THE PREDICTION OF ENGINEERING APTITUDE

\- - - - - - - - - - - -

M. G. de Jersey

\- - - - -

Thesis submitted to the Department of Psychology in

partial fulfillment of the requirements for the

degree of Master of Arts.

\- - - - -

McGILL UNIVERSITY

Montreal

1 9 4 6

PREFACE

In addition to the regular supervision by Dr. C.E. Kellogg, director of this research, the writer has benefitted from conversation with Professor G.A. Wallace of the Faculty of Engineering and with Dr. E.C. Webster of the Department of Psychology at McGill University.

CONTENTS

CHAPTER I.

INTRODUCTION

One of the main purposes of the science of Psychology is the prediction of human behaviour. To attain such a goal, the nature of the behaviour in question must be adequately and reliably known. Therefore, the initial step is one of observation and description, and, when the results of these activities are substantial, one may attempt to predict the behaviour of the individual.

As applied to the educational field, Psychology has made extensive efforts to estimate, in advance, the likely scholastic success of the college student. In particular, the present study seeks to predict the success of a sample of McGill students in the field of engineering training. A survey of the heavy academic mortality rate in engineering colleges of many different universities will make it apparent that the purpose set for this study is decidedly a practical one. For example, at McGill University, of the class entering the first year of Engineering in 1935, only about 43% received degrees in 1939, taking account of both the Spring and Fall Convocation. (The class of 1935 was selected for this estimate because it was the last class to complete the undergraduate course before the beginning of the War.) If we consider the students who were included in the present experiment, we find that, of the persons who wrote the final examinations for the first year of Engineering in the spring of 1945, about one-third did not register in the fall for the second year. Of course, in this last statement we have the influence of the War counting as one of the factors. In the University of Saskatchewan, only 65% of those entering the College of Engineering as freshmen reach the second year, 45% reach the third year, and 35% reach the fourth year.(27)* Holcomb and Laslett (24) have reported that 38.1% of freshman engineering students in Oregon State College did not re-enroll as sophomores in the next year. These authors quote further information showing that of 5,338 freshmen in Engineering in twenty-five large schools of Engineering in the United States, 39.1% did not register again as sophomores (for 1926).

---

*Throughout this thesis, numbers in brackets following authors' names or statements indicate the number of the reference as it is listed in the Bibliography.

It is not implied that these values indicate academic deficiency alone, but, at the same time, this is undoubtedly the greatest single cause of leaving the course. As Bingham has observed, engineering training is a specialized branch of college work setting high demands on intellectual capacities of the student (6). He refers to a student mortality rate of about 62%, among the reasons for which he includes interests as well as intellectual factors. At least it is sufficiently clear that, if prediction can be improved in this branch of college training, any efforts involved will be well justified. We should consider that these low graduating proportions point to many effects which are detrimental to the individual student: waste of human time, effort, and economic resources, plus the possible bad influence of failure upon the personality of certain students (27).

Interest in this problem stems from both the educational administrator and the student. As early as 1928, Mann remarked about the increasing interest taken by the administrations of American colleges and universities in "ascertaining at an early period in the college student's career just what sort of a creature he is, what his chances are of succeeding in university work, whether this will hold for one, for two, or perchance for four years of the course"(29). However, that this is not an easy accomplishment has been frequently pointed out since scholastic success does not depend upon a single quality of behaviour. It is thus unfortunate, for the cause of prediction, that scholastic achievement is not only based upon fundamental abilities, but is sprinkled generously with numerous other factors such as interests, motivation, distribution of time, study methods, personal problems of vocational choice and ambition, etc. Prediction in this field must be looked upon as a highly complicated task if it aims for high efficiency.

At this point, we should elaborate somewhat upon the implications of the thesis title: "The Prediction of Engineering Aptitude". According to Warren's "Dictionary of Psychology", an aptitude is "a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill, or set of responses, such as the ability to speak a language, to produce music, etc." Bingham expands this definition and refers to an aptitude as "a condition symptomatic of a person's relative fitness, of which one essential aspect is his readiness to acquire proficiency - his potential ability - and another is his readiness to develop an interest in exercising that ability"(6). This is a very broad view of aptitude in so far as it includes satisfaction in, or liking for, the work in question. In actual practice, specific aptitude tests usually fit Warren's definition since only the ability to profit from training is measured. The interests and likely satisfaction of the individual in a certain type of work must be estimated by means of other tests. As far as the use of the term aptitude is concerned in this thesis, evaluation of engineering aptitude is based upon two main factors: a psychological test measuring ability in elements considered necessary for engineering training success, and pre-engineering scholastic achievement. Interest and satisfaction, as in the Bingham definition,are not measured except to the extent that they are reflected in the previous achievement of the student in subjects similar to engineering courses.

As a concluding consideration in this introduction, let us anticipate that a fairly adequate method of prediction is available. In that event, in what ways should the information of prediction be used and what advantages would result? The first advantage,that seems probable, would be an important effect on the teaching methods and attitudes of the instructional staff as a result of an increase in the homogeneity of the student body with respect to aptitude. In a conversation with Professor Wallace of the Faculty of Engineering, the point was made that the decrease in the spread of ability,at the low end of the scale,might have the tendency to remove some of the hampering effects felt by the instructor when he must cater to the students who are of lowest aptitude. Having removed a proportion of those who are most likely to retard the class, the level of teaching might reflect the higher status of aptitude present among the students. It will be further observed that the morale or enthusiasm of the instructor might increase as he feels a greater responsiveness to his lectures. Undoubtedly these several factors would culminate in a somewhat higher standard of graduate.

From the student's point of view, the chief advantage of the same predictive instrument lies in its counselling powers. As a matter of fact, greater emphasis has been placed upon the guidance aspect of prediction than upon the selection aspect. The following quotations will describe typical views along this line:

> "The Iowa Placement Examinations are less a prognosis test than an educative procedure. Their aim is not primarily to predict academic success, but to render its attainment more likely; that is, to give aid in the setting up of educational conditions such that sound principles of selection, class-sectioning, and curriculum organization may be more effectively applied to the securing of maximum performance on the part of each student." - Stoddard (42, p. 214).

> "Testing should be regarded as a means of revealing to the student his own potentialities and weaknesses, of holding out to him desirable goals of accomplishment, and of stimulating him to further improvement. Tests achieve their fullest purpose when they reveal growth, act as spurs to improvement, and serve as instruments which students will find useful in their own evaluation." Davis (14, p. 44).

In a study of engineering students at the University of Washington, Wilcox (48) suggested gathering together the students who were low in two or more placement tests and informing them of their possibilities of success. Their cases should be studied in the hope of establishing readjustments,as a result of which some would be discouraged from taking Engineering while others might be salvaged. This procedure would lead to a desirably earlier adjustment to college environment than would be possible by the trial-and-error method. Compton (11) has found that slightly more than one-half of a large number of college students benefitted from a knowledge of their mental ability test score. None the less, it is to be strongly recommended that the procedure of informing the students about their predicted scores or test scores should be carried out through a guidance or counselling service in order to make this knowledge most valuable.

The goal of the present thesis is thus one of practical value both for the educator and the student. The findings of the investigation may be applied for the purpose of selection alone, or of counselling alone, or, preferably, for both purposes.

CHAPTER II

PREVIOUS EXPERIMENTATION

(A) Background of the Problem, and Typical, General Results.

Historical Note. In so far as the act of educational prediction is dependent upon the existence of testing devices, one could seek out its origin in the work of the men who were responsible for the emergence of the mental testing program. This would lead us back to the last part of the nineteenth century and the early years of the twentieth century (51). Wundt , Ebbinghaus, Galton, Stern, Cattell, are but a few of the names associated with the foundation of a great mental test movement. However, our chief interest is not in the actual groundwork of the movement but rather in the practical applications of the psychological tests, so developed, for educational prediction.

In this approach the name of Alfred Binet is conspicuous as one who produced a testing instrument which could be employed in practical work. His first scale for measuring general intelligence was published in 1905 and revised in 1908 and 1911. The value of this work originated in Binet's view of intelligence as an integration of many different mental processes. Thus the tasks and problems set in this test called for varied intellectual activities rather than narrowly restricted phases of behaviour as had the previous types of tests. In addition, Binet prepared the way for comparative studies of intelligence by making use of the Mental Age concept and by grouping the test items into an age scale. It is obvious that Binet contributed much to the production of tools necessary for educational prediction. In America these tests were revised and became widely used. One of the best known of these early revisions was prepared by Terman in 1916.

The use of the Army Alpha and Beta tests, which were administered to almost two million men, gave a great boost to the principle of group testing and, consequently, was followed by the appearance of many other types of group tests. However, as far as application was concerned, Leitch (28) has observed that, up to about 1920, mental tests had been utilized mainly in institutions, the U.S. Army, and public schools. He dates the real beginning of intelligence testing of college students in America from the 1918-19 academic year and points to the reason as a carry-over effect from the army testing program. This is an appropriate choice of date for the beginning of substantial predictive efforts of a psychological nature at the college level since earlier efforts covered very restricted phases of ability. For example, typical correlations between college grades and such tests as speed of learning, memory, association, card sorting, etc., extended from .09 to .44 during the early years of the twentieth century.

The first type of psychological prediction at the college level aimed at the estimation of general college success rather than achievement in any specific faculty or school. Moreover, the testing instrument invariably employed was the intelligence test. A very brief survey of typical findings along these lines will allow us to lead up naturally to the period when a greater differentiation was made with respect to both the testing instrument and the realm of prediction.

Early studies employing the Stanford Revision of the Binet were not particularly successful. For example, for a group of 48 college students, Caldwell reported a correlation of .44 between I.Q. and grades (10) while an investigation of two classes (98 to 103 cases) at Bryn Mawr yielded values of .298 and .197 with freshman marks (1). The rapid growth of group tests from 1920 - 1930 produced more and more efficient and practical instruments revealing somewhat higher relationships with college success. However, the increase did not usually exceed the limits mentioned by Whipple in a summary of the use of intelligence tests in colleges (47). His statement is dated 1922 and refers to a validity coefficient of .30 as of practical significance and the most likely range .40 to .60. Later in the same decade, Stoddard quoted from Pintner a series of sample correlations between intelligence tests and scholastic grades of the following values: .37, .31, .50, .44, .27, .65, .38, .43 and .55 (43). These values fall neatly within the earlier limits suggested by Whipple. A summary for the years 1930 - 37 shows similar values ranging from .33 to .64 (22). We can extend our inspection up to present times by quoting the following results of the use of the American Council on Education Psychological Examination: .34 (144 cases), .563 (244 cases), .48 (383 cases) and .624 (over 1000 cases) - bibliography references (27, 26, 32, 37). The conclusion,that tends to be more or less universal, is that intelligence tests have lower predictive value than the student's previous achievement ratings.

Toward the close of the second decade, significant attempts were made to predict in specific rather than general realms of college study. While intelligence tests gave a general measure of ability, there was a felt need for testing instruments which would bring out particular skills for specific types of training. The Iowa Placement Examinations (43) were designed to meet this need. In the Placement tests each subject, such as English, Mathematics, Chemistry, was represented by two examinations - one to reveal the aptitude of the student and the other to reveal his training in that subject. The prognostic value of these tests was demonstrated in an investigation at the Case School of Applied Science where three predicting instruments were employed: Army Alpha, Council of Education (Thurstone) and Iowa Placement Examinations (using all tests except ET-I). For 182 cases, the following correlations with pooled first semester grades were found: Army Alpha .491, Council of Education .623, and Iowa Placement .752, giving a margin of superiority to the pooled placement results.

Another specific phase of this movement is represented by Zyve's Stanford Scientific Aptitude Test (52). The author has reported a moderate correlation of .51 ± .07 with scholarship in science courses. In certain other studies of this test, no particularly encouraging validity coefficients have been revealed. Investigations based on samples of from 43 to 47 students show correlations of .30, .345 and .404 with science standing (3, 31). Indeed, this very test has been applied to McGill students and the results have not been any higher. As reported by Stevenson (41), the test yielded a statistically insignificant correlation of .340 with final year engineering grades for 21 cases. When fourth year average Physical Science mark was used as the criterion, the correlation was only .322 (60 students). The various investigators quoted above have concluded that the Stanford Scientific Aptitude Test is not of particular value in predicting science standing. However, such instruments as these helped to stimulate predictions in specific fields of college endeavour and this trend,from the prediction of general scholastic achievement to differential or specific prediction, has continued to gain ground.

Major Results. We have now reached the point where we can inspect the fruits of this movement as applied to the field of training in Engineering. Perhaps the least tedious way of revealing these facts is to present a systematic summary of typical findings. The studies will be reported in order of increasing merit and are intended to form a representative list, not an exhaustive one.

SUMMARY OF REPRESENTATIVE STUDIES
PREDICTING ENGINEERING SCHOLASTIC SUCCESS

(a) Tests: Thurstone Primary Mental Abilities Tests (experimental edition): abilities of number, verbal, space, induction, and deductive reasoning.

Criterion: Scholastic work for the first semester of the freshman year.

Cases: 170 freshman engineering students of the Pennsylvania State College.

Correlation: Multiple R .51

Reference: Bernreuter and Goodman (5).

(b) Tests: American Council Psychological Examination, 1939 ed.; Cooperative English Test Form OM; Cooperative Mathematics Test Form P.

Criterion: Grades during freshman year converted into point average ratio.

Cases: 383 students, North Carolina State College of Agriculture and Engineering.

Correlation: Multiple R .57

Reference: McGehee (32).

(c) Tests: Form Relations Test of the National Institute of Industrial Psychology of Great Britain; American Council Psychological Examination, 1937 ed.; Thurstone Interest Inventory, Physical Science scale; Grade XII average mark.

	Criterion:	Average mark in first year Engineering.
	Correlation:	Multiple R .66
	Reference:	Laycock and Hutcheon (27).
(d)	Tests:	Freshman Engineering; Freshman Mathematics; Freshman Mechanical Drawing; Otis Advanced Intelligence Scale.
	Criterion:	Average of honour points in all courses beyond the freshman year.
	Cases:	107 students of the College of Engineering of Oklahoma University.
	Correlation:	Multiple R .690
	Reference:	Wilson and Hodges (50).
(e)	Tests:	Percentile rank in high school graduate class; Iowa High School Content Examination, test parts on Mathematics and Science used separately; Thurstone Psychological Examination 1936 ed., test parts of Completion, Arithmetic, Artifical Language, and Analogies used separately.
	Criterion:	Average grade for first semester in Engineering College, Marquette University.
	Cases:	333 freshman students.
	Correlation:	Multiple R .719
	Reference:	Butsch (9).

(f) Tests: Iowa Placement Mathematics Training Test; American Council Psychological Examination; Purdue Placement Test in English.

Criterion: First semester grade point average.

Cases: Over 1000 freshman engineering students at Purdue University.

Correlation: Multiple R .719

Reference: Remmers and Geiger (37).

(g) Tests: Iowa Placement Mathematics Training; Cooperative Intermediate Algebra Form P; Tenth of high school graduating class; Thurstone "V" factor.

Criterion: First semester grade point average.

Cases: 244 freshman engineering students at Purdue University.

