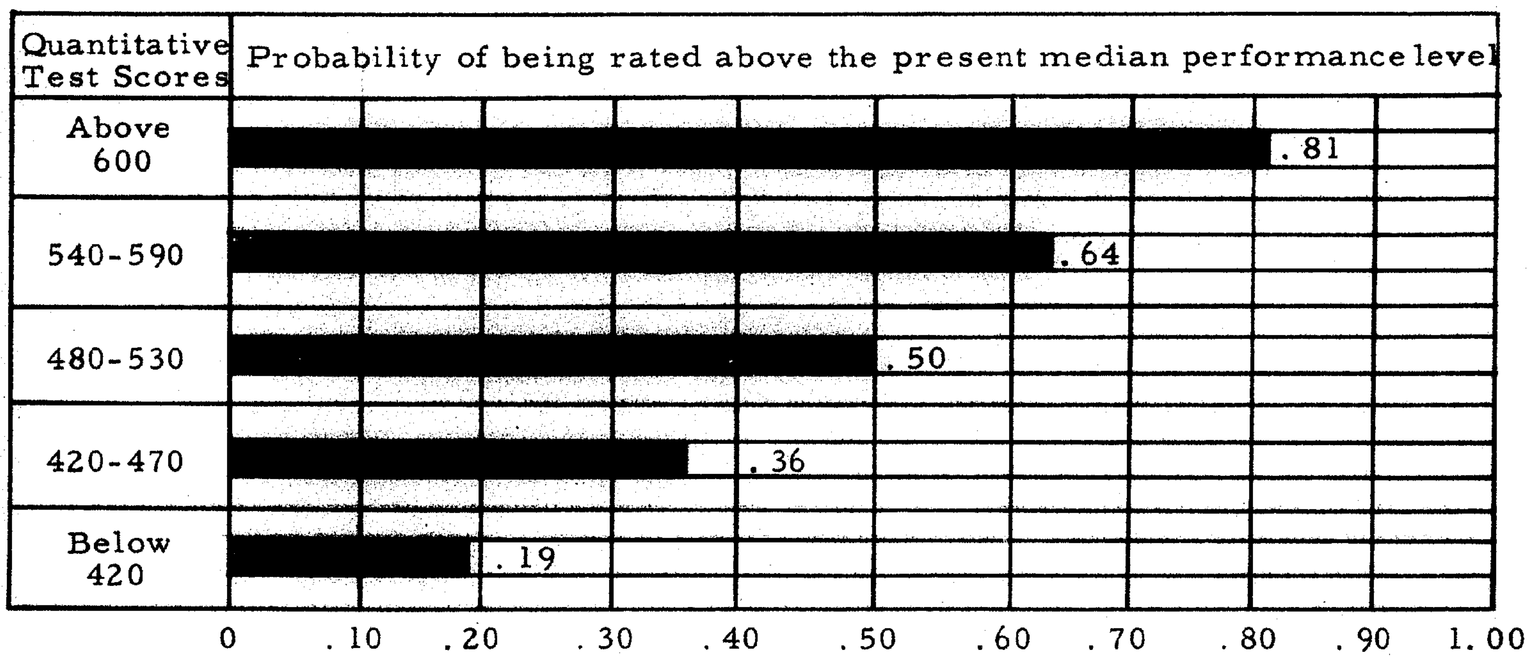
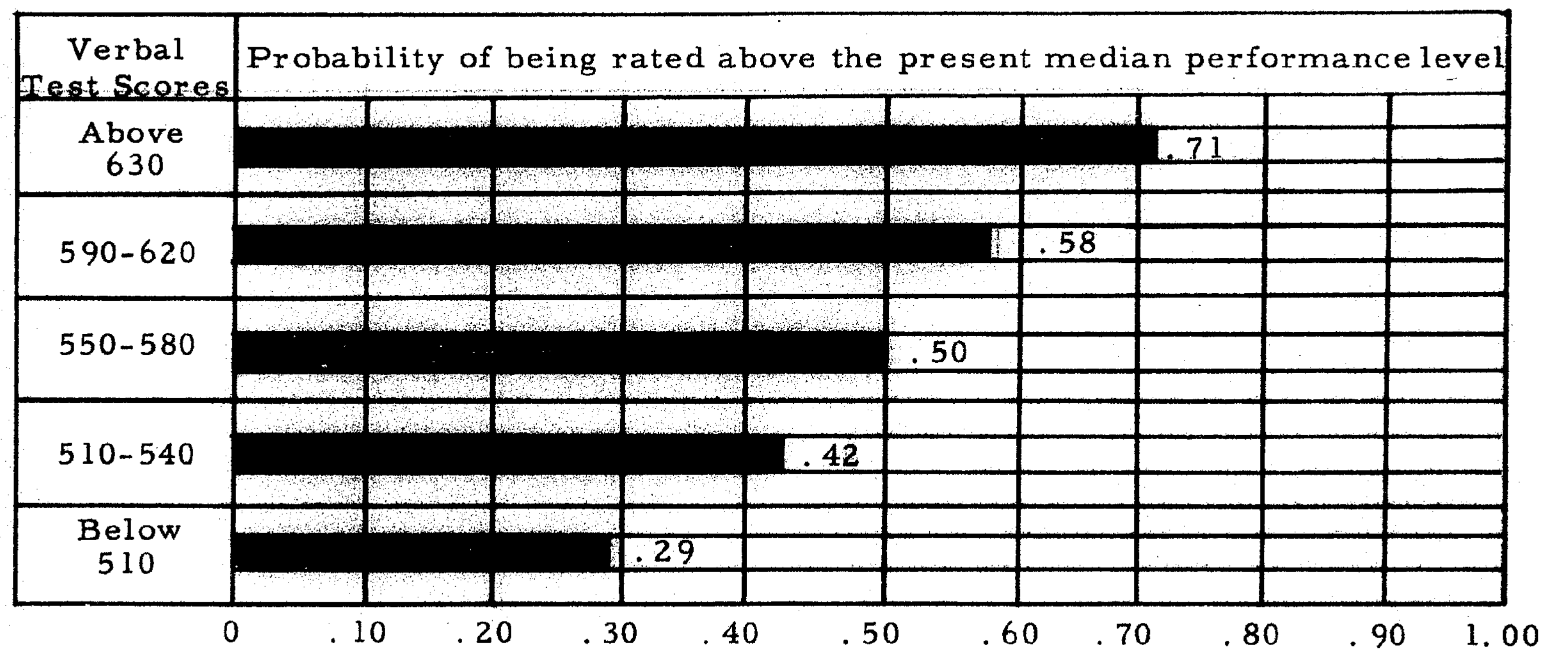


