A LONGITUDINAL STUDY OF THE RELATIONSHIP OF DIFFERENTIAL APTITUDE TEST SCORES WITH COLLEGE SUCCESS

By

EDWIN EARLE VINEYARD

Bachelor of Science
Oklahoma Agricultural and Mechanical College
Stillwater, Oklahoma
1949

Master of Science
Oklahoma Agricultural and Mechanical College
Stillwater, Oklahoma
1951

Submitted to the faculty of the Graduate School of
the Oklahoma Agricultural and Mechanical College
in partial fulfillment of the requirements
for the degree of
DOCTOR OF EDUCATION
August, 1955

A LONGITUDINAL STUDY OF THE RELATIONSHIP

OF DIFFERENTIAL APTITUDE TEST SCORES

WITH COLLEGE SUCCESS

Thesis Approved:

Harry K. Brobst
Thesis Adviser

Roy E. Sommerfeld

Robert MacVicar
Dean of the Graduate School

ACKNOWLEDGMENTS

The writer would like to acknowledge his indebtedness to those persons who through their assistance and cooperation made this study possible.

Sincere appreciation is extended to the members of the advisory committee, Professors Harry K. Brobst, Roy E. Sommerfeld, S. L. Reed, Eli C. Foster, and Carl E. Marshall, for their guidance of this research from inception to completion.

For services in statistical computation by I. B. M. the writer is indebted to Professor Carl E. Marshall and his staff, and for assistance in the factor analysis of test data to Dr. Pairlee J. Stinson. For cooperation in making necessary information available from student personnel records, the writer is indebted to the administrative staff of Stillwater High School, the staff of the Registrar's Office, and the staff of the Bureau of Tests and Measurements.

To those above, to Mrs. Vineyard, and to the many others who offered encouragement and helpful suggestions, the writer extends his deepest appreciation.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

CHAPTER I

INTRODUCTION

Schools in Oklahoma and elsewhere are becoming increasingly interested in the establishment and maintenance of adequate programs for academic and vocational guidance. Testing programs for purposes of guidance are becoming commonplace, and these programs are increasingly being placed under the direction of counseling personnel who have had some training in the techniques of test administration and interpretation. These counselors are being charged with the responsibility of advising students with respect to educational goals and vocational objectives. Not only are they expected to help the students make immediate choices and decisions, but they must also help the pupils set realistic long term goals toward which to strive.

The counselor is aware that the students whom he advises differ widely from one another in their abilities, interests, and ambitions. He is also aware that these students differ within themselves, that they have areas of strength and areas of weakness. As Bingham puts it:

> An individual's potentialities are not all equally strong. One can learn to do certain things more easily and better than other things, and can develop greater interest and satisfaction in some kinds of activity than others.
>
> Individuals differ from one another in their potentialities.
>
> Many of these differences are relatively stable. They tend to persist.[1]

---

[1] W. V. Bingham, Aptitudes and Aptitude Testing (New York, 1937), p. 72.

Jones seems to feel that while these principles of individual differences are commonly recognized, much guidance of pupils is done on the basis of other assumptions:

> It is generally believed that individuals differ widely in native capacity. We no longer hold to the opinion that all are born equal in capacities any more than that all are born boys, or with red hair, or with brown eyes, or with any other purely physical characteristic. In spite of this generally accepted principle, guidance, as well as other parts of education, sometimes seems to assume the opposite and bases its conclusions on the principle that anyone can become anything he chooses. All he needs is proper assistance and the "will to do."[2]

The decision of whether or not to plan toward a college education is one of the prominent student problems with which the counselor is concerned, and it is a problem toward which the present research is directed. Quite frequently the student and his parents, if not the counselor, may be operating under the latter assumption mentioned above by Jones. They feel that anyone can be successful in completing any college program he may choose if only enough persistence and effort are expended. Darley takes this point and expands upon it:

> Furthermore, those families that send their children on to higher education seem often to view the process as a specialized form of unemployment insurance or social sine qua non; it is a necessary finishing process which, in the opinion of the parents, can be completed if only their children work hard enough. The fact that approximately six out of ten college entrants fail to survive to graduation gives rise to disappointments and for parents frustrations that may have a serious effect on the personalities of their children.[3]

While there may be no more certain way for an individual student to learn that he does not possess the attributes needed to succeed in college

---

[2]Arthur J. Jones, Principles of Guidance (New York, 1945), p. 84.