Correlation: Multiple R .791

Reference: Johnson (26).

From this summary, it is apparent that prediction with reference to the specific field of Engineering Training has been carried well into the range of high validity coefficients, that is, high as compared with the general run of predictions. With these findings in mind, we may express accord with some of the results obtained by Douglass (15) who conducted a six year study at the University of Minnesota to discover the relationship between scholastic success in schools and colleges and other variables such as test scores and previous scholastic record. Among a number of his conclusions we may quote three directly:

(1) "There was no predictive variable which was of much importance in predicting success in all schools and colleges of the University."

(2) "General tests of mental ability were of limited and of widely varying prognostic value, not of great value alone, but of importance when used with special ability tests and previous scholastic record."

(3) "The variables found especially useful in predicting success in one school or college were not found to be equally prognostic for any other school or college."

In connection with this last point, Douglass has emphasized that, while we cannot set a single level of ability for success in a university which has several faculties or schools, even greater differences in this level will be found from one institution to another. For the evidence he presented a table of scores on the American Council Psychological Examination made at various institutions: the range of median scores extended from 54.69 to 120.00!

For our own part we feel justification in asserting that the results of predictive studies in education to-day should not be generalized beyond the immediate test situation. Every university represents a somewhat unique complex of variables with reference to teaching standards, marking standards, curricula, levels of ability among the students, etc. This need not be looked upon as a pessimistic note since it is relatively easy for each university to set up its own predictive test batteries. Substantial consistency is likely to be found within the separate faculties of the university in the above-mentioned variables so that, with proper periodic adjustments, these predicting instruments will maintain their value.

(B) The Specific Factors in the Problem.

It would seem to be an appropriate step, at this point, to examine the problem of predicting success in engineering training from the analytical approach. Just what factors constitute the ingredients of this aptitude? Is it possible to measure all of these forces? What is the relationship between ability and achievement? What are the causes of scholastic deficiency in engineering training? The answers to these and other questions are admittedly vital to the success of prediction. Therefore, we shall attempt to provide a general description of the total complex of factors which may be of importance in determining success or failure for the individual student.

Several authors have listed abilities or factors which they consider essential for success in engineering or physical science training. Bennett and Cruikshank have described the following characteristics: "ability to discriminate sizes, shapes, and locations of objects in space as well as the ability to manipulate spatial, numerical and verbal symbols" (2, p.12). Bernreuter and Goodman (5), in an application of the Thurstone Tests of Primary Mental Abilities to freshman engineering students, concluded that the use of four abilities - number, verbal, induction and reasoning - seemed justified. Zyve (52), the author of the Stanford Scientific Aptitude Test, has been more elaborate but perhaps less basic in his description of the elements active in scientific aptitude. He names ten elements which may be listed briefly as: clarity of definition; suspended vs. snap judgment; experimental bent; discrimination of values in selecting and arranging experimental data; detection of fallacies and contradictions; reasoning; accuracy of systematic observations; induction, deduction and generalization; accuracy of understanding and of interpretation; caution.

Perhaps the most systematic and inclusive description of factors in engineering aptitude is given by Bingham (6). Some of the elements in his discussion have been mentioned in the material above but it will be a convenient skeleton for us to employ for the purposes of discussion. We shall deal separately with each of the six headings suggested by Bingham and evaluate these by referring to the results of previous experimentation.

(1) Aptitude for Higher Mathematics. According to Bingham, this is the most significant single indicator of aptitude for engineering studies. He suggests that one who is low on this component might be advised to by-pass professional engineering training and to study a trade in a technical school. Although his claim that Mathematics is the best single indicator is not always corroborated, it is decidely one of the most important factors. Johnson (26) used a battery of nineteen psychological and physical tests with engineering students at Purdue University and found the Iowa Placement Examination in Mathematics training to be the best single predictor of first semester grade point average, with a correlation of .70 for 242 cases. The same conclusion was reached by Remmers and Geiger (37) who obtained a correlation of .687 (over 1000 cases) for the same test, and by Wilson and Hodges (50) who used previous achievement in Mathematics as the predictor and obtained a correlation of .630 for 107 cases. These are all high coefficients for single test instruments and they point to the heavy weighting of Mathematics in the engineering curriculum.

(2) Ability to Perceive Sizes, Shapes and Relations of Objects in Space. The ability to think quickly and clearly about these relations is considered to be an asset to the student in engineering, especially for success in drafting, descriptive geometry and mechanics. Mann (30) has observed that engineering drawing grades show no particular correlation with other freshman subjects and

therefore built up an objective type of test in Engineering Drawing with material drawn largely from the field of plane and solid geometry. This test was administered to 162 freshmen and was found to be satisfactorily effective. The correlation with term grades in drawing was .626.

The value of mechanical ability tests in predicting engineering success tends to be rather limited. Bennett and Cruikshank (2) have summarized a large number of studies employing such tests in predictive efforts. If we consider only the coefficients reported for a criterion of first year grades, then we have a range of values extending from .064 to .59. The authors conclude that assembling tests utilizing simple gadgets, and tests which measure speed primarily, show little relationship with engineering success. The type of tests showing a more marked degree of correlation with the criterion are those requiring "a high degree of visualization" and a "comprehension of principles and relationships underlying the operation of mechanical devices".

A rather effective sampling of mechanical ability tests was studied by Brush who included nineteen of these tests in his battery(8). By a combination of most of these, he obtained a multiple correlation of .544 with the criterion of point average in engineering courses only, the number of cases being 77 to 160.

He felt justified in stating that "mechanical ability, as measured by several current tests, may be regarded as a component of engineering aptitude", but, at the same time, "the actual predictive power of most single tests of mechanical ability is not great". In general, they are inferior to the achievement tests. This would appear to be a reasonable view of the status of mechanical ability tests in the field.

(3) School Grades and Achievement in Physics and Chemistry. The use of high school or pre-engineering grades(and content examinations) as a basis for prediction has been placed as the best single indicator by numerous studies (8, 9, 15, 27, 39, 46). The universality of this finding is not unexpected. It is obvious that measures of past achievement take into account, automatically, many of the factors not measured by an aptitude test. In achievement marks one has a direct sample of the individual's performance under the actual conditions of study. No wonder then, that his success in the past is most indicative of his probable attainment in the future. However, the relationship here is far from perfect and most of the references already cited on this topic agree that the addition of psychological tests to the achievement batteries usually brings an increase in predictive efficiency. Comparing the three subjects, Mathematics, Physics and Chemistry, of the pre-engineering year, Hurd found that Physics had the highest single relationship with mean achievement in Engineering courses (25). The actual correlation value reported was .57 for 103 cases.

(4) Ability in English. Oddly enough, a factor that we might not suspect as being related to engineering success proves to be significant. Bingham has stated that individuals who have a restricted vocabulary and who are deficient in their language usage might suffer from a lack of verbal intelligence. Several studies have shown the presence of the verbal factor (24, 32, 37). As a specific example (5), the experimental edition of the Thurstone Primary Abilities Tests was administered to 170 freshman engineering students at Pennsylvania State College and the relationships between the various abilities and success in engineering scholastic work for the first semester produced the following correlations: Perceptual .04; Number .32; Verbal .33; Spatial .23; Memory .10; Induction .34; Deduction .38. The verbal factor ranked third and was higher than the spatial factor. A stronger relationship was found by Johnson (26) who obtained a correlation of .522 with 244 cases for the Thurstone Verbal factor and first semester engineering grades. We may conclude reasonably that this is a significant element.

(5) Interests. The most conspicuous testing device in this field is the Strong Vocational Interest Blank which measures the degree of similarity between a student's expressed interests and the interest patterns of successful men in a specific occupation. In this case, the engineering occupation is the criterion and the test attempts to measure the similarity between the student's interests and those of successful engineers. That interest is

an important factor in achievement has been demonstrated by several investigations. Herriott (23) studied a group of high achievement or "honour" engineering students and in a personal interview he asked them to give the reasons for their success. Of the reasons offered by the first 25 students interviewed, out of a total of 48, one of the two reasons ranked first was: "Interest in engineering and engineering subjects". In a multiple factor analysis of mental abilities in the freshman engineering curriculum, Sisk (38) found that there were three factors present and one was common to all courses with the exception of woodwork. He tentatively named it a study or interest factor.

The Strong Vocational Interest Blank, engineer rating, has yielded a correlation with first year engineering grades of .23 (229 cases) and, on the chemist scale, a value of .345 (244 cases) with first semester grades(26). Two other studies obtained positive but statistically non-significant coefficients of correlation (4, 24). At the University of Saskatchewan, the Physical Science Interest Scale of the Thurstone Interest Inventory gave a correlation of .26 (141 cases) with first year engineering average mark (27). Using the Strong Vocational Interest Blank, Goodman has shown that Engineering students appear to have interests more similar to successful men in such occupations as Chemist, Engineer, Production Manager, Farmer, Mathematics-Physical Science teacher and Policeman, than do Liberal Arts students (17).

These studies point to the likely importance of the interest factor but do not present very marked relationships, possibly because of the difficulty of measuring interest.

(6) Miscellaneous Factors: Bingham mentioned a series of additional forces that probably enter into the determination of success, such as: necessary health, energy, drive, and constancy of purpose. These are the most difficult elements to track down and to measure, even though we may logically recognize them as significant variables. In general, a student of high achievement and high aptitude is likely to have satisfactory motivation, although there are exceptions. It is conceivable that his ability is great enough to obscure the effects of relatively poor drive. A student of high achievement and comparatively low aptitude must have a great deal of the qualities of perseverance and motivation. In the case of low achievement the most probable cause is lack of sufficient ability; however, as is frequently demonstrated, the student may have the ability but lack the drive. Gerberich has examined some of the factors which affect the college achievement of high aptitude students who do not measure up to expectation and the low aptitude students who exceed expectation (16). Although few of the differences found between these groups were statistically significant, there was a consistency of differences

in certain directions which the author considered suggestive. On this basis, "the low-scholarship groups were consistently found to have greater difficulty in paying attention in class, to have greater difficulty in using study techniques, to be fraternity or sorority pledges more frequently, to participate more often in freshman foot-ball and intramural athletics, and to recognize a greater degree of difference in difficulty between high school and university work than was true for the high-scholarship groups. The high-scholarship groups, on the other hand, more frequently than the low-scholarship groups, handed themes in promptly, crammed for examinations, had a good place to study, used the library for study, participated in freshman basketball, liked most of their classes and instructors, and felt that they got a 'square deal' in their classes". As far as time expenditure was concerned, "the high-scholarship students consistently spent more time in sleep, at meals, in the classroom, and studying, and less time at the movies and in optional reading than did the low-scholarship groups". All of which would seem to point conspicuously to the general factors of motivation and study habits. Indeed, Ohmann has ranked "motivation and interests" along with "intellectual factors" as the two most frequent major causes of scholastic deficiency (35).

The relationship between ability and achievement is seen thus to be a complicated one. While it may be difficult to measure the part played by the factor of motivation in educational achievement (12), its importance cannot be denied. At least one conclusion is justified: motivation is a variable which goes unmeasured in most of the attempts of the psychologist to predict scholastic success and thus accounts for some loss in predictive efficiency.

The factor of personality has not been found to be particularly important. From a review of the literature, Harris has observed that personality tests in general show low correlations with grades, while maladjustments show conflicting results (22). By illustration, we have the work of Griffiths who attempted to ascertain the relationship between scholastic achievement and personality adjustment, utilizing the Bell Adjustment Inventory (20). He concluded that students with very high scholastic records are no better adjusted in personality than men of lowest academic achievement. The differences found did suggest some degree of positive correlation between scholastic achievement and personality but none was statistically significant. In a study of engineering aptitude the Bernreuter Personality Inventory was used as one of the testing instruments but it yielded nothing of significance in correlations with first year marks (27). The use of the Allport A-S Reaction Study in a different investigation did reveal a statistically significant relationship (24). In this case a negative correlation, -.404, was found between the A-S scores and college grades. This indicated a greater submissiveness on the part of the best students. Poorer students tended toward the ascendance side, especially those having a high entrance psychological score.

Darley set up the hypothesis that if measured maladjustment or radicalism affected academic achievement adversely, then severe cases should be working below their capacities and hence would show a lower correlation between achievement and ability than would be found among students free from maladjustment. In experimentation, the author was able to substantiate his hypothesis for female students but was not able to do so for the men. One explanation for the latter failure was found in the fact that the group of maladjusted men had received a greater number of counselling interviews than had the maladjusted women. As a result, the achievement of these men might have been brought more in line with their capacity.

Our last consideration will be given to study habits. In Herriott's study of "honour" engineering students (23), these high achievement students gave their reasons for personal success. The two reasons ranked in first place were: "Interest in engineering and engineering subjects" (which we have already mentioned) and "Regularity of daily preparation, including persistent effort and 'hard work'". In second place were: "Good study conditions, including physical conditions, influence of roommates, general atmosphere and the like" and "High School work had much in common with the work in the College of Engineering". Third place was held by "Concentrated effort when studying". Ohmann found that, of the study habit problems listed as causes of scholastic deficiency, the single item receiving greatest weight was: "No study habits formed in high schools" (35). Finally, in an investigation of 229 engineering students, the Wrenn Study Habits Inventory revealed a correlation of .28 with first year point average grade (26).

It is apparent from this sampling of the literature that study habits are to be considered important in the determination of scholastic success.

<u>Conclusion</u>. We have examined the many factors that enter into this very complex problem of predicting success in engineering training. While certain distinct abilities are fundamental to this success, we have observed that additional factors operate to disturb the relationship between achievement and ability. Indeed, strong agreement can be expressed with the principle of a hypothesis set forth by Thornton in this very connection (45). He would seek to gather together all the significant factors, other than scholastic ability, under the title of "scholastic efficiency". Having a measure of scholastic efficiency, one could devise a formula for predicting achievement in college by making use of an index of scholastic efficiency $\frac{A}{B}$, where A represents average T score for grades in a given school and B represents T score for scholastic ability in comparison with other students in the same school. This procedure might be capable of anticipating a discrepancy between later attainment and ability. Another aid to the efficiency of educational prediction has been described by Williamson (49). This author has suggested that the standard regression equation might be supplemented by means of individual clinical diagnosis and counselling.

In this way, factors which operate only in the cases of particular students, and for which no allowance has been made in the regression equation, may be discovered. If readjustment could be accomplished for these students, then their achievement would be more nearly commensurate with their ability. It is undoubtedly true that such schemes as these would aid in the raising of predictive efficiency.

We are now in a position to consider the present study and its attempt to predict engineering aptitude. The results may be evaluated in the light of the discussion in this Chapter, taking into account the fact that prediction is based upon two main factors: a psychological test of engineering aptitude and the records of pre-engineering scholastic attainment.