Table 20

Expectancy Chart: Sociology

Probability that a Student with the Indicated Test Scores will be Rated Above the Present Median Performance Level in Sociology



veniences in interpretation that arise when empirically derived score distributions are used. Further research of this procedure by Tiffin and Vincent (1960) has verified the conclusions and recommendations of Lawshe et al in their paper.

Recommendations

It is recommended that the departments involved in this study require applicants to submit test scores on the Graduate Record Examination Aptitude Test as part of the admission procedure. These departments should establish minimum required scores for admission of applicants. Most departments could eliminate a large proportion of their low-performing and unsuccessful students by establishing minimum cut-off scores around the 20th percentile of test scores within their respective departments. Except for the Civil Engineering and Industrial Engineering departments, these cut-off scores should only be established and used with that section of the test which had a significant relationship with performance.

SUMMARY

The primary purpose of this study was to investigate the relationship between various criteria of academic and research performance in graduate school and scores on the Aptitude Test of the Graduate Record Examination. The secondary purpose was to determine the factor structure of the various criteria of success in graduate school.

Using grade point averages and faculty ratings of various traits of graduate student performance, only one overall, general performance factor was found. All of the performance measures were highly intercorrelated and there was not even a suggestion of a second factor of graduate school performance.

Scores on the Graduate Record Examination Aptitude Test were significantly and positively related to the overall performance factor. It was concluded that the Aptitude Test would be a useful and reliable addition to the selection procedure for the departments involved in this study.

LIST OF REFERENCES

APA, AERA, and NCMUE. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.* (supplement), 1954, *51*, 13-18.

Buros, O. K. *The fifth mental measurements yearbook.* Highland Park, New Jersey: Gryphon Press, 1959. Pp. 467-468.

Brogden, H. E. & Taylor, E. K. The theory and classification of criterion bias. *Educ. psychol. Measmt*, 1950, *10*, 159-186.