[3]John G. Darley, "The Functions of Measurement in Counseling," In E. F. Lindquist, Educational Measurement. (Washington, D. C., 1951), p. 69.

than to attempt and fail, this practice has its rather obvious drawbacks. As brought out above by Darley, one rather serious effect is the psychological consequences of frustration and failure upon the student and his family. Chauncey and Fredericksen elaborate further in this vein of thought:

> The procedure of eliminating a large proportion of students on the basis of grades in the first term or half-term, that is, of tryout selection, has obvious advantages. The validity of the procedure is unquestionable insofar as course grades are valid and reliable; correlational studies show a higher relationship between first semester grades and measures of later academic success in college than is usually found between selection test scores and college grades. For example, in a recent study at Princeton a correlation of .78 was found between first semester average and average grade at graduation. On the other hand, the objections to the tryout are not only that it is inefficient and costly, but, even more important, that it may cause a serious sense of failure and frustration in many students.[4]

While the vantage point from which the high school counselor is likely to view this problem of the advisability of higher education for his counselees is likely to be that of the individual welfare of the students involved, there is another angle from which the problem may be viewed. Chauncey and Fredericksen have made the point that selection by tryout is an inefficient and costly practice, in addition to the psychological stress which may accrue to the individual students involved. Black is rather emphatic in his denunciation of the wastefulness of human and economic resources involved in student mortality in college:

> It seems highly questionable, wasteful of human resources, inefficient, and costly for a university .... to admit 2868 students of whom 45% withdraw from the university before

---

[4]Henry Chauncey and Norman Fredericksen, "The Function of Measurement in Educational Placement," In Lindquist, p. 86.

> the end of their second year with records primarily of low grades, few credits completed, and many credits "failed," "withdrawn," and "incomplete."[5]

While the problem in the discussion thus far has appeared to be one mainly concerned with who should go to college, it is not so simple. There is a corollary to the problem which concerns the field of concentration or the course of study to be pursued if college training seems to be advisable. The solution of the corollary hinges somewhat on the resolution of the problem itself, but there is a converse relationship as well. While it is a commonly accepted principle that there is a minimum amount of general academic aptitude necessary for success in any college curriculum, it is also true that the question of whether college enrollment is advisable for a given student may well depend upon the curriculum he chooses to take. That is, there is a differential aspect of ability as well as a general one, and the different curricula place emphasis on different abilities.

While the problem of who should go to college is not a new one by any means, the two-fold aspect of the problem has mushroomed in importance with the emergence of the guidance movement and the appearance of tests purporting to measure differential abilities. Recent studies have tended to place more emphasis upon the differential nature of the advisement problem. The approach of Morris is an example:

> The problem of who should go to college and in what special fields of endeavor the students should concentrate is not one of recent origin. For many years now most colleges have been concerned about the high mortality rate of their students, due often to the lack of capacity of some individuals for

---

[5]Donald B. Black, "Prediction of Academic Success in the University of Washington," Unpublished Doctoral Thesis, University of Washington (Seattle, 1951), p. 81.

> certain types of endeavor. Moreover, psychologists have long known (1) that different mental processes are required for different subject matter fields and (2) that individuals differ within themselves as well as among others in utilizing the mental processes they possess. Educational growth, then, would seem to depend a large extent on intelligent and adequate educational guidance.
>
> The growing enthusiasm for higher education makes the problem of effective differential guidance, and consequent reduction of student mortality, even more pressing.[6]

Students should be helped to make educational and vocational choices in keeping with their levels of ability and the kinds of ability they possess. A student should be helped to choose, if there are no mitigating circumstances beyond his control, a vocation or profession on the highest level on which it is possible for him to operate. He should be helped to choose the type of training from which he has the greatest capacity to profit. Thus the greatest good and satisfaction will come to the individual, and society will receive the greatest return.