CHAPTER III

THE TEST AND PROCEDURE EMPLOYED

(A) A Description of the Engineering and Physical Science Aptitude Test

The Engineering and Physical Science Aptitude Test is a psychological measuring instrument which owes its origin to the emergency of the Second World War. In the interests of the national defense effort, United States officials found it necessary to train large numbers of men for technical work (19). To accomplish this task, advanced educational facilities were required and so the colleges and universities had to assume much of the burden. The federal program involved was known as Engineering, Science and Management Defense Training. The training of these men,effectively, called for some psychological test which would evaluate the degree of aptitude that the trainee might have for engineering and physical science studies. In a survey of the field of existing tests, it was concluded that these tests measured but relatively narrow phases of the aptitude in question whereas a number of abilities were considered to be active in the determination of success in engineering training. To provide such an instrument, research workers at Pennsylvania State College constructed a battery of tests which would measure the aptitude of an individual to handle successfully courses of instruction in engineering or physical science subjects. The finished product was called the Engineering and Physical Science Aptitude Test and was compiled under the direction of Bruce V. Moore, C. J. Lapp, and Charles H. Griffin (34). A manual of directions was prepared by Henry Borow (7). The test was first administered widely in March 1942 to persons commencing technical training as a part of the Engineering, Science, and Management Defense Training program.

(N.B. Inasmuch as we will have occasion to use the name "Engineering and Physical Science Aptitude Test" most frequently in the remaining discussions, it will be convenient and appropriate to abbreviate this title and hereafter to employ the symbol "EPSAT".)

The EPSAT was built up on the basis of existing standardized tests. Concerning this development the following quotation is pertinent:

> "The selection of the test parts was based on a physics criterion while validation of the final battery was determined with reference to attainment in a series of technical subjects. Male high school graduates

> enrolled in the Introductory Engineering Subjects course of the Pennsylvania State College's war training program were given an experimental battery of tests. The average age of this group was approximately 18 and the average intelligence level as measured by the Otis S-A Test of Mental Ability was somewhat superior to that of college freshmen as found by Otis. The population samples used in the analysis of the several tests ranged in size from approximately 300 to 600 subjects. Each test was studied in relation to scores on the Cooperative Physics Test for College Students: Mechanics, Form 1936 B, which was administered to I.E.S. students at the close of their physics course." (19).

With this criterion in use, three parts of the Revised Iowa Physics Aptitude Test were selected to form the Mathematics, Formulation, and Physical Science Comprehension test parts of the EPSAT. The Arithmetic Reasoning part and the Mechancial Comprehension test part of the EPSAT were developed from the Moore Test of Arithmetic Reasoning and the Bennett Test of Mechanical Comprehension, Form AA, respectively. Finally, the Verbal Comprehension test part of the EPSAT was composed of technical terms selected from the vocabulary section of the Moore-Nell Examination for Admission used at Pennsylvania State College. This last-named part was included because of the discovery that many highly verbal items of the Otis S-A Test of Mental Ability were substantially related to the criterion. The test items and directions are printed in booklet form while the subjects record their answers on special answer sheets by filling in the space between dotted lines under the letter of their choice of correct answer. These may be scored either by hand or by machine. Since the booklets are not marked in the administration of the test, they may be re-used.

The Test Parts. In the following brief description of the kind of material and directions used in each of the six test parts, no item from the test proper will be included. This step is taken to maintain the validity of the test should it be used as a part of any testing program at McGill.

Part I: Mathematics

Time allowance: 15 minutes; 25 items.

Directions: The exercises in this test part represent commonly used arithmetical and algebraic skills in first year college physics. Solve each of the problems, find the answer among the five choices, and record your answer by filling in the space between the dotted lines under the letter of the correct choice on the answer sheet.

Samples: X. $3x = 15$; what does x equal?
(A) 2 (B) 3 (C) 5 (D) 7 (E) 10
Y. $4a + 7 = 35$; what does a equal?
(A) 3 (B) 4 (C) 5 (D) 6 (E) 7

Part II: Formulation

Time allowance: 10 minutes; 10 items.

Directions: In this part you are to read each statement or short paragraph and do what it tells you to do. In most cases this involves writing an algebraic expression for what the statement says. As in Part I, select the correct answer from among the five choices and record your answer by filling in the space between the dotted lines under the letter of the correct choice on the answer sheet. Study the samples.

Samples: X. If X is a number, twice that number would be expressed algebraically as:

(A) x (B) x^2 (C) $2x$ (D) $2x^2$ (E) $4x$

Y. If x and y represent two numbers, their sum would be expressed as:

(A) $x + y$ (B) xy (C) $x-y$ (D) x^2y^2 (E) $\frac{x}{y}$

Part III: Physical Science Comprehension

Time allowance: 10 minutes; 45 items.

Directions: Examine each statement below and decide whether it is true or false. If the statement is true, fill in the space under T on the answer sheet. If the statement is false, fill in the space under F.

Samples: X. Iron is a metal.

Y. Water is a solid.

Part IV: Arithmetic Reasoning

Time allowance: 15 minutes; 10 items.

Directions: Solve these problems. Use the extra sheet of paper provided for any figuring you need to do. Do not write on the booklet. Do not spend too much time on any one problem. Work rapidly but be accurate. When you have solved each problem, find the alternate which agrees with your choice and fill in its space on the answer sheet.

Actually,no sample is provided in this test part but one may easily be invented for the sake of illustration.

Sample: If 3 pencils cost 18 cents, what is the cost of 5 pencils?

(A) 25 (B) 21 (C) 30 (D) 24 (E) 35

Part V: Verbal Comprehension

Time allowance: 10 minutes; 43 items.

Directions: On the next two pages are forty-three words in CAPITAL LETTERS. After each word there are five choices of other words, one of which is most nearly equivalent in meaning to the word in capital letters. Select the most appropriate answer and fill in the space under its letter on the answer sheet.

Samples: X. FLOOD (A) denude (B) deluge (C) buried (D) delude (E) destroy

Y. IRON (A) animal (B) vegetable (C) social (D) metal (E) religious

Part VI: Mechanical Comprehension

Time allowance: 12 minutes; 22 items.

Directions: Look at Sample X on this page (i.e. of the test). It shows pictures of two rooms and asks, "Which room has more of an echo?" Room "A" has more of an echo because it has no rug or curtains, so the space under A is filled in on the answer sheet.

Sample Y pictures 2 cutting shears: A has long cutting jaws and short handles while B has very short cutting jaws and long handles. The question is: "Which would be the better shears for cutting metal?"

On the following pages of the EPSAT there are more pictures and questions to which the subject must find the correct answer.

Total working time on the test is 72 minutes. It is estimated that from 80 - 90 minutes should be ample time for the administration of the entire test. Superficially at least, we can observe that these 6 test parts of the EPSAT cover several of the elements considered basic to engineering aptitude, such as aptitude for mathematics, a verbal factor, and a comprehension of relationships in physical science and in mechanics.

Interrelationships. Concerning the interrelationship of the 6 test parts, one seeks evidence that these parts actually measure different abilities or phases of engineering and physical science aptitude. The conventional check on this feature is to look for low coefficients of intercorrelation. As far as the EPSAT is concerned, the authors claim that the expectation of relatively low interrelationships among the test parts "appears to be moderately well founded". The median coefficient of the test part intercorrelations is .460. The highest degree of interrelationships occurs between Mathematics and Formulation with a coefficient of correlation of .609. Each of these subtests emphasizes mathematical content. The lowest degree of interrelationship is found between Verbal Comprehension and Mechanical Comprehension with an intercorrelation of .280. The values of these correlations were computed from the test scores of 188 high school graduates receiving war training in Introductory Engineering Subjects.

Validity. In judging the validity of any test, we wish to know how well the test measures what it is supposed to measure. The authors make it clear that the EPSAT has been designed basically as a measure of training aptitude, or the aptitude which fits an individual to profit from training in engineering and physical science studies. "As such, its validity must rest first upon its effectiveness in segregating those persons who are likely to progress rapidly in technical training from those who are likely to falter during such training. A second and more critical demonstration of the test's validity concerns its discriminating power in industrial situations." (19). As yet, information is only available with reference to the first, most basic form of validation. Consequently, the criterion, in the light of which the EPSAT must be measured, will be the grades which students obtain in engineering subjects. If the test shows a high relationship with attainment as measured by these grades, then it is a valid test of training aptitude in this field. To obtain these facts, the EPSAT was administered to 188 Pennsylvania high school graduates, most of whom were males. They were enrolled in the 1942 Summer Course in Introductory Engineering Subjects. The nine-week course included the following subjects: engineering mathematics, engineering chemistry, engineering physics, engineering drafting, and manufacturing processes. Of interest to us at this point are the correlations found between scores on the EPSAT and Final Average grade in the Introductory Engineering Subjects. These are reproduced below:

	Average Grade
EPSAT (total score)	.733
Part 1: Mathematics	.626
Part 2: Formulation	.587
Part 3: Physical Science Comprehension	.561
Part 4: Arithmetic Reasoning	.522
Part 5: Verbal Comprehension	.542
Part 6: Mechanical Comprehension	.485

The correlation of .733, between the composite test and final average grade, points to a high validity coefficient for the EPSAT. When the 6 subtests are weighted by multiple regression technique, this correlation is raised to .763 with the mathematics test part possessing the highest weight. It is worth while noting that in considering the five engineering subjects separately, the correlations between the total test score and the subjects of Manufacturing Processes and Drafting were by far the lowest at .376 and .346 respectively. Concerning the low correlation with Drafting the authors point out that the EPSAT does not sample "certain visualization and motor skills" which are important for success in Drafting.

Reliability. The reliability of a test is determined in answer to the question: "Does the test measure whatever it is supposed to measure in a consistent fashion?" In other words, if an equivalent form of a test were to be administered to a group of individuals, the same relative ranks on the score scale should be maintained by these individuals if the test is reliable. The correlation which expresses this tendency is called a reliability coefficient and for the composite EPSAT has been computed by the split-half method (having only one form of the test) and corrected by the Spearman-Brown prophecy formula. The reliability coefficient thus reported is .96.

Standardization. Percentile norms for both sexes have been established using large groups of men and women who began their training in engineering and science courses in the spring of 1942. According to the authors, a "reasonably accurate" description of the characteristics of these persons includes the following information: Average age approximately 30; median salary between 30 and 40 dollars per week; 85% completed high school and a substantial proportion proceeded further; approximately one-half held professional, skilled, and clerical positions while about 10% of the employed were classified as unskilled workers. The group was considered to be "somewhat superior educationally, economically, and, in all probability, mentally, to a general, unselected population." (7). Since the present study deals only with the male sex, we will reproduce the percentile norms for men based on 6,695 cases:

EPSAT Norms for Men (N = 6,695)

Score	Percentile	Score	Percentile
138	99	57	50
120	95	52	45
105	90	48	40
95	85	44	35
88	80	40	30
82	75	35	25
76	70	31	20
70	65	26	15
66	60	21	10
61	55	13	5
		3	1

A full comparison of the relationship between these norms and the sample of McGill students studied will be found in Chapter IV.

(B) The Details of the Procedure

Obtaining Data. At the beginning of the fall session 1944, the EPSAT was administered to 141 students who were entering the first year of Engineering. This number represents approximately 86% of the total class of that year. These students had completed either the first year of Science at college or they had taken senior matriculation courses in the high school. In itself, this stands for a level of education advanced one year beyond the high school graduate status. However, even further educational selection operated in the choice of these students since they had to satisfy the entrance requirements set by the Faculty of Engineering. Involved in this selection are certain definite prerequisites. First of all, according to the official regulations for 1944, any student desiring to enter Engineering from the first year of the B. Sc. Course must obtain a passing mark (50%) in his examinations in Mathematics, Physics and Chemistry. A student who has failed in one subject only (other than Mathematics, Physics and Chemistry) may be admitted with a condition in that subject if his average standing in Mathematics, Physics and Chemistry is at least Second Class, i.e. 65 - 79%. A student who has failed in more than one subject is not admitted. It is evident, from these requirements, that a stronger factor of academic selection functions in the choice of this group of students than is active in the general Faculty of Arts and Science.

As concerns the total group of 141 students who were given the EPSAT, all but three were males. It may be stated therefore, that sex is not a variable factor in the present study.

All of the completed answer sheets for the EPSAT were scored by hand and later re-scored to insure accuracy before any computations were undertaken.

In the spring of 1945 this class of first year engineering students wrote their final examinations and, with the kind permission of Dean J. J. O'Neill, the results of these examinations were made available. For each student who had taken the EPSAT in the fall, marks on the following 12 courses of the first year of Engineering were recorded:

	Maximum Mark
1. Elementary Physical Chemistry and Laboratory	125
2. Engineering Problems	50
3. Spherical Trigonometry and Mensuration	50
4. Algebra	100
5. Calculus	100
6. Mechanics	100
7. Analytical Geometry	50
8. Mechanical Drawing	50
9. Descriptive Geometry	150
10. Surveying	50
11. Physics	100
12. Physics Laboratory	50

The heaviest weighting among the above subjects occurs in Mathematics with almost one-half of the total marks assigned to this general subject.

During the academic year, fifteen students withdrew from the course for various reasons, chief among which was enlistment in the Armed Forces. This left a total of 126 students for whom records of both achievement and aptitude test scores were available. Such data form the basis for evaluating the relationship between the aptitude test and academic attainment in the first year of Engineering. The total average percent standing in all subjects of the first year was taken as the criterion which we are seeking to predict. This meaning of the term "criterion" will be employed consistently with reference to the results of the present study.

The next step in the procedure consisted of the accumulation of information with respect to the student's academic standing in the year prior to entering Engineering, so that the relationship between pre-engineering scholastic success and achievement in the criterion could be computed. One difficulty that was encountered in this part of the procedure involved the fact that these students had been trained in several different institutions and educational systems before commencing the Engineering course. This threatened to contribute two sources of error to the study: in the first place, variations in the methods of reporting grades and secondly, variations in the standards of marking. The first source of error becomes most serious when institutions report grades in terms of very course units of measurement, for example, Class A, B or C; Class I, II, III; V.G., G., passed, etc. For the McGill students and Quebec senior matriculation students (who, fortunately, accounted for a large majority of the total) academic records revealed the exact percent standing of each individual in each course taken. Such a form of finer discrimination between marks is required since one must employ relatively exact quantitative measurements in an analysis such as this. Therefore, it was decided to eliminate from the investigation any students for whom percent gradings were not available. This act was justified inasmuch as any arbitrary method of assigning values to the crude letter grades would represent low reliability and extremely poor discrimination between individuals.

The second source of error mentioned above, namely, variations in the standards of marking, could not be controlled. Fortunately, it is considered that these variations are not large enough to be serious in the present study, since, after the elimination of the students for whom only crude grades were at hand, a fairly homogeneous group remained, consisting largely of senior matriculants from the high schools of the Province of Quebec and McGill first year B. Sc. students. All in all, allowing for a few other individuals whose past records were incomplete, 103 students were included for whom the following information was recorded: EPSAT scores, first year engineering marks, and pre-engineering marks.

A few words concerning an additional problem in dealing with pre-engineering grades is in order. A survey of the curricula of these students revealed the fact that a wide variety of subjects had been taken. In order to obtain a stable basis for computing pre-engineering standing, only those courses of study common to all students were selected. The four subjects satisfying this requirement were Chemistry, English, Mathematics and Physics. The weighting of these four subjects also varied with the institution, and, therefore, to preserve uniformity, a total average percent grade for the pre-engineering year was based upon an average of the four courses weighted equally. In subsequent discussion of the results of this study, wherever first year Science average standing is mentioned, the above meaning will be implied, namely, average percent standing in Chemistry, English, Mathematics and Physics weighted equally.