Buckner, D. N. The predictability of ratings as a function of interrater agreement. *J. appl. Psychol.*, 1959, *43*, 60-64.

Capps, M. P. & DeCosta, F. A. Contributions of the Graduate Record Examinations and the National Teacher Examinations to the prediction of Graduate School success. *J. educ. Res.*, 1957, *50*, 383-389.

Cochran, W. G. & Cox, Gertrude M. *Experimental Designs.* New York: Wiley, 1957.

Denenberg, V. H. & Besco, R. O. A method of obtaining a common rank order scale from incomplete data. *J. appl. Psychol.*, in press.

Durnall, E. J. Predicting scholastic success for graduate students in education. *School and Soc.*, 1954, *80*, 107.

Dvorak, Beatrice J. Advantages of the multiple cut-off method. *Personnel Psychol.*, 1956, *9*, 45-47.

Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, *16*, 407-423.

Educational Testing Service. *Prospectus for deans and advisors.* Princeton, New Jersey: Educ. Testing Service, 1959.

Edwards, A. L. *Techniques of attitude scale construction.* New York: Appleton-Century, 1957.

Fisher, J. T. The value of tests and records in the prediction of college achievement. Unpublished doctoral dissertation, Univer. of Nebraska, 1954.

Fruchter, B. Introduction to factor analysis. New York: Von Nostrand, 1954. Pp. 99-105.

Ghiselli, E. E. Dimensional problems of criteria. J. appl. Psychol., 1956, 40, 1-4.

Grimsley, G. A comparative study of the Wherry-Doolittle and a multiple cutting-score method. Psychol. Monogr., 1949, 63 (2).

Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954. Pp. 263-301.

Guilford, J. P. When not to factor analyze. Psychol. Bull., 1952, 49, 26-37.

Guilford, J. P. & Michael, W. B. Approaches to univocal factor scores. Psychometrika, 1948, 13, 1-22.

Harmon, L. R., Flanagan, J. C., Thurstone, L. L., & Taylor, C. W. Criteria of success in science. National Academy of Sciences, 1954, technical report # 4, 1-55.

Harvey, P. R. The use of the Graduate Record Examination by colleges and universities. Princeton, New Jersey: Educ. Testing Service, 1957.

Hountras, P. T. The use of the Miller analogies test in predicting graduate student achievement. College and University, 1956, 32, 65-70.

Kendall, M. G. Rank correlation methods. New York: Hafner, 1955. Pp. 144-154.

King, D. C. & Besco, R. O. The Graduate Record Examination as a selection device for Purdue Research Foundation graduate research fellows. Lafayette, Indiana: Purdue Univer. Occupational Research Center, 1960. Pp. 1-14.

Kubany, A. J. Use of sociometric peer nominations in medical education research. J. appl. Psychol., 1957, 41, 389-394.

Lannholm, G. V. Letter to the editor. College and University, 1957, 32, 410-411.

Lannholm, G. V. & Schrader, W. B. Predicting graduate school success. Princeton, New Jersey: Educ. Testing Service, 1951.

Lawshe, C. H. Principles of personnel testing. New York: McGraw-Hill, 1948. Pp. 13-19.

Lawshe, C. H., Bolda, R. A., Brune, R. B. & Auclair, G. Expectancy charts II. Their theoretical development. Personnel Psychol., 1958, 11, 545-559.

Lawshe, C. H., Kephart, N. C., & McCormick, E. J. The paired comparison technique for rating performance of industrial employees. J. appl. Psychol., 1949, 33, 69-77.

Manning, W. H. & DuBois, P. H. Gain in proficiency as a criterion in test validation. J. appl. Psychol., 1958, 42, 191-194.

Parkhurst, N. M. & Suddarth, Betty. Potential enrollment for Indiana colleges and universities. Lafayette, Indiana: Purdue Univer., 1955.

Patterson, J. D. The prediction of graduate student potential in chemistry. Unpublished M.S. thesis, Purdue Univer., 1958.

Robinson, D. W. A comparison of two batteries of tests as predictors of first year achievement in the Graduate School of Bradley University. Yearb. nat. Council on Measmts used in Educ., 1958, 15, 118-127.

Roper, Elmo & Associates. Why college enrollments may triple by 1970. College Board Review, 1960, 40, 18-19.