Pupils with superior scholastic aptitude should be encouraged to prepare for and to enter college. They need to be helped in their choice of areas of concentration so that strengths may be capitalized upon and weaknesses minimized. Those without definite plans to attend college should, through counseling, be helped to develop a point of view which will take them into college.

Those pupils without sufficient ability to succeed in college should be helped in the selection of educational and vocational goals suitable to their levels and types of ability. These students need to be helped to adjust their aspirations to realistic levels and to set

---

[6]Louis H. Morris, "The Relationahip Between Certain Factors and Academic Success in College Mathematics," Unpublished Doctoral Thesis, Oklahoma A. and M. College (Stillwater, Oklahoma, 1954), p. 1.

their sights upon realistic and attainable goals. Knapp suggests that the counselor should have at hand some evidence concerning student aptitudes and should use it in the advisement process:

> It might be well to present concrete evidence to students with superior scholastic aptitude, as well as to those with limited scholastic aptitude, and to their parents, indicating the pupils' chances of doing successful college work.[7]

This, of course, implies that such concrete evidence should have been collected previously by the counselor and be readily accessible. Jones states, "The first job of guidance would then be to discover the abilities ... possessed by the individual."[8] In order to properly perform his function of guiding students in their decisions, the counselor must be cognizant of recent theory and research concerning aptitudes and aptitude testing. Traxler suggests that there are certain assumptions concerning aptitudes which have been justified by observation and research, and which are fundamental to scientific guidance:

> There are three important assumptions concerning aptitudes which are justified by general observation and research, and which are among the basic tenets of a program of guidance which lays any claim to being scientific. The first assumption is that few, if any, individuals have equally strong aptitudes in all directions. ... A second assumption is that individuals differ from one another in every aptitude they possess regardless of whether broad aptitudes or very specific aptitudes are being considered. ... A third hypothesis is that differences among individuals and within individuals tend to persist within limits. They are fairly, although not completely, stable.[9]

---

[7]Robert H. Knapp, Practical Guidance Methods (New York, 1953), p. 11.

[8]Jones, Principles of Guidance, p. 91.

[9]Arthur E. Traxler, Techniques of Guidance (New York, 1945), p. 43.

These thoughts by Traxler seem to crystallize certain points rather well. First, individuals differ in the amounts of a given aptitude or ability they possess. Second, each individual differs within himself in the amounts of different kinds of aptitude he possesses. Third, aptitudes tend to be fairly stable. With these observations in mind concerning the nature of aptitudes, their distribution and behavior, the next problem concerns their measurement. Measurement usually means testing, and, as Darley points out, testing is an integral part of counseling:

> Any discussion of counseling must necessarily involve a discussion of tests, which should be considered as an economical and accurate means of studying the student preparatory to effective guidance. The rising importance of tests cannot be stressed too heavily.[10]

> Testing is one of the more important devices to help the individual student in solving his problems. It is quite possible that the availability of increasing numbers of tests has contributed to the guidance bandwagon's speed.[11]

Tests then are legitimate and useful instruments for evaluating the level of ability or abilities possessed by students. Bowles appeared to be very enthusiastic concerning the possibilities of aptitude testing when he wrote the following in 1951:

> Aptitude tests can be used by themselves to predict college success and they will predict it, regardless of the pattern of subjects studied or to be studied. Aptitude tests can be improved, and as tests and prediction are improved, use of aptitude tests as the controlling criterion for college entrance will increase.[12]

---

[10] John G. Darley, Testing and Counseling in the High School Program (Chicago, 1943), p. 14.

[11] Ibid., p. 20.

[12] Frank H. Bowles, "Crystal Balls and College Admissions," American Council on Education Studies, XV (1951), p. 8.