Statistical Methods. With the accumulation of the above data completed, statistical analysis was next undertaken. While the results will be presented in detail in the following chapter, a preliminary word is appropriate. The study and its efforts to predict academic attainment in the first year of Engineering has been based mainly upon simple correlation and multiple correlation techniques (21, 36). The method of computing the coefficients of correlation utilized the Durost-

Walker Correlation Chart published by the World Book Company. This Chart facilitates the computation of a Pearson product-moment coefficient of correlation and has a very important feature which allows for checks on the arithmetical processes involved. These occur at frequent intervals in the course of the computations. The checking procedure is carried through to the end of the work so that two independent derivations of the coefficient are possible. Thus the accuracy of the calculations is assured. In all cases, the use of the term "Mean" shall refer to the Arithmetic Mean, the symbol "N" will indicate number of cases, and "SD" the standard deviation.

CHAPTER IV

THE RESULTS.

The plan of this chapter involves two distinct steps in presentation. (a) We shall first examine the EPSAT scores and the scholastic achievement grades,basing our observations on means, standard deviations and percentiles, and drawing comparisons with other data where possible. (b) By means of correlation technique, the efficiency of predicting first year engineering attainment will be analysed. A practical view of the question of prediction will be emphasized at the end of this section.

(A) General Statistical Analysis.

EPSAT Scores. In Table 1,below, are shown the means and standard deviations of the six test parts and the composite scores of the EPSAT. Along with this information is included an additional set of comparison means and standard deviations which are based on the results of 188 high school graduates enrolled in the Introductory Engineering Subjects. These latter values have been reported in the Manual of Directions for EPSAT (7).

Table I.

EPSAT SCORES.

Test	McGill:1st Yr.Engineering			Comparison Values (see Text).		
	N	Mean	SD	N	Mean	SD
Part I Mathematics	126	22.44	2.13	188	12.99	6.02
Part II Formulation	126	8.92	1.25	188	4.98	2.56
Part III Phys.Sc.Comp.	126	25.44	9.44	188	13.67	9.24
Part IV Arith.Reason.	126	7.44	2.00	188	4.12	1.92
Part V Verbal Comp.	126	29.31	6.93	188	19.12	8.33
Part VI Mechan.Comp.	126	15.83	5.13	188	10.96	5.14
Composite Test	141	109.08	20.36	188	65.84	25.92
	126	109.98	20.77	-	-	-

Before making any interpretative statements,it must be emphasized that the comparison values used are based on the results of high school graduates whereas the McGill students represent an additional year of educational selection. The difference between these two groups in mean value for each test part and the composite test is shown to favor the McGill students consistently by a wide margin. Although it did not seem to be necessary, all of these differences were tested for statistical significance and were found to be extremely significant,that is, none of these differences could have arisen by chance. Under such circumstances,it is obvious that some bias or force is operating in the selecting of the McGill students. This factor is not hard to find. The superiority of the McGill group is to be taken as a demonstration of the effectiveness of one more year of educational selection,plus the selection forces resident in the entrance requirements set by the Faculty of Engineering. Put in another way, we are in a position to conclude that, as a result of one year of educational training past the high school graduate level, the McGill students that are able to meet all academic requirements for entrance to Engineering are significantly superior to high school graduates enrolled in the Introductory Engineering Subjects at Pennsylvania State College. Our comparison as above, must, of course, be limited to the question of selection. We cannot proceed further to any justifiable conclusions as to the basic abilities of Pennsylvania and McGill students since they do not represent equivalent educational levels and are thus affected differently by the variable of selection.

An inspection of the standard deviations for the two groups yields further evidence of the factor of selection which functions strongly at McGill. In the subtests on Mathematics, Formulation, Verbal Comprehension and on the composite test, the McGill distributions show less variability, that is, in terms of smaller standard deviations. Subtests on Arithmetic Reasoning and Mechanical Comprehension show but very small differences either way while in the Physical Science Comprehension test part the difference tends to be a little larger with the McGill group showing the higher value of standard deviation. This is due to the peculiarity of the distribution of McGill scores,as may be seen later.

For the composite test the McGill,average value is strikingly high. It will be noticed that two sets of mean and standard deviation are given for two different values of N. The mean of 109.08 and SD of 20.36 are given for all of the students who took the EPSAT at the beginning of the fall session (N is 141). During the session, fifteen dropped out so that the highest number that could be used in the computation of correlations was reduced to 126. It is interesting to observe that the fifteen who left the course tended to have a depressant effect upon the mean value of the group since this value rose to 109.98 when their scores were removed.

No further observations can be made effectively until the actual frequency distributions of the McGill scores on the EPSAT are revealed. These trends are displayed in the following Figures 1 to 7. The abscissa values are the lower score limits of each class interval wherever these intervals contain more than one score unit. For example, in the case of Figure 5,the first interval on the abscissa reads 11 to 14 and includes all scores of 11 or more, but less than 14. When the class intervals contain only one score unit, then the midpoint score value of each interval appears on the abscissa as in Figure 1.

Fig. 1. Distribution of Scores on EPSAT Test Part I, Mathematics, N = 126

Maximum possible range: -6 to 25

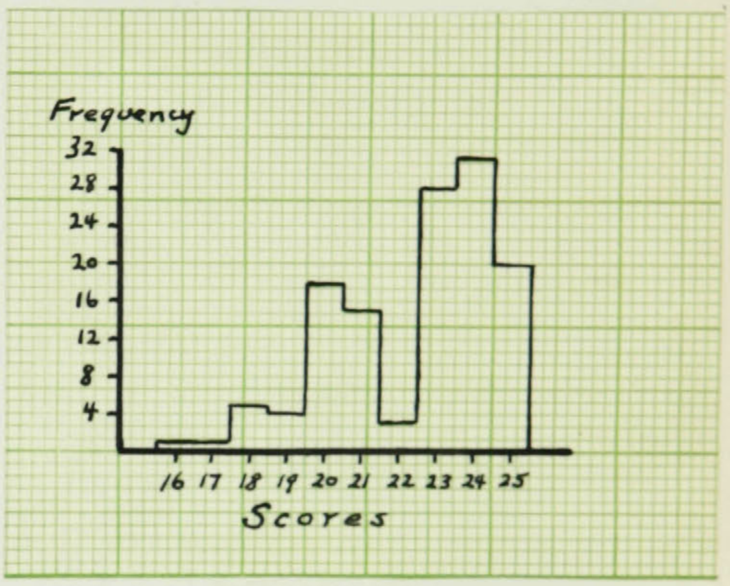


Fig. 2. Distribution of Scores on EPSAT Test Part II, Formulation, N = 126.

Maximum possible range: -2 to 10

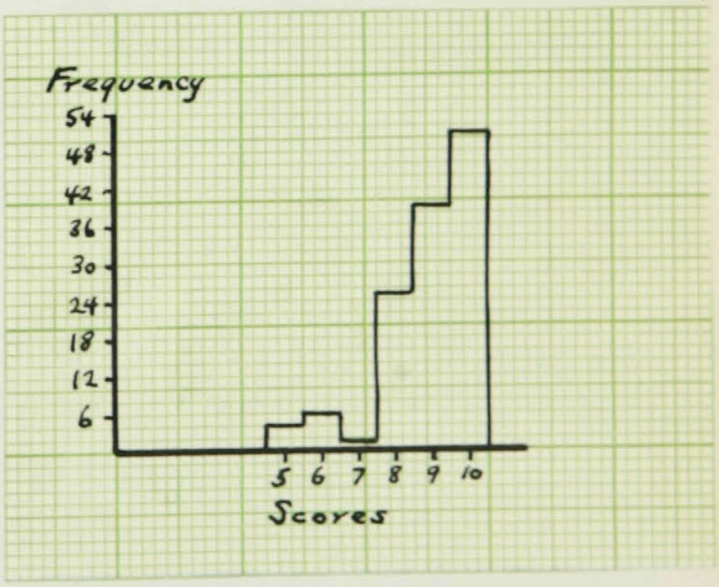


Fig.3. Distribution of Scores on EPSAT Test Part III, Physical Science Comprehension: N = 126. Maximum possible range: -45 to 45

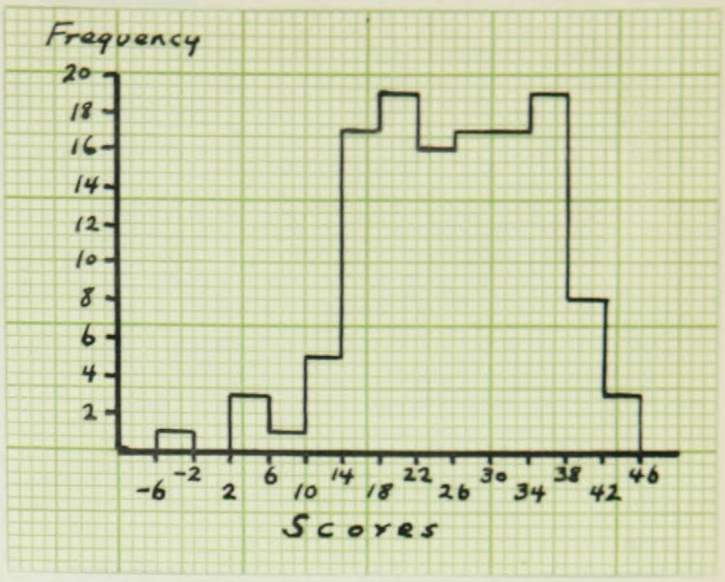


Fig. 4. Distribution of Scores on EPSAT Test Part IV, Arithmetic Reasoning: N = 126 Maximum possible range: -2 to 10.

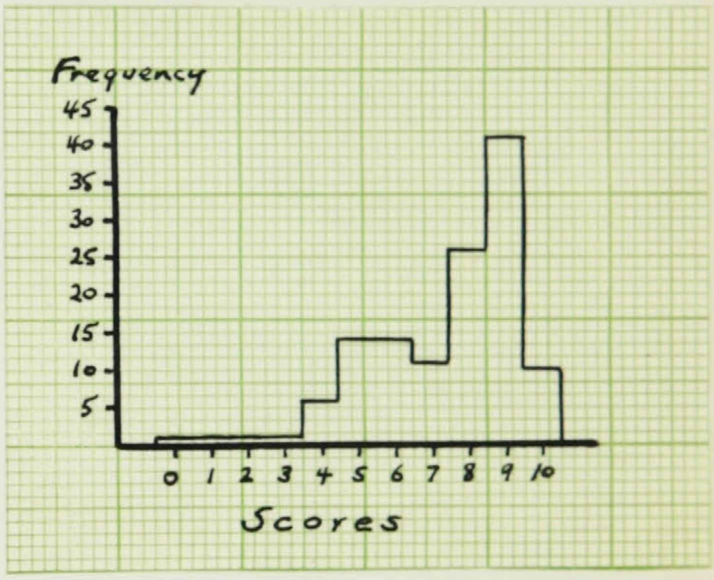


Fig.5. Distribution of Scores on EPSAT Test Part V, Verbal Comprehension: N = 126. Maximum possible range: -11 to 43.

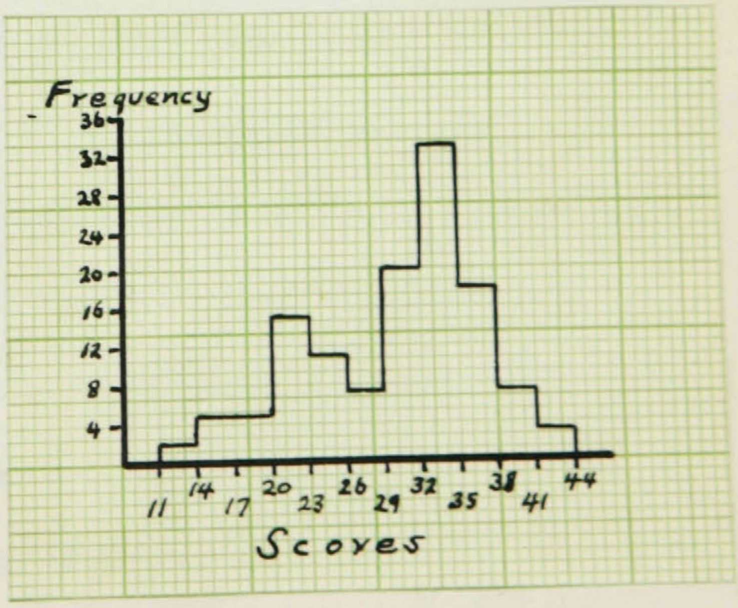


Fig. 6. Distribution of Scores on EPSAT Test Part VI, Mechanical Comprehension, N = 126.

Maximum possible range: -11 to 22

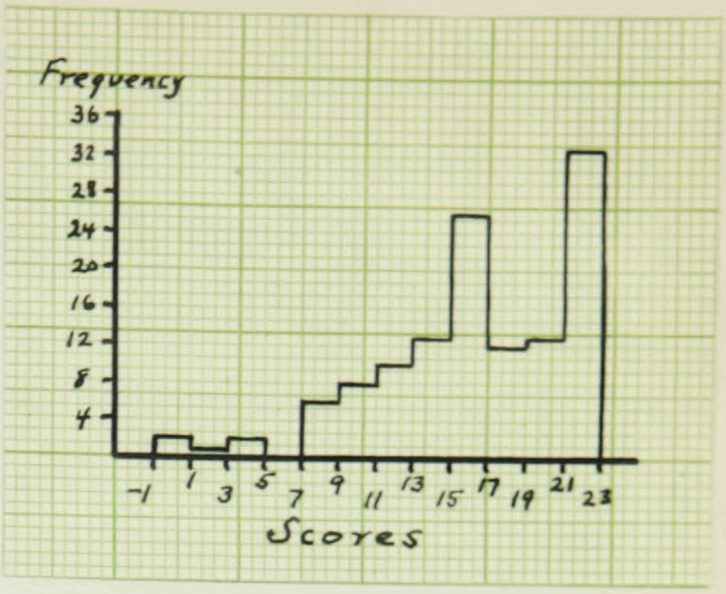
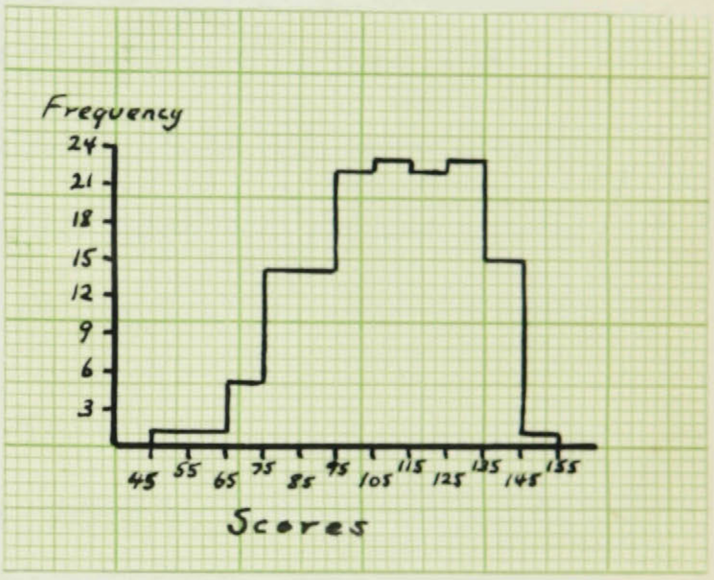


Fig. 7. Distribution of Scores on EPSAT composite test, N = 141.