Ryans, D. C. An analysis of comparisons of certain techniques for weighting criterion data. Educ. psychol. Measmt, 1954, 14, 449-458.

Schulz, M. K. & Angoff, W. H. The development of new scales for the aptitude and advanced tests of the Graduate Record Examination. J. educ. Psychol., 1956, 47, 285-294.

Thompson, R. B. Impending tidal wave of students. Columbus, Ohio: Ohio State Univer., 1954.

Thorndike, R. L. Personnel selection. New York: Wiley, 1949. Pp. 119-159.

Thurstone, L. L. _Multiple-factor analysis_. Chicago: Univer. of Chicago Press, 1947.

Tiffin, J. & McCormick, E. J. _Industrial psychology_. Englewood Cliffs, New Jersey: Prentice-Hall, 1958. Pp. 91-94.

Tiffin, J. & Vincent, N. L. Comparison of empirical and theoretical expectancies. _Personnel Psychol_., 1960, _13_, 59-64.

Walton, W. W. The role of evaluation in special scholarship selection programs. _J. exp. Educ_., 1956, _25_, 41-42.

Webb, Sam C. Differential prediction of success in a graduate school. _J. educ. Res_., 1956, _50_, 45-54.

Webb, Sam C. Predictors of achievement in graduate school. _J. appl. Psychol_., 1951, _35_, 265-271.

APPENDIX A

Table 21

Graduate Record Examination Aptitude Test Norms

Based on 80 Foreign Students from Non-English

Speaking Countries

Quantitative Scores	Percentile	Verbal Scores	Percentile
840	99	580	99
820	99	560	97
800	97	540	96
780	96	520	95
760	95	500	91
740	92	480	86
720	88	460	84
700	87	440	82
680	86	420	79
660	82	400	73
640	75	380	65
620	73	360	57
600	63	340	46
580	52	320	32
560	49	300	29
540	47	280	19
520	38	260	13
500	33	240	9
480	23		
460	13		
440	7		
420	6		
400	4		
380	1		

Table 22

Graduate Record Examination Aptitude Test Score Norms Based on 850 Enrolled Purdue Graduate Students to September 1959*

Quantitative Scores	Percentile	Verbal Scores	Percentile
900		900	
880	99	880	
860	99	860	
840	98	840	
820	97	820	
800	95	800	99
780	93	780	99
760	90	760	99
740	86	740	99
720	84	720	98
700	79	700	95
680	70	680	91
660	65	660	87
640	60	640	81
620	52	620	76
600	46	600	74
580	39	580	63
560	34	560	56
540	28	540	46
520	22	520	38
500	18	500	32
480	14	480	25
460	9	460	18
440	6	440	13
420	5	420	9
400	4	400	7
380	3	380	5
360	2	360	4
340	1	340	2
320	1	320	1

*Foreign Students not included.

APPENDIX B

To Professor ______________________

As part of a Purdue University sponsored project to study selection techniques for admission to Graduate School, the Department of Industrial Engineering and the Department of Psychology are cooperating to investigate the possible use of the Graduate Record Examination as a device for assisting [illegible] the selection of graduate students in Industrial Engineering. One big problem in this type of [illegible] is to define success as a graduate student. [illegible] do not reflect a complete picture of a graduate student's [illegible] and research ability would be desirable. The only practical method of gathering such information is through evaluations [illegible] staff members. Before this can be done, it is necessary to know who can evaluate whom. Below is a list of the graduate students in Industrial Engineering. Please indicate, by placing a check after the name of each student whom you feel qualified to evaluate on fairly broad categories of performance as a graduate Industrial Engineering student. [illegible] return this form to Dr. J. H. Greene at your earliest [illegible]

Aingst[illegible], [illegible] L.
Barn[illegible], [illegible]
Bodworth, David L.
Bhatia, Harik[illegible] P.
Bindstiedler, Andre [illegible]
Blakesley, James T.
Bevapragasam, Christopher [illegible]
Foley, Charles H.
Gilbert, Michael S.
Glickstein, Aaron
Herman, Engelbert W.
James, Charles F.
Kurfess, Roland H.