Although all authorities are not as enthusiastic as the above would indicate, aptitude testing is thought by some to be the answer to the problem of predicting success in college, and to the problem of differential choice of curricula. The need for just such a set of measuring devices and the apparent logic of their usefulness has brought tremendous popularity in recent years to standardized batteries of guidance tests purporting to measure differential capacities. The testing movement appears to be firmly entrenched in the public schools of this country. According to Traxler:

> Results of a survey show that testing programs now have an important and well established place in the school systems of the majority of the cities in the United States. Among the uses of test results, guidance stands first; instructional use, second.[13]

Tests are tools which serve the guidance function. They help the counselor to know and better understand the strengths, weaknesses, and needs of individual pupils. Tests measure individual differences, which are the very heart of the guidance program. They measure differences within the individual and thus facilitate self-analysis, another of the primary functions of guidance.

Two basic qualities necessary in any test are reliability and validity. Closely akin to these qualities and of basic interest to the counselor is that of prediction. Essentially, prediction is the function that the academic or vocational counselor is attempting to perform on the basis of test data. The predictive value of the test instruments used, then, is of utmost importance in the successful counseling program. These points are brought into focus by Darley:

---

[13]Arthur E. Traxler, "The Status of Measurement and Appraisal Programs," Educational Records Bulletin, LXI (February, 1953), p. 75.

> There is first the property of reliability or accuracy. The current methods of test construction maximize the precision and consistency with which behavior can be measured. In counseling situations, as in many other situations, this increased precision is a necessary antidote to tendencies toward over or under estimation of relevant behavior by either the counselor or the student.
>
> There is second the property of validity or meaning or predictive power. In spite of all the difficulties of locating adequate criteria of vocational success, or educational success, or personal adjustment, good psychological tests carry some forecasting power for things to come in the life of the individual student, and this predictive value is a balance wheel again in the wishful thinking and perennial optimism with which students approach many long range decisions.[14]

Cronbach enunciates rather emphatically upon the close relationship of the predictive function to testing:

> An attempt to predict underlies every use of testing. Whenever a test is given to two people, it tells about some difference between their performances at this moment. But the fact would be of no significance, would not be worth knowing, if from it one could not predict that these two people would differ in some future activity.[15]

In another context Cronbach elaborates further:

> Prediction emphasizes differences between individuals, or between an individual and some standard. A second function of testing, diagnosis, emphasizes differences among characteristics of the same individual. Diagnosis in guidance and individual case work identifies the particular strengths of the individual so that he may capitalize on them, or the particular weaknesses so that he may adapt to or correct them. Diagnosis involves prediction, since the test user must decide what behavior the present pattern of characteristics permits and how stable that pattern will remain.[16]

It is with the predictive efficiency of one such group of guidance tests, the Differential Aptitude Battery (DAT),[17] that the present study

---

[14]John G. Darley, "The Functions of Measurement in Counseling," p. 77.

[15]Lee J. Cronbach, Essentials of Psychological Testing (New York, 1949), p. 9.

[16]Ibid., p. 11.

[17]George K. Bennett, Harold G. Seashore, and Alexander G. Wesman, A Manual for the Differential Aptitudes Tests (New York, 1952).

is concerned. This battery which first appeared in 1948 has attained a rather unusual popularity to have been available for so short a time.[18] Designed for use with the high school population, it has gained wide acceptance and use in secondary school guidance programs. From informal verbal observations made by persons in charge of bureaus which are the two largest suppliers of guidance and test materials to the public schools in Oklahoma, the Differential Aptitude Battery is one of the most popular guidance items in use in state schools today.[19]

It is a practice in some state high schools to administer guidance tests, such as the DAT, to entering freshmen and to do no further testing of this type. These scores are subsequently used throughout the high school period in the selection of major curricula, long range advisement, and in terminal counseling of seniors.

Data are rather plentiful concerning the relationship of DAT scores and other guidance test scores with high school grades. Data are also rather plentiful concerning the relationship between scores on scholastic aptitude tests given to entering college freshmen and college grades. However, there are yet no data reported in the common sources on the relationship of DAT scores obtained early in high school with academic success in college. Nor are any data reported on the relationship of such DAT scores and academic success in different areas of college study.