Maximum possible range: -77 to 155



From these distributions, two consistent qualities are apparent. Firstly, all distributions show some degree of negative skewness, that is, the bulk of the cases tends to fall near the high end of the scale with relatively fewer cases occurring near the low end. This condition is most pronounced in Figures 1, 2,4 and 6. The second common quality is that these distributions occupy but a part of the maximum possible range in each case. Evidence to this effect is readily available by comparing the maximum possible score range given beside each figure with the actual range of scores as plotted.

Of all the test parts, Part II, Formulation, is the weakest element inasmuch as it discriminates least well among the students. Only about half of the possible range is covered with approximately 91% scoring either 8,9 or 10 points. Part I, Mathematics, shows use of about one-third of the possible range but, because of the larger number of items, for one thing, it tends to discriminate slightly better than Part II. As for the other test parts, we may conclude that discrimination has been fairly good even though only a part of each range has been covered. This condition of relatively high cut-off value for the occurrence of low scores is to be expected since we are dealing with a highly selected group.

In considering the composite test scores on EPSAT for the entire McGill group (N = 141), a most revealing form of comparison can be made by translating the raw scores obtained by the McGill students into percentile values, as given in the Manual of Directions and shown earlier in this study in Chapter III. This comparison is indicated graphically in Figure 8 below.

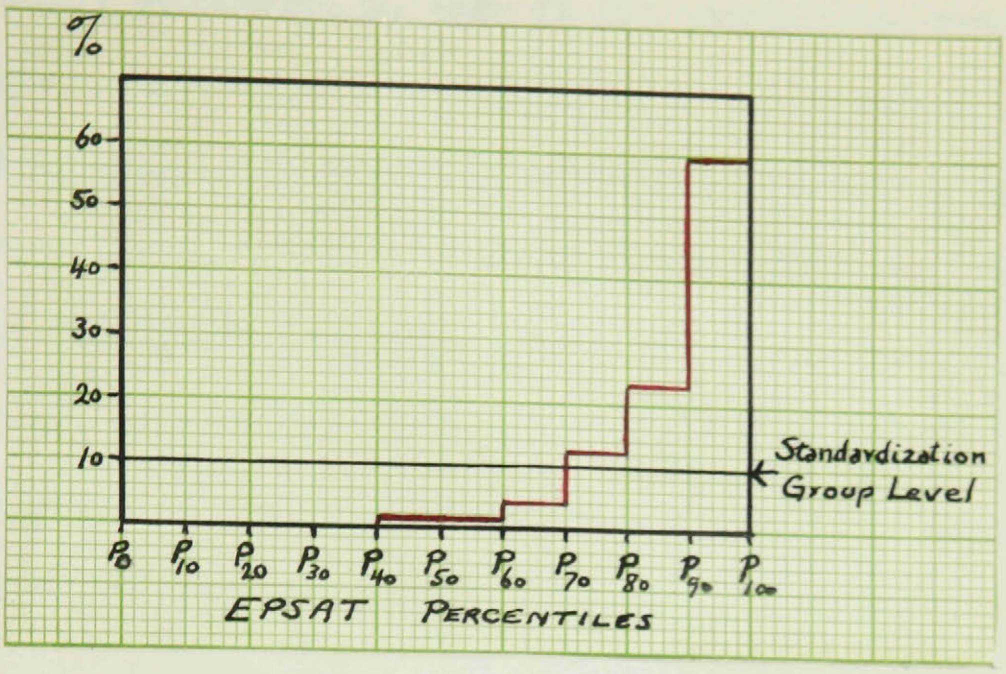


Fig. 8 McGill Scores compared with EPSAT Percentile Norms.

Figure 8 should be read as follows. Abscissa units mark off the various deciles (P_{10} to P_{20}, etc.) based on the EPSAT norms. On the ordinate is shown the percentage of the McGill group whose percentile scores fall within a specified decile. The line cutting the graph at the ordinate value of 10%, and designated "Standardization Group Level", represents the percentage of students who would fall within each decile if that particular group were identically constituted to the EPSAT standardization group. Thus a deviation from this 10% level will indicate a difference of constitution if marked enough. From Figure 8, it will be apparent that the deviation of the McGill group from the EPSAT standarization population is most significant. It should be recalled that the EPSAT population consisted of 6695 men who, for the most part, had completed high school training and were judged to be "somewhat superior educationally, economically, and, in all probability, mentally, to a general, unselected population" (7). This being true, the McGill group would differ even more markedly from a random sample of the general population. In Figure 8, 59.6% of the McGill first year Engineering students obtain EPSAT scores which are equivalent to percentiles of 90 and above. Thus 59.6% of these students obtain scores which equal or exceed 90% of the EPSAT population. Similarly, we have the remaining readings:

22.7%	obtain scores	from	P_{80}	-	P_{89}
12.1%	"	"	"	P_{70}	-	P_{79}
4.2%	"	"	"	P_{60}	-	P_{69}
0.7%	"	"	"	P_{50}	-	P_{59}
0.7%	"	"	"	P_{40}	-	P_{49}

No McGill student obtains a percentile score below 40 while only two get scores below the 60th percentile. A further breakdown of the scores for students falling within the percentiles 90 to 100 reveals that 36.2% obtain scores between P_{95} to P_{100} with the remainder, 23.4%, falling within the percentiles 90 to 94.

Before ending our discussion of Figure 8, one last observation is in order. If the EPSAT does measure engineering aptitude, then the rapid drop in the trend of EPSAT scores, from very high percentile ranks to the average ranks, points to the relative effectiveness of the present method of selecting students. The sharp cut-off of students who get scores below P_{90} in the aptitude test shows that the present selection instruments do differentiate between students of low aptitude and those of high aptitude. This is accomplished on the basis of past academic achievement.

Since the McGill Engineering students differ so markedly from the standardization population used in EPSAT, it will be of practical value to present a set of percentile norms based on the 141 cases sampled at McGill. All but three of this group are males. Such a set of norms will be of value for any testing and counselling of McGill students with respect to Engineering.

Table 2.

McGILL PERCENTILE NORMS : FIRST YEAR ENGINEERING N = 141

Score	Percentile	Score	Percentile.
144	99	110	50
140	95	107	45
136	90	104	40
132	85	101	35
129	80	98	30
126	75	95	25
123	70	90	20
120	65	85	15
117	60	80	10
113	55	75	5
		59	1

Any percentile read from Table 2 refers to the percentage of the first year Engineering students who fall below a given score value on the EPSAT.

Achievement Scores. Consideration must now be given to the scholastic standing of the McGill students both in their pre-engineering year (Chemistry,English,Mathematics and Physics) and in their first year of Engineering (total average percentage). Table 3, below, shows these values.

Table 3.

ACHIEVEMENT SCORES.

Subject	N	Mean	Standard Deviation
Chemistry Pre-Engineering	103	71.71	12.96
English Pre-Engineering	103	65.30	10.53
Mathematics Pre-Engineering	103	69.33	12.79
Physics Pre-Engineering	103	72.05	12.67
Total Average % of 4 Pre-Engineering subjects.	103	69.54	9.38
Total Average % of First Year Engineering	126	60.49	11.94

The lowest subject, of the four considered, is English, the reason being that a passing mark in this subject is not compulsory for entrance to Engineering while it is essential in the other three subjects. However, the difference is not as great as might have been expected. Another interesting feature is the decrease in average class standing from the Pre-Engineering year to the first year of Engineering. This decrease is a substantial one of 9.05 points and is but slightly altered when the two groups are made equal with respect to the total number of cases. The explanation most probable is a higher standard of marking maintained in the first year of Engineering. Unfortunately, no comparison values were available and yet it seems reasonable to assume that the average standing in at least Chemistry, Mathematics and Physics will be above the average standing for all first year college students.

In Figures 9 to 14 are shown the frequency distributions of the six elements of Table 3. The same conventions hold as in the case of Figures 1 to 7.

Fig 9. Distribution of Chemistry % Marks for Pre-Engineering year of Science. N = 103

Fig. 10. Distribution of English % Marks for Pre-Engineering year of Science.

N = 103

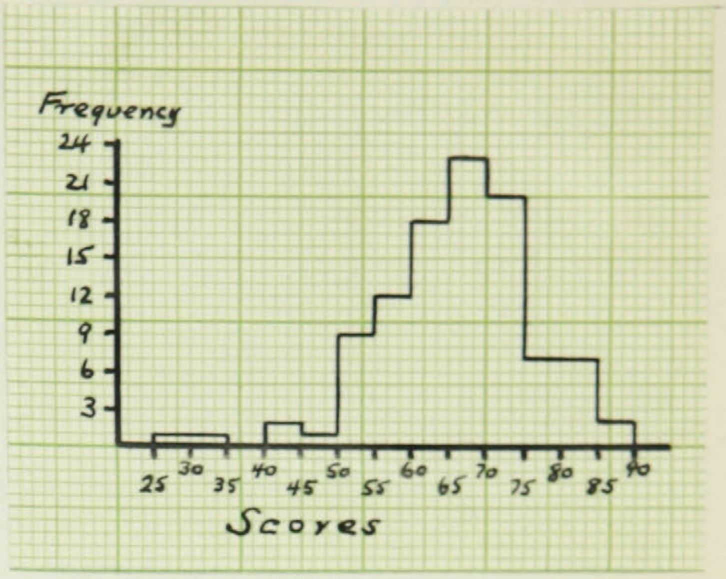


Fig. 11. Distribution of Mathematics % Marks for Pre-Engineering Year of Science.

N = 103.

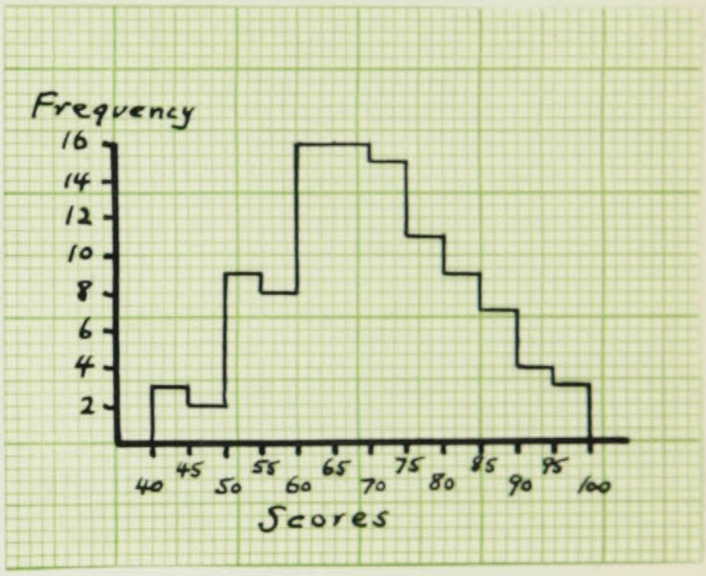


Fig.12 Distribution of Physics % Marks for Pre-Engineering Year of Science.

N = 103.

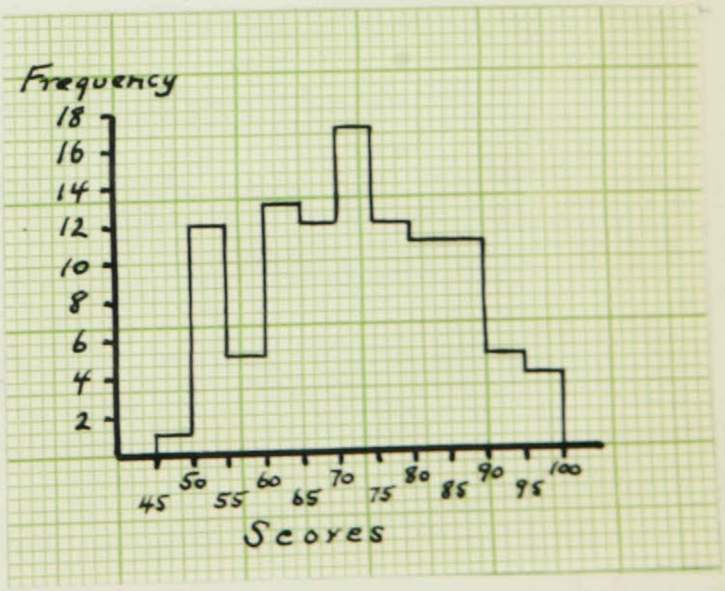


Fig. 13. Distribution of Total Average % Mark for Four Pre-Engineering Year Science Subjects. N = 103.

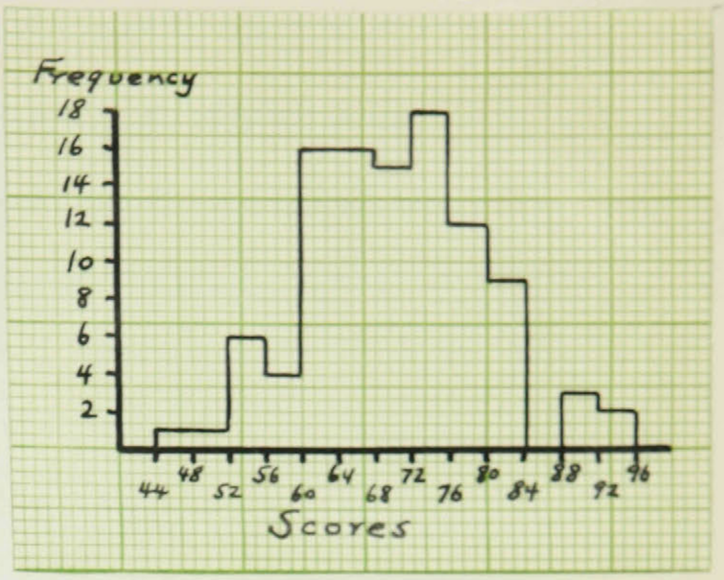


Fig. 14. Distribution of Total Average % Marks for All Subjects of First Year Engineering. N = 126.

In each of the Figures 9 - 13 there is a significant drop in frequency for scores below 50. Apparently a few students were admitted with less than 50% in some of the essential subjects. However, these distributions show, for the most part, a highly restricted coverage of the score scale below 50% and a generous sampling of scores at the high end of the scales. In Chemistry, Mathematics and Physics the scores run substantially to the maximum limit of the scale. Figures 13 and 14, representing the totals for Pre-Engineering and first year Engineering respectively, show rough similarity in trend, but in Figure 14 this whole trend appears on a slightly lower part of the scale.

(B) Statistical Evaluation of Prediction.

As has been pointed out, the present study seeks to predict attainment in first year Engineering on the basis of two factors: EPSAT test scores, and achievement in four subjects of the pre-engineering year of Science - Chemistry, English, Mathematics and Physics. The efficiency of each of these factors for prediction may be determined separately and also in combination. Three distinct steps are thus involved:

(1) Prediction of the criterion, first year Engineering attainment, on the basis of EPSAT scores.