Linn, Ian Barry
Mets, Theodore D.
Miller, [illegible] B.
[illegible]
[illegible], Roland [illegible]
[illegible], [illegible]
[illegible], [illegible]
Srinivasan, [illegible] K.
Taylor, James T.
Undercuffler, Edwin N
Wagner, Larry W.
Wilkin, John F.
Zaccarelli, Sergio

Professor ______________________

The Purdue Research Foundation and the Graduate School are sponsoring a study to improve selection techniques for graduate students at Purdue. The Industrial Engineering Department and the Psychology Department are cooperating to investigate the value of the Aptitude Test of the Graduate Record Examination for predicting success in graduate work in Industrial Engineering.

Grade Point indexes are a measure of the student's success in graduate school, but it is felt, especially in Industrial Engineering that some measures of self-direction and research ability are needed to evaluate a student's performance. For this reason we are asking you to complete the enclosed rating forms. Only the students whom you indicated you could evaluate are included.

On the 8" x 10" rating forms describe each student by placing a check after each trait to be evaluated, in the box that most nearly represents your evaluation of the student. Compare each student with all graduate students you have known.

There is also a deck of IBM cards with two names on each card. Place a pencil check mark under the name of the student you consider to be the better student of the pair. Make your judgment on the basis of over all performance in graduate work including research.

Please return both sets of ratings to Dr. J.H. Greene at your earlist convenience. Thank you for your cooperation.

Sincerely,

Dr. D.C. King
Robert Besco

Student ____

Rater ____

A typical group should be distributed like this:	Below Average	Average	Above Average	Good	Unusually Good	Out Standing
	Lower 40%	Middle 20%	Next 15%	Next 10%	Next 10%	Top 5%
His degree of mastery of the fundamental knowledge in his field						
His possession of a fertile imagination and originality in his field						
His self-reliance and independence in his work						
His motivation toward a successful productive career						
His ability to plan and design research in his field						
His ability to carry out research in his field						

In the rating scales below please describe the student by checking the box that most nearly represents your global evaluation of the student. Compare this student with all other graduate students with whom you have recently had contact. How do you regard him in terms of his overall graduate performance?

Below Average Lower 40%	Average Middle 20%	Above Average Next 15%	Good Next 10%	Unusually Good Next 10%	Out Standing Top 5%

If at the time of this student's application for admission, you knew that he was to perform in graduate work as he has, what recommendation would you have made concerning his admission to graduate study in your department? Check the one category below that most nearly describes the action you would have recommended.

Admission denied	Conditional Admission as a Master's candidate only	Admission as a Master's candidate only	Conditional Admission as a Ph.D. candidate	Admission as a Ph.D. candidate

Example of Paired Comparison Rating Form

APPENDIX C

Table 23

Matrix of Intercorrelations and Residuals: Agronomy*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Relia-bility
1. Knowledge	83	02	-10	-06	-02	-05	-03	-01	-07	-02	00				28
2. Imagination	85	88	-03	02	01	02	-03	-01	-14	-06	-13				31
3. Independence	76	85	91	03	-02	-03	-04	01	-05	-06	-01				35
4. Motivation	78	88	92	93	02	04	-04	-04	-15	-10	-08				36
5. Design Research	84	89	89	92	93	05	-02	-05	-10	-08	-12				29
6. Conduct Research	81	89	88	92	95	93	-03	-08	-09	-09	-09				28
7. Overall Performance	87	89	92	89	93	92	95	00	-05	-06	00				36
8. Readmission	58	59	63	57	57	54	65	49	02	-04	-10				32
9. Paired Comparison Z Score	70	64	76	64	71	71	81	57	87	09	10				50
10. Paired Comparison Matrix Score	83	81	85	78	82	81	89	57	89	91	09				72
11. Grade Point Average	42	30	43	35	32	35	46	20	49	53	41				
12. Months in School	23	23	15	15	24	25	19	09	15	15	-11	--			
13. GRE Quantitative	22	25	25	17	19	20	28	00	39	31	31	-03	--		
14. GRE Verbal	-02	-07	-14	-27	-10	-13	-05	-13	14	-01	-07	14	46	--	

N = 42

Correlation of .29 significant at the .05 level.