---

[18] Donald E. Super, Appraising Vocational Fitness (New York, 1949), p. 368.

[19] Informal conversational comments made by Dr. Harry K. Brobst, Director of the Bureau of Tests and Measurements at Oklahoma A. and M. College and Dr. W. R. Fulton, Director of the Educational Materials Division of the University of Oklahoma.

Purpose of the Study

The purpose of the current study is to investigate whether the scores made by high school freshmen on the DAT offer information which may be used with confidence by the high school counselor for long range advisement of these students. Under particular consideration is the value of the battery as a tool of the counselor in assessing the long range advisability of higher education and in the selection of major college curricula.

While the avowed purpose of this study as just outlined is general in nature and arises from a general need, the attack upon the problem will be specific. This will be a predictive study of a particular battery of guidance tests administered to a particular group of freshman students in high school who subsequently attend one college. The problem is attacked in this specific manner because practicality precludes its attack on a wider scope. Accruing to this approach are certain limitations. These will be considered in the final section of this chapter.

Statement of the Problem

The problem of this study is to identify the relationship between scores made by a sample of Stillwater High School freshmen on the Differential Aptitude Tests and the subsequent grade averages made by the same students as freshmen at Oklahoma A. and M. College.

In the course of the study, the following specific outcomes will be obtained for males and females separately:

A. Pearsonian correlation coefficients between scores on each Differential Aptitude Test and college freshman over-all grade average based upon whole year academic performance.

B. Pearsonian coefficients of correlation between scores on each of the Differential Aptitude Tests and college freshman grade averages in the fields of language, social science, science, mathematics, and home economics.

C. The coefficient of multiple correlation between all or the optimum combination of tests of the Differential Aptitude Battery and college freshman over-all grade average.

D. The multiple regression equation for prediction of college freshman over-all grade average from known scores on the Differential Aptitude Tests.

E. The coefficient of correlation between over-all grade averages predicted by the obtained multiple regression equation for members of a validation group from the same population and the actual over-all grade averages made by this validation group as college freshmen.

F. Test groupings or factors by factor analyzing the test intercorrelation matrix, and multiple coefficients of correlation between the component tests of each factor and the over-all grade average criterion.

Need for the Study

Since the Differential Aptitude Tests are now being used widely by counselors for long range advisement of high school youth, it is imperative that research be forthcoming which will shed some light on the efficacy of this practice. The general need for information concerning the long range predictive efficiency of this battery has given rise to the undertaking of this study. The general dearth of such information in the literature gives the study a certain uniqueness and an added importance.

Speaking of new tests of this general type, Traxler called for research:

> There is need for much research with these newer tests that are designed to make diagnostic measurements of mental functions. Their guidance possibilities are great, but as yet so little research has been done with the results that it is impossible for even the specialists in measurement to say just what the various scores mean or how they may be used in the practical job of counseling pupils. Counselors and teachers may use the scores experimentally, but until further studies are available, predictions based on them must be made cautiously.[20]

From the point of view of the practicing counselor, Fredericksen in 1951 lamented the lack of adequate predictive data:

> The longer I have tried to counsel students, the more I have been impressed with the lack of adequate information on which to base interpretations of test scores. Even when dealing with prediction of academic success, which has been studied more thoroughly than any other prediction problem, I often feel the handicap of a lack of useful information to give to the student sitting across my desk who wants to know his chances of being successful in some academic undertaking. To such problems as the choice of a college, the choice of a curriculum, a client might legitimately expect me to make useful statements about the probabilities of success in those academic programs which he is considering.
>
> What I am trying to say is that I, for one, feel frustrated by the lack of adequate information for the interpretation of a test score.[21]

Being specific again, information is needed concerning the relationship of the scores on each of the Differential Aptitude Tests and college success. Since the battery is based upon the principle of differential measurement of aptitude, and since success in various college curricula is thought to involve certain patterns of abilities, information is

---

[20]Traxler, Techniques of Guidance, p. 48.