(2) Prediction of the criterion on the basis of pre-Engineering scholastic standing.

(3) Prediction of the criterion by combining these two factors.

In order to show the relative value of each, we shall deal with these three steps individually. Having completed the separate analyses, a total summary will be presented for the purpose of comparison. In subsequent discussions of the results of this study, use of the term "criterion" without qualification will imply first year Engineering total average percent standing in all subjects.

(1) Prediction on the Basis of the EPSAT. In Table 4 will be found the correlations and intercorrelations required for this analysis.

Table 4.

INTERCORRELATIONS FOR 6 TEST PARTS OF EPSAT N = 126.

Variable	Part I	Part II	Part III	Part IV	Part V	Part VI	Composite.
Part I	-	.3886	.2131	.2254	.2644	.2031	-
Part II	.3886	-	.2630	.2041	.3232	.2145	-
Part III	.2131	.2630	-	.4433	.5272	.7033	-
Part IV	.2254	.2041	.4433	-	.3258	.4325	-
Part V	.2644	.3232	.5272	.3258	-	.3711	-
Part VI	.2031	.2145	.7033	.4325	.3711	-	-
Criterion	.3510	.3372	.4411	.4545	.4118	.4010	.5564

Any correlation in Table 4 is statistically significant at the .01 level if it equals or exceeds a value of .228 (21).

It is seen from Table 4 that the validity of the EPSAT, as measured against a criterion of total average % for the first year of Engineering, is fairly high at .5564. However, this value is decidedlylower than that reported by the authors, namely, a correlation of .733 between total EPSAT score and average engineering grade. Explanations for the lower validity of the test at McGill involve two important considerations. First of all, in the original choice of items for the test, the subjects used in the experimental battery of tests were male high school graduates who were taking Introductory Engineering Subjects at Pennsylvania State College. When the test was completed it was administered to, presumably, a later group of high school graduates taking the same subjects in the 1942 Summer Course, the validity of the test being reported against average grade. It is thus highly likely that these two groups were very similar in constitution, including educational background and instructional environment. As remarked by Dr. Webster of the McGill Department of Psychology, these conditions would tend to produce a validity coefficient of maximum value. Thus when the test is removed from the environment in which it was constructed and validated, its effectiveness tends to drop. Secondly, we must take account of the factor of strong educational selection at work in the McGill sample studied. It has been shown that the distribution of composite test scores on the EPSAT occupies approximately only the upper half of the total range of scores given in the norms. This occurrence affects the correlation between criterion and test in a detrimental way. That is, if the range were not so restricted there is a likelihood that the correlation would be higher.

Concerning the correlations between the criterion and the test parts, all are statistically significant but not very high, being lower than those reported for the EPSAT. In the McGill group, the Arithmetic Reasoning test part correlates highest with the criterion (.4545) while in the Pennsylvania group the Mathematics test part yielded the highest correlation with the criterion (.626). The highest interrelationship among the test parts occurs between the Physical Science Comprehension and the Mechanical Comprehension test parts with a correlation of .7033.

As a predictive measure, the total score on the EPSAT offers the following qualities:

Coefficient of Determination (r^2)..........	.3096
Coefficient of Non-Determination ($1-r^2$).....	.6904
Forecasting Efficiency, E, ($100\,(1-\sqrt{1-r^2})$)...	16.91%
Standard Error of Estimate ($\sigma_c\sqrt{1-r^2}$)	9.92

The coefficient of determination shows that the measuring instrument used accounts for 30.96% of the variance (mean square deviation) of the criterion thus leaving 69.04% of the variance unaccounted for by the test. A forecasting efficiency of 16.91% indicates that, with knowledge of the EPSAT score for each individual, one is able to predict his score on the criterion 16.91% better than if prediction had to be made without knowledge of these test scores. In this latter case, knowing only the average of the criterion values, the best prediction for each individual would be the mean of the criterion values. Knowing the EPSAT scores thus allows for a 16.91% reduction in errors of prediction over the condition of predicting without any other knowledge except the likely mean criterion value. The standard error of estimate has a value of 9.92 and may be interpreted to mean that in approximately two-thirds of the predictions we make, we may expect to find that our prediction will not be in error by more than 9.92 points either way. In one-third of the cases it may be in error by more than that amount , which constitutes a rather wide margin of error for individual predictions (36).

At this point it is reasonable to exploit the possibility that the EPSAT can be improved as a predicting instrument by treating the test parts as separate variables rather than merely to sum them and use a single composite score. As a matter of fact, the authors of EPSAT definitely advise validating the test itself in the specialized environment in which it is to be used and to derive the best system of weighting the various parts of the test to increase predictive power. Multiple correlation technique, employing the Doolittle solution (36), was applied to the correlations of Table 4 and the regression weights for the various test parts were determined. The following values were obtained:

Multiple Correlation (R)..............	.6035
Coefficient of Multiple Determination (R^2)	.3642
Coefficient of Multiple Non-Determination $(1-R^2)$.....	.6358
Forecasting Efficiency, E, $(100(1-\sqrt{1-R^2}))$	20.26%
Standard Error of Estimate $(\sigma_c \sqrt{1-R^2})$	9.52

These values have the same kind of interpretation as described previously. It is obvious that noticeable improvement in predicting efficiency is obtained by weighting the six test parts in multiple regression. Forecasting efficiency is increased from 16.91% to 20.26% and the standard error of estimate reduced by .40 points. This value of multiple correlation (which is, of course, statistically significant) is still much smaller than the multiple correlation value of .763 reported by the authors. It is of interest to add that the test part contributing the most to the relationship was Part 4, Arithmetic Reasoning, which accounted for 11.4% of the criterion variance while the whole battery accounted for 36.42%.

The actual weighting of the six test parts is given by the partial regression coefficients:

B_1 =	.1627	B_4 =	.2508
B_2 =	.1231	B_5 =	.1505
B_3 =	.1165	B_6 =	.0953

These may be used,as they stand,only with scores having equal variability such as standard scores.

The correlation value of .6035 between criterion and EPSAT weighted scores does not display the test's discrimination powers to best advantage. Figure 15 gives visual proof of the fact that one relatively consistent quality, of the students who are most successful scholastically, is the quality of high attainment on the EPSAT.

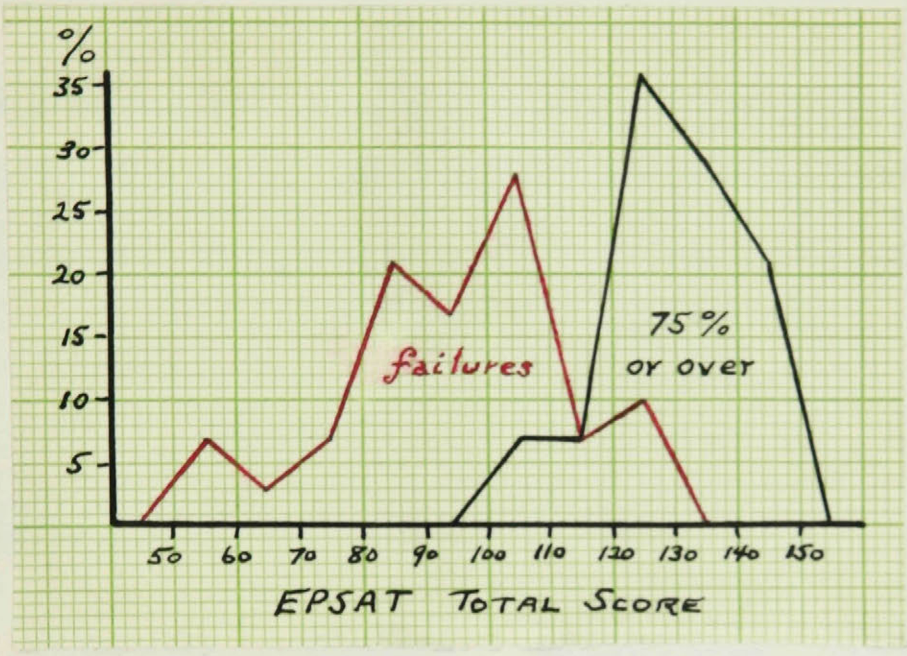


Fig. 15. A Comparison of High Achievement and Low Achievement Students.

The red graph line shows the distribution of EPSAT scores for 29 failing students while the black graph line indicates the distribution for 14 students who obtain 75% or over in first year Engineering. Two conclusions are significant. It takes a relatively high EPSAT score in order to reach the high achievement ranks. Only two students of the"75% or over" group obtained EPSAT scores below 120 (the mean of the entire class being 109). Secondly, the greater spread of the failing group,so as to overlap the high achievement group,does not demonstrate any necessary weakness in the

BPSAT. It should be taken as an indication of the uncontrolled factors of interest and application that also enter into success. In other words, given an average or high aptitude test score, we cannot be certain of success unless the individual exercises that aptitude.

(2) Prediction on the Basis of Pre-Engineering Achievement. In Table 5 are listed the intercorrelations between the four first year Science subjects: Chemistry, English, Mathematics and Physics.

Table 5.

INTERCORRELATIONS FOR PRE-ENGINEERING SUBJECTS N = 103.

Variable	Chemistry	English	Mathematics	Physics	Total Av.
Chemistry	-	.3805	.4465	.6286	-
English	.3805	-	.2042	.4245	-
Mathematics	.4465	.2042	-	.4564	-
Physics	.6286	.4245	.4564	-	-
Criterion	.4363	.4286	.4830	.5897	.6457

Correlations in Table 5 are statistically significant at the .01 level if they equal or exceed .254.

Our first inspection of Table 5 reveals the fact that the interrelationships among the subjects tend to the relatively marked, especially between Physics and Chemistry. Physics shows the highest correlation of all the subjects with the criterion of first year Engineering total average percentage, the value being .5897. English correlates with the criterion almost as highly as does Chemistry.

Using the total average percentage of these four subjects as a single evaluating instrument, the efficiency of prediction is given by :

Correlation between criterion and total average %	.6457
Coefficient of Determination..........................	.4169
Coefficient of Non-Determination......................	.5831
Forecasting Efficiency................................	23.64%
Standard Error of Estimate............................	9.12

It is evident that this method of prediction is slightly better than the use of EPSAT scores alone. Forecasting efficiency is 3.38 units higher and the standard error of estimate is reduced by .40 points. The application of multiple correlation technique to the four subjects (i.e., treating them as separate variables) occasioned a very small increase in efficiency as may be seen below :

Multiple Correlation	.6661
Coefficient of Multiple Determination..................	.4437
Coefficient of Multiple Non-Determination..............	.5563
Forecasting Efficiency................................	25.41%
Standard Error of Estimate............................	8.91

The weighting of the subjects appears in terms of the partial regression coefficients :

B_1 (Chemistry)	=	$-.0035$	B_3 (Mathematics)	=	.2679
B_2 (English)	=	.2150	B_4 (Physics)	=	.3789

Using the regression weights, our percentage reduction in errors of prediction is 25.41 as compared with 23.64 when we used the total average percentage of Chemistry, English, Mathematics and Physics weighted equally. It should be noted that in the differential weighting of the subjects, Physics contributes the most to the relationship. For example, of the 44.37% of the criterion variance accounted for by the total battery of four subjects, 22.36% is contributed by Physics. It is thus the most important single predictor of the pre-engineering year.

(3) Prediction by Combining Achievement and EPSAT Scores. The simplest method of combining these two factors is to use the total average percentage of the pre-engineering year and the EPSAT composite score, as two single variables. Only three correlations are required for this analysis, two of which have already been presented.

Correlation between criterion and pre-engineering total average (4 subjects)........................	.6457
Correlation between criterion and EPSAT total score.....	.5564
Intercorrelation of pre-engineering average and EPSAT (N = 103)..................	.2946

The results of using these two variables in multiple correlation are as follows :

Multiple Correlation..............................	.7508
Coefficient of Multiple Determination............	.5638
Coefficient of Multiple Non-Determination........	.4362
Forecasting Efficiency...........................	33.95%
Standard Error of Estimate.......................	7.89

These results are indeed gratifying. We see that the efficiency of prediction has increased remarkably as a consequence of combining these two factors of achievement and aptitude. The coefficient of multiple determination indicates that we are now accounting for 56.38% of the criterion variance whereas with the weighted EPSAT scores alone this percentage was only 36.42, and with the weighted pre-engineering subjects it was 44.37. Forecasting efficiency has been raised to 33.95% from the highest previous level of 25.41%. The standard error of estimate has been decreased from 8.91 to 7.89 or 1.02 points. The explanation for this large increase in efficiency of prediction is dependent upon the fact that the intercorrelation between the two predicting variables was low, with a value of .2946, while each one correlated substantially with the criterion, .6457 and .5564. In other words, these two factors cover somewhat different ground, all of which is important in the criterion, and thus, when combined, they show high validity.

The possibility that prediction can be pushed even further lies in the process of using the six test parts of

EPSAT and the subjects, Chemistry, English, Mathematics and Physics, as ten separate variables. In Table 6 are recorded all the necessary coefficients of correlation including the correlations already computed for previous predictions plus the intercorrelations of the four Science subjects and the six test parts.

Please see following page.

Table 6.

INTERCORRELATIONS OF SIX EPSAT TEST PARTS AND FOUR PRE-ENGINEERING SUBJECTS.

Variable	Part I	Part II	Part III	Part IV	Part V	Part VI	Chemistry	English	Mathematics	Physics.
Part I	-	.3886	.2131	.2254	.2644	.2031	.1742	.1326	.3108	.2454
Part II	.3886	-	.2630	.2041	.3232	.2145	.1874	.1865	.2714	.1478
Part III	.2131	.2630	-	.4433	.5272	.7033	.1209	.1467	.0323	.1933
Part IV	.2254	.2041	.4433	-	.3258	.4325	.1414	.1680	.1464	.1332
Part V	.2644	.3232	.5272	.3258	-	.3711	.2854	.3639	.1066	.2963
Part VI	.2031	.2145	.7033	.4325	.3711	-	.0481	-.0071	.0567	.2033
Chemistry	.1742	.1874	.1209	.1414	.2854	.0481	-	.3805	.4465	.6286
English	.1326	.1865	.1467	.1680	.3639	-.0071	.3805	-	.2042	.4245
Mathematics	.3108	.2714	.0323	.1464	.1066	.0567	.4465	.2042	-	.4564
Physics	.2454	.1478	.1933	.1332	.2963	.2033	.6286	.4245	.4564	-
Criterion	.3510	.3372	.4411	.4545	.4118	.4010	.4363	.4286	.4830	.5897

All of the additional correlations between pre-engineering subjects and EPSAT test parts are based on 103 cases and are significant at the .01 level if they equal or exceed .254.