Correlation of .39 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 24

Matrix of Intercorrelations and Residuals: Chemistry*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	84	05	-06	-02	-07	-05	01	00	-03	-03	14				77
2. Imagination	81	77	04	-01	04	01	03	-03	-09	-07	05				51
3. Independence	74	80	88	11	09	09	03	07	-11	-08	-11				61
4. Motivation	66	62	77	67	04	06	03	-18	-07	-03	-07				29
5. Design Research	74	81	90	71	87	07	02	-02	-08	-06	-10				61
6. Conduct Research	76	78	90	73	89	87	03	00	-11	-09	-06				72
7. Overall Performance	86	84	88	74	88	89	90	-01	-07	-04	00				67
8. Readmission	62	56	55	34	61	63	65	58	08	02	11				63
9. Paired Comparison Z Score	55	47	48	42	51	49	55	53	63	28	-01				44
10. Paired Comparison Matrix Score	65	58	60	53	63	60	68	55	78	71	02				63
11. Grade Point Average	79	61	54	47	56	60	69	62	47	52	67				
12. Months in School	04	08	03	01	06	14	08	06	19	15	19	--			
13. GRE Quantitative	32	28	27	21	31	28	26	25	18	15	35	-06	--		
14. GRE Verbal	29	14	10	03	12	11	17	35	18	19	21	11	43	--	

N = 82

Correlation of .22 significant at .05 level.

Correlation of .28 significant at .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 25

Matrix of Intercorrelations and Residuals: Civil Engineering*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	86	02	00	00	03	05	09	04	-08	-05	01				44
2. Imagination	94	96	01	01	04	02	03	00	-05	-05	-06				33
3. Independence	89	92	96	08	00	-02	-01	02	-06	-12	05				19
4. Motivation	88	88	92	92	-02	-01	03	01	-09	-09	-07				14
5. Design Research	88	94	88	82	97	12	03	01	-10	-05	-12				29
6. Conduct Research	87	89	83	80	96	96	03	02	-13	-03	-16				29
7. Overall Performance	97	97	91	90	93	90	92	05	-08	-08	-04				46
8. Readmission	84	82	82	76	80	78	87	80	-13	-15	04				53
9. Paired Comparison Z Score	64	69	65	59	61	55	66	51	91	31	08				27
10. Paired Comparison Matrix Score	68	70	61	60	67	67	68	50	90	92	00				72
11. Grade Point Average	62	57	66	51	49	43	59	59	57	50	69				
12. Months in School	36	39	46	43	42	48	37	27	29	39	14	--			
13. GRE Quantitative	25	26	22	15	27	25	23	18	-04	00	34	20	--		
14. GRE Verbal	26	19	08	08	12	18	21	07	33	42	13	08	38	--	

N = 26

Correlation of .37 significant at .05 level.

Correlation of .47 significant at .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 26

Matrix of Intercorrelations and Residuals: Industrial Engineering*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	90	03	-03	-06	-08	01	00	00	06	06	-03				12
2. Imagination	76	93	04	00	09	04	-02	-07	-03	-02	-12				15
3. Independence	73	81	91	10	05	-09	06	-07	-10	-10	-02				14
4. Motivation	74	81	87	95	-07	-09	02	-03	-02	07	-07				04
5. Design Research	68	86	78	69	97	12	02	-07	-10	-24	04				02
6. Conduct Research	83	88	71	74	91	97	-02	-01	01	-07	-01				02
7. Overall Performance	85	85	89	88	84	88	98	-02	-04	-10	04				17
8. Readmission	83	77	74	81	73	86	88	97	01	08	10				55
9. Paired Comparison Z Score	84	76	66	77	65	83	81	83	88	12	-02				24
10. Paired Comparison Matrix Score	78	71	60	80	46	69	68	84	83	97	-11				76
11. Grade Point Average	71	63	70	68	75	77	84	88	72	57	91				
12. Months in School	62	41	39	30	49	49	54	49	22	26	53	--			
13. GRE Quantitative	-06	-27	-13	-21	-26	-34	30	-06	-30	02	-08	25	--		
14. GRE Verbal	26	28	30	35	12	12	14	30	13	47	11	13	42	--	

N = 16

Correlation of .47 significant at .05 level.