[21]Norman O. Fredericksen, "Making Test Scores More Useful for Prediction," Educational and Psychological Measurement, XI (Winter, 1951), p. 783.

also needed concerning the relationship of scores on each of the tests in the battery with academic success in different subject matter fields in college.

In each case the need is for longitudinal research relating test results obtained early in the student's high school career with later success in the different college programs. Travers is crystal clear on this point:

> Attempts to predict college success have been made too often at a time in the student's career when it is too late for him to make use of such predictions. The common practice of administering tests of scholastic aptitude after the student has been admitted to college represents one of the worst examples of this folly. If such college freshmen tests should indicate that the student is an exceptionally poor college risk, it is too late to save the student from many of the frustrations and disappointments which necessarily accompany college failure. If college success could be predicted early in a child's development then much could be done to provide him with the necessary background for college, and even more important, it would be possible to provide a well-planned secondary school curriculum for those who have little aptitude for college work.[22]

Limitations of the Study

While this study is undertaken for the purpose of gathering information which will contribute to the meeting of the above-mentioned needs, it is not proposed that this piece of research will establish the complete and final answer to these problems. The study is limited in its scope and coverage and for this reason cannot do more than make a beginning in this area of research.

The general problem outlined, a study of the longitudinal relationahip of DAT scores and college success, is herein attacked in a

---

[22]Robert M. Travers, "Significant Research on the Prediction of Academic Success," In Wilma T. Donahue, The Measurement of Student Adjustment and Achievement (Ann Arbor, Michigan, 1949), p. 153.

specific situation. The groups of students on which the study is based attended the same high school and later attended the same college. Because of the selected rather than random nature of the group, statistical inference on the basis of the results would be unsound. However, this does not preclude logical inference, although such inference must be made cautiously and with some degree of uncertainty.

This question of universality of application of results is an important one in any research study and one which must be resolved in some manner. No claim of universality is made for this research since the problem is not attacked universally. Nevertheless, this specific study of a universal problem will contribute to an ultimate solution, and more will be known of the relationships studied than is now known. Further research of a similar nature in other situations will need to be done.

CHAPTER II

PROCEDURE

The purpose of this chapter is to present the general conditions under which the present research was done and the procedure employed. The subject groups will be described, the conditions of test administration discussed, and the tests and criteria indicated and defined. Methods of treating the data will be reviewed briefly.

Subjects

Two groups of subjects were utilized in this research, a study group (Group S) and a validation group (Group V). The composition of the groups is shown in Table I.

TABLE I

COMPOSITION OF STUDY AND VALIDATION GROUPS

Sex	Group S	Group V
Male	55	30
Female	53	26
Total	108	56

The study group, Group S, consists of those students who graduated from Stillwater High School in the years of 1952 and 1953 and who subsequently enrolled and received marks in two semesters of college work at Oklahoma A. and M. College, and for whom complete data were available

on the DAT administered when the subjects were high school freshmen. Data on this group are to be used to compute the correlation coefficients and the prediction equations indicated earlier. There are 108 students in the group, 55 of whom are male and 53 of whom are female. These will be referred to as the male and female study groups, or Group S Males and Group S Females.

The second group of subjects involved in this study is the validation group, Group V. The purpose in using this second group is to make a longitudinal validation of the regression prediction equations based on the data from the original study group. This validation group consists of all students graduating from Stillwater High School in 1954 who have enrolled and completed two semesters of college work at Oklahoma A. and M. College and for whom complete data were available from the DAT administered when these subjects were high school freshmen. This group contains 56 students, 30 of whom are male and 26 of whom are female. These will be referred to as the male and female validation groups, or Group V Males and Group V Females.

The question of representativeness of the subjects in the study groups in the general trait of academic aptitude, while not bearing directly on the study, is of some interest. Random samples of 55 boys and 53 girls, equal in number to the study groups, were selected from the freshman population at Oklahoma A. and M. College. In order that these samples might be proper comparison groups for the study groups described in the first two paragraphs of this section, only those students who were freshmen during the same years as the study groups and who completed two semesters of work at the college were considered in the sampling. Since all new students at Oklahoma A. and M. College

are required upon entrance to take the American Council on Education Psychological Examination (ACE),[1] a standardized test of academic aptitude, total scores on this examination were obtained from the files of the Bureau of Tests and Measurements for both the study groups and the comparison groups. Resulting data are shown in Table II.