The multiple correlation solution, using all ten variables, yields the following results:

Multiple Correlation......................	.7839
Coefficient of Multiple Determination.....	.6145
Coefficient of Multiple Non-Determination.	.3855
Forecasting Efficiency....................	37.91%
Standard Error of Estimate................	7.41

Still another noticeable gain in the efficiency of prediction is thus in evidence. This stage represents the highest level of efficiency reached in the study. Our forecasting efficiency now indicates a 37.91% reduction in errors of prediction while the standard error of estimate has been reduced from 9.92 (using only EPSAT total scores) to 7.41 with a weighting of the ten variables above. This latter value is interpreted to mean that in predicting a student's standing in first year Engineering, our prediction will not be in error by more than plus or minus 7.41 percent units in two-thirds of the cases. In one-third of the cases it may be in error by more than this amount. The total criterion variance accounted for by this ten variable battery is 61.45%. The three parts contributing most to the correlation are Physics, 18.40%; Mathematics, 11.16%; and EPSAT Part IV, Arithmetic Reasoning, 9.66%.

If the battery of ten variables as described is to be used, we must present the weights by which each variable must be multiplied before obtaining the predicted score. The form of the regression equation is :

$$X_0 = b_I X_I + b_2 X_2 + \ldots\ldots\ldots b_{10} X_{10} + \text{constant.}$$

Where X_0 is the predicted score for a single case ,
b_I, b_2 etc. are the weights for each variable,
X_I, X_2 etc. are the raw scores obtained in each variable for a single case, as follows :

X_I = EPSAT Test Part I
X_2 = " " " II
X_3 = " " " III
X_4 = " " " IV
X_5 = " " " V
X_6 = " " " IV
X_7 = Pre-Engineering Chemistry %
X_8 = " " English %
X_9 = " " Mathematics %
X_{10} = " " Physics%

Substituting the values obtained we have :

$$X_0 = .284\ X_I + .668\ X_2 + .188\ X_3 + 1.267\ X_4 + .025\ X_5 + .230\ X_6 - .006\ X_7 + .192\ X_8 + .216\ X_9 + .294\ X_{10} - 18.69$$

These weights differ from those presented in previous cases since they have been adjusted in order to be used directly with actual scores in each of the ten variables. Thus, to obtain a predicted criterion score for any student, it is simply necessary to substitute his actual scores for each variable in the above equation and solve.

For the convenience of comparison, we may summarize, in Table 7, our findings with reference to the essential measures of predictive efficiency.

Table 7.

A SUMMARY OF THE RESULTS OF PREDICTION IN ORDER OF INCREASING MERIT.

	Basis of Prediction	r or R	E	SE of Est.
1.	Total EPSAT Scores	.5564	16.91	9.92
2.	Six EPSAT Test Parts Weighted	.6035	20.26	9.52
3.	Total Average % of 4 Pre-Engineering Subjects	.6457	23.64	9.12
4.	Four Pre-Engineering Subjects Weighted	.6661	25.41	8.91
5.	EPSAT and Pre-Engineering totals Weighted	.7508	33.95	7.89
6.	Six EPSAT Test Parts and Four Pre-Engineering Subjects Weighted	.7839	37.91	7.41

It is to be concluded that the final result of this study of prediction in engineering ranks among the highest validity coefficients reported in the field for this kind of criterion. The results support the findings of many other similar studies of prediction in that past academic achievement is the best single instrument of prediction. At the same time it has been clearly demonstrated that the inclusion of the EPSAT makes for an important increase in predictive efficiency. The validity coefficient of .7839 represents a statistically significant value and a decidedly high value. And yet, one might wonder at the extent of errors of prediction which still exists, namely, a standard error of estimate of 7.41 units. In predicting any individual's likely criterion score the error of that prediction is relatively large. In view of this aspect of the problem, perhaps another method of interpreting our validity coefficient will prove to be of greater practical significance.

A Practical Interpretation. As an introduction to this method, let us observe that the statistical expressions, employed thus far to evaluate the efficiency of prediction, have been based upon the principle of individual prediction. Thus they express, for a given validity coefficient, the accuracy of the prediction for all the individual cases. Suppose now, that we are not interested in predicting a specific value for each individual, but only interested in estimating whether he is likely to "succeed" or not. In this case our requirements for efficiency are much less exacting and yet they are certainly practical enough. Educators are undoubtedly more interested in knowing whether a student can measure up to a certain standard than they are in knowing whether he will make 65% or 70%. From the counselling point of view the emphasis is the same. In conversation with a consultant psychologist, the point was made that the counsellor, knowing a student's standing in scholastic achievement and his aptitude test score, wishes to predict his chances of success or failure in engineering studies rather than his exact mark. Taylor and Russell (44), in an article in 1939, have emphasized the fact that validity coefficients even of low value, will prove to be more effective if this sort of approach is adopted. The idea, of course, is not to be considered basically new. In 1922, in the course of a summary of the work that had been done in administering intelligence tests to college students, Whipple stated : "It is quite possible, in theory, and sometimes happens, in practice, that a moderate or low statistical correlation may co-exist with a high predictive value if the object is to cull out very inferior or very superior mentalities; in other words, a mental test might fail to differentiate neatly among students of medium ability and still select with considerable precision, students of poor or excellent ability. Suppose that the primary object of testing were to locate the men who ought not to be allowed to enter the freshman class, it would then be relatively an indifferent matter if the testing did not locate in the order in which they afterward were located by their actual classroom accomplishments the men who were admitted. From this point of view, it will be seen that numerical expressions of the degree of correlation obtained are not always of final significance; ......" (47, p.262).

Taylor and Russell (44) have actually presented a method of evaluating predictive efficiency in these more practical terms. They claim that the common use of such measures as the index of forecasting efficiency leads to an excessively low interpretation of the value of a given validity coefficient. Although they have approached the problem from the standpoint of tests of selection in industry, it would seem reasonable to adopt their plan in this aspect of educational psychology. In the application of this method, we must have three items of information: first of all, it is necessary to ascertain the proportion of individuals in the class who are considered satisfactory students; secondly, we must make use of the concept of the "selection ratio", or the ratio of the number accepted to the total number seeking admission; thirdly, we must know the validity coefficient of the test or test battery to be used in the prediction. Taylor and Russell exhibit the principle of their scheme graphically as in Figure 16:

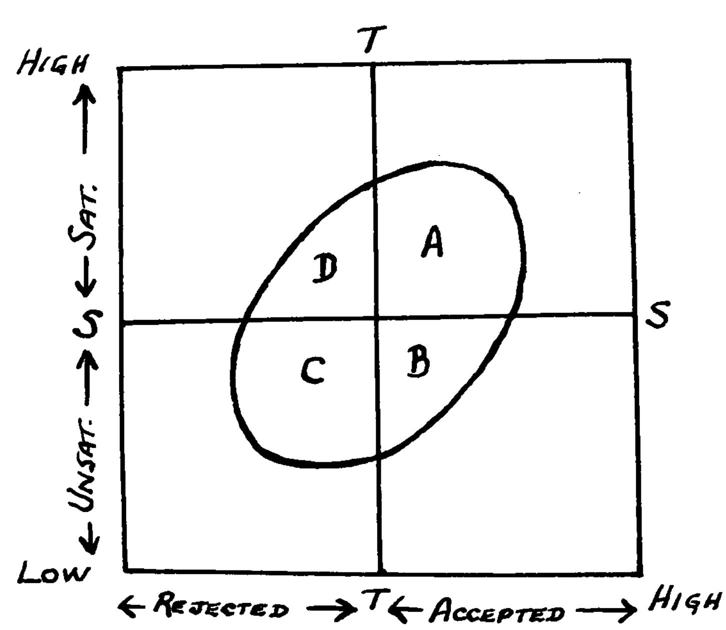


Fig. 16 Taylor - Russell Scheme

The oval-shaped area represents the scatter diagram produced when the test scores of the predicting instrument (abscissa) are plotted against the actual criterion scores (ordinate). This area becomes narrower as the validity coefficient increases. If it is decided that all students above a certain criterion score are satisfactory, then the line SS, may be located. The proportion of satisfactory students is thus given by the area A + D divided by the area A + B + C + D. If, in a future situation, more students seek admission than may be accepted, then we are able to set up a selection ratio of definite value. This locates line TT, such that the proportion obtained by dividing area A + B by A + B + C + D will equal the selection ratio. Only students above line TT will be accepted. Of this new group, the proportion of satisfactory students will differ from the original proportion. It is found by dividing area A by A + B and will obviously yield a higher proportion of successful students than in the existent group. Therefore, the excess of the ratio $\frac{A}{A + B}$ over the ratio $\frac{A + D}{A + B + C + D}$ will evaluate the efficiency of the test instrument as compared with existing means of selection. The common assumption must be made that the applicant group and the present employee (student) group are similarly constituted.

In order to facilitate the use of this scheme the authors have published tables covering all possible relationships between these factors. Knowing three values - proportion of employees (students) considered satisfactory under the existing conditions, selection ratio that can be used, and the validity of the test employed for prediction - these tables give the fourth value sought: the proportion of applicants so selected who are expected to be successful. In the present study the validity coefficient has a value of .78, the selection ratio is not yet known, and the proportion of students now in the class who are successful depends upon a decision as to what percent standing must be attained before the student is considered satisfactory or successful. A conversation with a Professor of the Faculty of Engineering led to the conclusion that a passing mark is not enough for the classification of satisfactory. The critical level should be placed higher, perhaps at 60%. Since the matter cannot be decided definitely here, we shall present the results of applying the Taylor- Russell technique for 3 different proportions of students considered satisfactory. If the critical point is set at the level of passing, then the proportion satisfactory will be .77 for the class studied (based on the group of failures as ruled by the Faculty after the final examinations,rather than on a minimum 50% standing); if it is set at 55% average standing, then the proportion satisfactory will be .65; and lastly, if it is set at 60%, the proportion will be .51. As for the selection ratio, since this is not decided until the exact moment of prediction occurs, we shall display results for any given selection ratio. All of this information will be found in Figure 17 below.

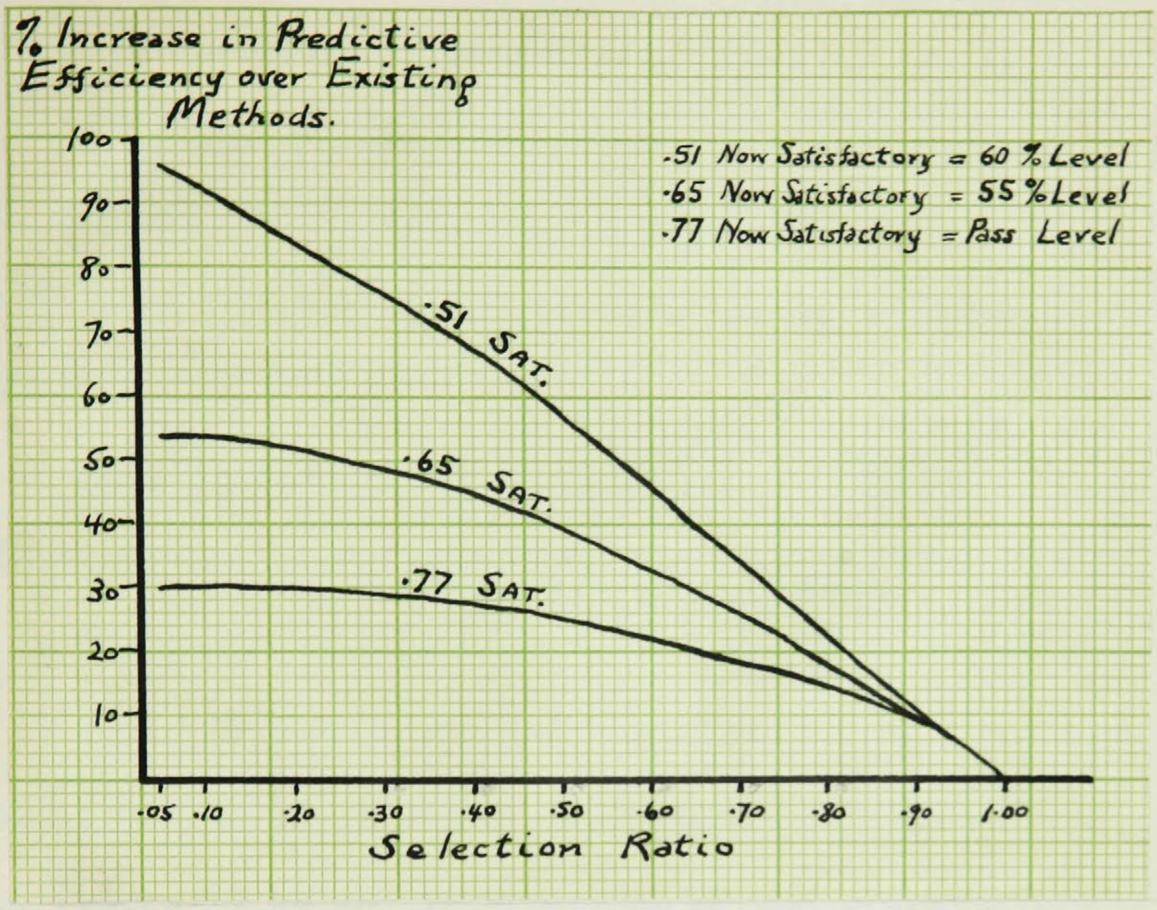


Fig. 17 The Efficiency of the Ten-Variable Test Battery Relative to Existing Means of Selection.

Figure 17 is best explained by choosing an hypothetical example. Let us suppose that we have decided to employ the present test battery, and so, for all the applicants, we derive a predicted score. This entails administering the EPSAT to these students and gathering information as to their previous standing in Chemistry, English, Mathematics and Physics. In view of the total number of candidates and present facilities, we will assume that only one-half may be accepted, giving us a selection ratio of .5. We thus consult the list of predicted scores and select only the upper 50%. In addition, we shall stipulate that a student must make 60% standing before being considered satisfactory. We can now enter Figure 17 with selection ratio .5, we shall choose the uppermost of the three graph curves designated .51 satisfactory (60% level), and on the ordinate the reading is 57%. This means that, under the conditions described, the use of the test battery will result in a 57% increase in the proportion of students expected to be satisfactory as compared with that proportion previously obtained. In this case the previous proportion was .51 and it is increased 57% by our new means of selection so that it becomes .80. Thus 80% of the students selected on the basis of predicted scores will attain or exceed a standard of 60%. If we did not set our standard so high and selected a critical value of 55% or .65 now satisfactory, our reading would be taken from Figure 17 at the intersection of the ordinate, at selection ratio .5, and the middle graph line designated .65 sat. The reading is 38% increase in efficiency of prediction over existing means of selection which has yielded .65 as satisfactory. By increasing .65 by 38% we find that .90 of the students selected by means of the test battery are expected to be satisfactory. However, the immediate aim of Figure 17 is to show, for various conditions, just how much more effective the ten-variable test battery is than the existing method of selection. This is given directly by the ordinate reading.

General trends will be obvious. As the selection ratio increasees, the effectiveness of the test battery, relative to the present method, becomes lower and lower until, when it becomes unity, we are in effect selecting everyone anyway so that the test battery would not be applied at all. The smaller the selection ratio the greater the benefit of using the test battery. Another general relationship is seen in the three graph curves. As the proportion of students now considered satisfactory becomes lower, i.e., as the critical level is set higher, the benefit of using the test battery increases.