Correlation of .59 significant at .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 27

Matrix of Intercorrelations and Residuals: Pharmacy*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	84	05	03	-01	01	-06	00	-07	-02	00	07				46
2. Imagination	84	85	06	01	05	03	03	-07	-09	-04	-06				49
3. Independence	83	87	93	02	11	10	03	12	-12	-10	-13				57
4. Motivation	81	84	86	91	03	07	03	-01	-09	-09	-01				57
5. Design Research	77	87	94	88	94	12	01	-06	-07	-07	-15				57
6. Conduct Research	74	83	91	90	94	92	01	-07	-08	-09	-15				51
7. Overall Performance	86	89	91	93	90	88	94	00	-05	-04	-02				58
8. Readmission	53	54	49	62	56	54	66	58	08	07	12				65
9. Paired Comparison Z Score	74	68	66	71	72	69	78	66	89	18	10				50
10. Paired Comparison Matrix Score	77	74	69	72	73	69	80	66	92	88	04				73
11. Grade Point Average	58	42	39	52	37	36	53	51	59	53	54				
12. Months in School	25	27	24	31	38	33	24	23	31	34	01	--			
13. GRE Quantitative	43	38	31	41	32	37	41	21	37	34	31	-08	--		
14. GRE Verbal	15	07	03	05	06	02	05	03	16	12	27	-20	63	--	

N = 44

Correlation of .29 significant at the .05 level.

Correlation of .37 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 28

Matrix of Intercorrelations and Residuals: Clinical Psychology*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	92	02	-01	-12	01	00	01	00	-04	-01	05				50
2. Imagination	86	83	-01	-02	01	-02	04	-04	-02	-02	01				42
3. Independence	90	82	92	07	-01	-01	-01	-05	-01	01	-01				39
4. Motivation	84	76	91	88	01	-02	00	-07	-03	04	-05				34
5. Design Research	93	85	90	86	97	06	00	00	-04	-01	-09				35
6. Conduct Research	91	81	89	82	97	96	-01	01	-02	-01	-06				38
7. Overall Performance	94	89	91	86	93	91	93	-02	-01	-01	01				44
8. Readmission	77	67	71	65	77	77	76	69	13	-01	08				37
9. Paired Comparison Z Score	79	73	80	73	79	79	82	75	82	06	-02				45
10. Paired Comparison Matrix Score	91	79	91	88	90	89	91	76	88	93	-04				66
11. Grade Point Average	61	54	56	48	49	51	59	56	50	53	49				
12. Months in School	12	18	18	23	10	09	14	03	15	19	09	--			
13. GRE Quantitative	03	-09	-09	-06	-08	-09	-03	03	-03	-02	03	07	--		
14. GRE Verbal	39	22	21	28	34	27	36	20	12	25	02	19	34	--	

N = 40

Correlation of .30 significant at the .05 level.

Correlation of .39 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 29

Matrix of Intercorrelations and Residuals: Experimental Psychology*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	98	14	-06	-11	02	-06	02	-09	06	04	-09				58
2. Imagination	93	97	-04	-11	06	00	02	-14	02	02	-21				56
3. Independence	83	78	99	10	-01	-10	-08	03	-06	-04	-02				45
4. Motivation	72	66	96	96	-03	05	-02	02	-07	-05	07				53
5. Design Research	92	89	92	84	97	01	00	-06	-02	00	00				51
6. Conduct Research	84	83	96	92	95	99	-02	01	-03	-01	-02				44
7. Overall Performance	94	87	95	87	96	94	99	00	00	-01	-01				57
8. Readmission	66	55	80	74	72	79	80	83	-02	-01	09				57
9. Paired Comparison Z Score	93	83	84	77	90	89	93	74	96	07	-03				61
10. Paired Comparison Matrix Score	91	82	86	79	91	90	92	75	96	95	-06				82
11. Grade Point Average	57	40	66	68	69	67	70	66	64	61	83				
12. Months in School	06	02	05	10	03	08	00	13	02	09	06	--			
13. GRE Quantitative	47	53	39	26	51	45	44	37	42	51	17	08	--		
14. GRE Verbal	48	51	25	20	40	32	37	26	42	46	03	26	39	--	

N = 20

Correlation of .42 significant at the .05 level.