TABLE II

RESULTS OF THE ACE PSYCHOLOGICAL EXAMINATION FOR STUDY AND COMPARISON GROUPS

	Study Groups		Comparison Groups	
	Male	Female	Male	Female
Mean	108.96	98.32	93.35	91.70
Standard deviation	21.52	23.81	20.74	20.60
S. E. of the mean	2.90	3.27	2.80	2.83

There appear to be significant differences between the means of certain of these groups, and comparisons are worthy of note. These are shown in Table III. Also shown are the "t" values obtained in testing the significance of differences between the group means compared.

The mean of the male study group appears to be different from the means of all the other groups, and such is the case. When the difference between the mean of the male study group and the mean of the male comparison group is treated statistically, a "t" of 3.87 is obtained, a value which is significant at the .01 level of confidence. Thus, the mean score of the male study group is significantly above the mean

[1]Published by Cooperative Test Division, Educational Testing Service, Princeton, New Jersey.

score of the randomly chosen comparison group of male freshmen enrolling and completing one year of work at Oklahoma A. and M. College during the same years.

TABLE III

RESULTS OF STATISTICAL TEST FOR DIFFERENCES BETWEEN GROUP MEANS

Groups compared	Difference between means	"t"
Male Study Group and Male Comparison Group	15.61	3.87**
Male Study Group and Female Study Group	10.64	2.43*
Female Study Group and Female Comparison Group	6.62	1.53
Male Comparison Group and Female Comparison Group	1.65	.41

*Significant at the .05 level of confidence.
**Significant at the .01 level of confidence.

The difference between the means of the female and male study groups is also statistically significant. A "t" of 2.43, significant at the .05 level, is obtained when the difference between the means of these two study groups is tested.

When the difference between the means of the female study group and the female comparison group is tested, a "t" of 1.53 results, a value which is not statistically significant. Likewise, there is no statistically significant difference between the means of the male and female comparison groups. As would appear by examining the standard

deviations of the different groups, there are no significant differences in variability among them.

Conditions of Test Administration

The Differential Aptitude Tests were administered, along with other tests and inventories, to freshmen of Stillwater High School in September of 1949, 1950, and 1951 as part of an established orientation program. The test materials and the scoring services were supplied by the Bureau of Tests and Measurements of Oklahoma A. and M. College. Although the aptitude measures are compared with norms and profiles are drawn as counseling aids, all data processed in this research have been treated in raw score form.

The Test Battery

The Differential Aptitude Battery consists of a series of eight different tests. These eight tests are entitled Verbal Reasoning, Numerical Ability, Abstract Reasoning, Space Relations, Mechanical Reasoning, Clerical Speed and Accuracy, Spelling, and Sentences. A brief description of each of these tests including some of the expectations of the test authors concerning them will be given in the following chapter. At that time there will also be presented excerpts of opinions of the battery written by test experts, evidence concerning the reliability of the tests, and validity data gathered from research with the battery.

The Criteria

All criteria employed in this investigation are college grade point averages. The official files of the Registrar, Oklahoma A. and M. College, were consulted in gathering these data. All averages are

computed from the records made by the students in the study group, Group S, during the two semesters of their freshman year at the college. Two types of grade point averages were computed: general averages of grades obtained in all courses taken, called over-all grade average, and averages of grades obtained in different subject matter areas. In computing grade averages the following weights were given to each letter grade: A = 4, B = 3, C = 2, D = 1, and F = 0.

Included in the over-all averages are grades in all courses in which the ordinary five point letter grading system is used. This includes such course areas as military science and physical education.

Separate averages were computed for the fields of language, social science, science, mathematics, and home economics. Included under language were courses in English, speech, journalism, and business correspondence. Social science included history, political science, geography, sociology, and psychology. The category of science included both physical sciences and biological sciences. The areas of mathematics and home economics included the usual courses for freshmen listed under their respective departmental titles in the college catalogue.