It should be pointed out that we must use this scheme of interpretation with caution. The common assumption, mentioned before, is that the applicant group is similarly constituted to the present group studied, otherwise these expected values will not hold. Applying the method strictly, we probably should pre-select the applicant group itself according to the 1944 standards so that it becomes reasonably similar in constitution to the group studied. If approximate similarity can be achieved, then the application of the Taylor - Russell values is valid.

Using the Regression Formula. As a last form of statistical investigation, we can observe the results of applying the regression equation, developed for the prediction of scores, to the data gathered for the study. This will serve the dual purpose of illustrating the scatter diagram which forms the basis of the Taylor - Russell method and of illustrating graphically the effectiveness of the test battery. For each of the 103 students who had complete records in all ten variables, a predicted percentage **score** has been estimated and plotted against their actual percentage in the first year Engineering. In Figure 18 the predicted score is on the abscissa and the actual score or criterion on the ordinate.

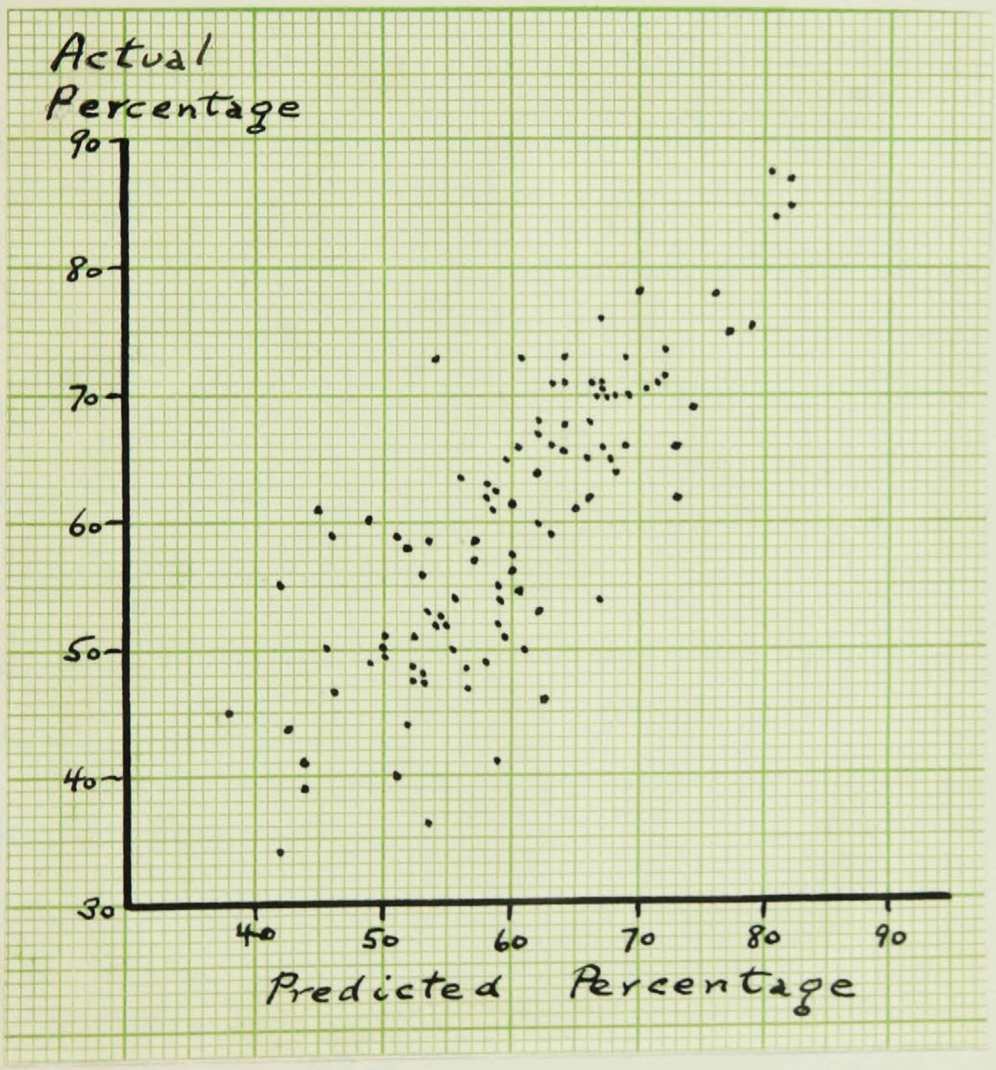


Fig. 18 A Scatter Diagram of Predicted and Actual Average Percentage in First Year Engineering.

CHAPTER V

SUMMARY AND CONCLUSIONS

Emphasis must be placed upon the heavy academic mortality rate in engineering colleges, and the consequent need of a substantial method of predicting aptitude for engineering training. Unfortunately, this is not a simple task inasmuch as success in engineering studies is determined by a number of factors in addition to basic ability. The advantages of a reliable form of prediction would be felt in the field of student counselling as well as educational selection.

Early effort in educational prediction utilized the intelligence test as the measuring instrument and general scholastic success as the criterion. The typical range of validity coefficients for such studies extended from .30 to .60. However, a trend toward a more refined type of prediction soon gained ground. This movement called for testing devices capable of measuring specific aptitudes that were required for success in different schools and colleges of the university. The typical studies predicting achievement in engineering colleges have employed batteries of tests or test parts and have revealed their efficiency in terms of the multiple correlation coefficient. The range of such values may be approximated as .51 to .79.

In analysing the specific factors which determine engineering success, six main headings may be adopted. These are: (1) Aptitude for higher mathematics; (2) Ability to perceive sizes, shapes, and relations of objects in space; (3) School grades and achievement in Physics and Chemistry; (4) Ability in English; (5) Interests; (6) Miscellaneous factors including motivation and study habits.

The test employed in this study is the Engineering and Physical Science Aptitude Test by Moore, Lapp and Griffin. It was administered to 141 McGill students entering the first year of Engineering in the fall, 1944. The average mark made by the students at the end of this first year is used as the criterion. Pre-engineering achievement in Chemistry, English, Mathematics and Physics is also included among the predicting instruments.

The general statistical analysis of results is based upon the means, standard deviations, and percentiles for the aptitude test scores, and constitutes evidence of high performance on the part of McGill Engineering students.

Prediction is estimated by means of correlation and multiple correlation technique, and is presented, in three steps, as based upon: (1) the aptitude subtest scores, (2) pre-engineering achievement in Chemistry, English, Mathematics and Physics, (3) a combination of (1) and (2). These three bases of estimate occupy the same positions when arranged in order of increasing predictive efficiency (validity coefficients :.6035, .6661 and .7839 respectively). The multiple correlation coefficient of .7839 found in step (3) is given a practical evaluation in terms of its efficiency to predict "satisfactory" or "unsatisfactory" performance of the student in the first year of Engineering.

On the basis of the results of the present study, the following general conclusions appear to be justified:

1. This sample of first year students in Engineering at McGill University represents a high degree of educational selection.

2. The Engineering and Physical Science Aptitude Test has a lower validity (for this group of students) at McGill University than at Pennsylvania State College.

3. Pre-engineering average scholastic standing in Chemistry, English, Mathematics, and Physics, is the best single indication of success in first year Engineering.

4. Of these four subjects, Physics shows the highest correlation with the average mark for first year Engineering.

5. The addition of the Engineering and Physical Science Aptitude Test scores to pre-engineering standing results in a valuable increase in predictive efficiency. This aptitude test is therefore seen to be a very useful psychological instrument.

6. A ten-variable test battery consisting of the six test parts of the Engineering and Physical Science Aptitude Test and the four pre-engineering subjects, Chemistry, English, Mathematics, and Physics, constituted the best basis for prediction, with a multiple correlation coefficient of .78 with the average mark in first year Engineering.

7. The use of this predicting test battery is expected to be of marked practical value for selection and guidance.

BIBLIOGRAPHY

1. Arlitt, A.H. and Hall, M. Intelligence tests versus entrance examinations as a means of predicting success in college. J. of Applied Psychology, 1923, 7, pp. 330-338.

2. Bennett, G.K. and Cruikshank, R.M. A Summary of Manual and Mechanical Ability Tests. The Psychological Corporation, New York, 1942.

3. Benton, A.L. and Perry, J.D. A study of the predictive value of the Stanford Scientific Aptitude Test (Zyve). J. of Psychology, 1940, 10, pp. 309-312.

4. Berdie, R.F. The prediction of college achievement and satisfaction. J. of Applied Psychology, 1944, 28, pp. 239-245.

5. Bernreuter, R.G. and Goodman, C.H. A study of the Thurstone Primary Mental Abilities Tests applied to freshman engineering students. J. of Educational Psychology, 1941, 32, pp. 55-60

6. Bingham, W.V.D. Aptitudes and Aptitude Testing. Harper and Brothers, New York, 1937.

7. Borow, H. Manual for Engineering and Physical Science Aptitude Test. The Psychological Corporation, New York, 1943.

8. Brush, E.N. Mechanical ability as a factor in engineering aptitude. J. of Applied Psychology, 1941, 25, pp. 300-312.

9. Butsch, R.L.C. Improving the prediction of academic success through differential weighting. J. of Educational Psychology, 1939, 30, pp. 401-420.

10. Caldwell, H.H. Adult tests of the Stanford Revision applied to college students. J. of Educational Psychology, 1919, 10, pp. 477-488.

11. Compton, R.K. Student evaluation of knowing college aptitude test score. J. of Educational Psychology, 1941, 32, pp. 656-664.

12. Corey, S.M. The effect of motivation upon the relationship between achievement and intelligence. School and Society, 1935, 41, pp. 256-257.

13. Darley, J.G. Scholastic achievement and measured maladjustment. J. of Applied Psychology, 1937, 21, pp. 485-492.

14. Davis, R.A. Testing for aptitudes. J. of Educational Psychology, 1945, 36, pp. 39-45.

15. Douglass, H.R. Different levels of patterns of ability necessary for success in college. Occupations, The Vocational Guidance Magazine. 1943, 22, pp. 182-186.

16. Gerberich, J.R. Factors related to the college achievement of high-aptitude students who fail of expectation and low-aptitude students who exceed expectation. J. of Educational Psychology, 1941, 32, pp. 253-265.

17. Goodman, C.H. A comparison of the interests and personality traits of engineers and liberal arts students. J. of Applied Psychology, 1942, 26, pp. 721-737.

18. Goodman, C.H. Prediction of college success by means of Thurstone's Primary Abilities Tests. Educational and Psychological Measurement. 1944, 4, pp. 125-140.

19. Griffin, C.H. and Borow, H. An Engineering and Physical Science Aptitude Test. J. of Applied Psychology, 1944, 28,pp. 376-387.

20. Griffiths, G.R. The relationship between scholastic achievement and personality adjustment of men college students. J. of Applied Psychology, 1945, 29, pp. 360-367.

21. Guilford, J.P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1942.

22. Harris, D. Factors affecting college grades: a review of the literature 1930-1937. Psychological Bulletin, 1940, 37, pp. 125-166.

23. Herriott, M.E. "Honor" engineering students: their characteristics and reasons for success. J. of Engineering Education, N.S. 1929, 19, pp. 871-883.

24. Holcomb, G.W. and Laslett, H.R. A prognostic study of engineering aptitude. J. of Applied Psychology, 1932, 16, pp. 107-115.

25. Hurd, A.W. A study of achievement in various engineering and pre-engineering courses. J. of Educational Research, 1931, 23, pp. 25-30.

26. Johnson, A.P. The prediction of scholastic achievement for freshman engineering students at Purdue University. Purdue Univ. Studies in Higher Education, No. 44, 1942.

27. Laycock, S.R. and Hutcheon, N.B. A preliminary investigation into the problem of measuring engineering aptitude. J. of Educational Psychology, 1939, 30, pp. 280-288.

28. Leitch, A. The development of mental measurements in American Colleges and Universities. J. of Educational Psychology, 1943, 34, pp. 407-419.

29. Mann, C.V. An evaluation of placement examinations. J. of Engineering Education, N.S., 1928, 19, pp.288-300.

30. Mann, C.V. Objective type tests in engineering drawing and descriptive geometry. J. of Engineering Education, N.S., 1929, 19, pp. 979-992.

31. Marshall, M.V. A study of the Stanford Scientific Aptitude Test. Occupations, 1942, 20, pp. 433-434.

32. Mc Gehee, W. The prediction/of differential achievement in a technological college. J. of Applied Psychology, 1943, 27, pp. 88-92.

33. Moore, B.V. Analysis of results of tests administered to men in Engineering Defense Training Courses. J. Applied Psychology, 1941, 25, pp. 619-635.

34. Moore, B.V., Lapp, C.J., and Griffin, C.H. Engineering and Physical Science Aptitude Test. The Psychological Corporation, New York, 1943.

35. Ohmann, O.A. A study of the causes of scholastic deficiencies in engineering by the individual case method. Univ. of Iowa Studies in Education, Vol. 3 No. 7.

36. Peters, C.C. and Van Voorhis, W.R. Statistical Procedures and Their Mathematical Bases. McGraw-Hill, New York, 1940.

37. Remmers, H.H. and Geiger, H.E. Predicting success and failure of engineering students in the School of Engineering in Purdue University. Purdue Univ. Studies in Higher Education, No. 38, 1940, pp. 10-19.

38. Sisk, H.L. A multiple factor analysis of mental abilities in the freshman engineering curriculum. J. of Psychology, 1939, 9, pp. 165-177.

39. Smith, F.F. The use of previous record in estimating college success. J. of Educational Psychology, 1945, 36, pp. 167-176.

40. Stalnaker, E.M. A four year study of the freshman class of 1933 at the West Virginia University. J. of Educational Research, 1945, 39, pp. 81-101.

41. Stevenson, J.A.F. The measurement of scientific aptitude in the field of student personnel work. McGill University Thesis, 1938.

42. Stoddard, G.D. Iowa Placement Examinations. School and Society, 1926, 24, pp. 212-216.

43. Stoddard, G.D. Iowa Placement Examinations. Univ. of Iowa Studies in Education, Vol. 3, No. 2.

44. Taylor, H.C., and Russell, J.T. The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. J. of Applied Psychology, 1939, 23, pp. 565-578.

45. Thornton, G.R. The use of tests of persistence in the prediction of scholastic achièvement. J. of Educational Psychology, 1941, 32, pp. 266-273.

46. Weintraub, R.G., and Salley, R.E. Graduation prospeats of an entering freshman. J. of Educational Research, 1945, 39, pp. 116-126.

47. Whipple, G.M. Intelligence tests in colleges and universities. 21st Yearbook of National Society for the Study of Education. 1922, Chap. 10.

48. Wilcox, E.R. Placement tests as a means of predicting failures in first year work in engineering at the University of Washington. J. of Engineering Education, N.S., 1930, 20, pp. 791-794.

49. Williamson, E.G. The decreasing accuracy of scholastic predictions. J. of Educational Psychology, 1937, 28, pp. 1-16.

50. Wilson, M.O., and Hodges, J.H. Predicting success in the engineering college. J. of Applied Psychology, 1932, 16, pp. 343-357.

51. Young, K. The history of mental testing. The Pedagogical Seminary, 1923, Vol. 31, No.1.

52. Zyve, D.L. A test of Scientific Aptitude. J. of Educational Psychology, 1927, 18, pp. 525-546.