Correlation of .54 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 30

Matrix of Intercorrelations and Residuals: Industrial Psychology*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	98	06	00	-03	05	04	04	-05	-03	-03	-11				46
2. Imagination	96	94	00	-01	04	03	04	-08	-03	-04	-08				49
3. Independence	92	90	94	08	00	00	01	-03	-02	-01	-06				50
4. Motivation	80	81	91	86	-02	-01	00	-04	-03	-03	-03				38
5. Design Research	98	96	92	82	98	04	04	-09	-03	-02	-07				49
6. Conduct Research	97	95	93	83	98	98	03	-05	-03	-02	-05				49
7. Overall Performance	98	96	94	85	98	98	99	-05	-01	-07	-07				54
8. Readmission	79	75	81	72	76	81	81	84	05	14	11				55
9. Paired Comparison Z Score	83	81	84	75	84	85	87	84	91	-06	07				59
10. Paired Comparison Matrix Score	84	82	87	76	87	87	83	94	76	91	08				78
11. Grade Point Average	60	62	65	61	65	67	65	76	73	76	74				
12. Months in School	30	34	48	49	29	33	34	40	30	44	48	--			
13. GRE Quantitative	-28	-25	-27	-30	-23	-23	-24	-17	-08	-16	02	-17	--		
14. GRE Verbal	15	20	12	17	15	14	14	20	19	18	11	02	34	--	

N = 37

Correlation of .32 significant at the .05 level.

Correlation of .41 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

Table 31

Matrix of Intercorrelations and Residuals: Sociology*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Reliability
1. Knowledge	98	02	00	-02	-01	01	00	-06	00	03	-02				75
2. Imagination	96	96	-01	-03	-01	-02	01	01	01	00	03				71
3. Independence	93	93	96	04	02	03	01	-02	-06	-02	-02				63
4. Motivation	83	82	88	83	02	01	00	-03	-03	06	-08				56
5. Design Research	94	94	97	88	99	04	01	-04	-04	01	-04				70
6. Conduct Research	94	92	96	86	99	99	01	-04	-05	-01	-06				70
7. Overall Performance	95	96	96	86	97	96	99	00	-05	-03	02				74
8. Readmission	77	85	81	73	81	79	85	88	05	-09	07				78
9. Paired Comparison Z Score	86	86	80	75	84	81	83	82	92	04	04				77
10. Paired Comparison Matrix Score	77	74	72	73	76	75	72	57	72	67	-09				92
11. Grade Point Average	80	86	80	67	80	76	86	80	80	56	87				
12. Months in School	03	01	05	01	02	02	07	11	06	17	05	--			
13. GRE Quantitative	59	53	63	52	64	61	60	55	63	62	52	12	--		
14. GRE Verbal	37	43	50	36	52	47	51	36	32	18	48	08	57	--	

N = 24

Correlation of .39 significant at the .05 level.

Correlation of .49 significant at the .01 level.

*The intercorrelations among the variable are all below the main diagonal, i.e. upper left to lower right. The residual correlations remaining after the extraction of the first factor are all above the main diagonal. The communalities appear on the main diagonal and all the decimal points are omitted.

VITA

Name: Robert Orin Besco

Born: May 11, 1932, Albert Lea, Minnesota

Academic Career:

1950, Diploma, Roosevelt High School, Cedar Rapids, Iowa

1954, B.A., Psychology, University of Iowa, Iowa City, Iowa

1959, M.S., Industrial Psychology, Purdue University

1960, Candidate for the Ph.D. Degree, Purdue University

Work Experience:

1954, Job Analyst, Collins Radio Co., Cedar Rapids, Iowa

1954 to 1958, Pilot, USAF

1958 to 1960, Research Fellow, Purdue University

Professional Societies:

American Psychological Association

Midwestern Psychological Association

Psychometric Society

Sigma Xi

Publications:

"Foreman leadership as perceived by superiors and subordinates," (with C. H. Lawshe) Personnel Psychology, 1959, 12, No. 4, 573-582.

"Evaluation techniques for management development programs," (with J. Tiffin and D. C. King) Journal of the American Society of Training Directors, 1959, 13, No. 10, 13-28.

"The direction and stability of responses in toggle switches," (with J. D. Gilmour and E. J. McCormick) Journal of Applied Psychology, in press.

"The Graduate Record Examination as a selection device for Purdue Research Foundation graduate research fellows," (with D. C. King) Occupational Research Center, Purdue University, February, 1960.

"A method of obtaining a common rank order scale from incomplete data," (with V. H. Denenberg) Journal of Applied Psychology, in press.