Much has been said and written concerning the value or lack of value of grades or school marks as validating criteria. Since grades and ratings are known to be somewhat unreliable and are many times influenced by irrelevant factors, they leave much to be desired as criteria. Cronbach points out some of the weaknesses of grades and ratings as criteria in the validation of aptitude tests:

> A common type of criterion is the rating or grade. Aptitude tests are validated against marks earned in school. These ratings are weak criteria, since the judge may not know the facts about the person, and different judges disagree. When

> a test fails to predict a grade, one cannot say whether this is the fault of the test or of the rating.[2]

Wallace recognizes the criticisms made of grades as criteria, but continues to use grades as the criteria in his study. His reasoning in so doing is given as follows:

> In reporting previous studies on the predictive value of test results some authors have argued that grades were an unsatisfactory criterion of achievement and have sought to substitute a different one, such as achievement test scores. However, scholastic success continues to be judged entirely on the basis of grades, and it is the accuracy of the prediction of this type of success that the study is designed to investigate.[3]

This writer is inclined toward the same resolution of the problem as that reached by Wallace above. It is not the purpose of this study to enter into the controversy concerning the relative merit or lack of merit of grades as criterion measures. This is a study of the relationship of certain aptitude measures made early in the secondary school careers of the subjects with academic success in college some four years later. Grades are assumed to be a measure of academic success. It is not assumed, however, that grades are perfectly consistent, that is, that they have perfect reliability. The lack of perfect reliability of grades as criteria, as well as the lack of perfect reliability of the test instruments, will serve to limit the maximum validity coefficient attainable between the tests and the criteria. This is a point which should be kept in mind as one views the results of any predictive study such as this one.

---

[2]Cronbach, p. 56.

[3]W. L. Wallace, "Differential Predictive Value of the A.C.E. Psychological Examination," School and Society, LXX (July, 1949), p. 24.

Procedure

The purpose of this section is to acquaint the reader with the steps which were taken in gathering the data for this study and the methods used in processing it.

The first step was the identification of the subject groups and the location of the relevant information on each of the members of these groups. Since it is a custom in Stillwater High School, as in most others, to make a notation on the permanent record sheet of pupils who request transcripts to be sent to college registrars preparatory to enrollment, these records were consulted to identify persons graduating from Stillwater High School who had intentions of attending Oklahoma A. and M. College. The files of the high school and the files of the Bureau of Tests and Measurements at Oklahoma A. and M. College, which furnished the scoring services, were then consulted to obtain scores on the DAT for each of the subjects. Those for whom complete records were not available were dropped from the list. Next, the files of the Registrar, Oklahoma A. and M. College, were consulted for the grade records made by the subjects as freshmen in college. Any persons on the list who failed to enroll or failed to complete two semesters of college work were dropped from the list.

It was arbitrarily decided to use the 1952 and the 1953 graduates as a study group on which to compute the various statistical measures of relationship, and to hold the 1954 graduates for later use as a group in the validation of certain findings from the study of the first group. This first group is called the study group, or Group S, and the latter group is called the validation group, or Group V. Boys and girls form two sub-groups of each type and all statistics are computed separately for the sexes.

After the subject groups had been thus identified and the data collected, the next step was that of processing. The Pearson product-moment method was used in computing the correlation coefficients between scores on each Differential Aptitude Test and college freshman over-all grade average. The Pearson product-moment method was used to compute the coefficients of correlation between scores on each of the tests and college freshman grade averages in the fields of language, social science, science, mathematics, and home economics. The Pearson product-moment method was also used in the computation of the test inter-correlations.

The Wherry-Doolittle method of test selection to obtain the highest multiple R was used in calculating the coefficients of multiple correlation between the optimum combination of the tests and over-all grade average. The Wherry-Doolittle technique was also employed in the formation of the multiple regression equations for prediction of college freshman over-all grade average from known scores on the Differential Aptitude Tests.

That this procedure is not unusual in research design is revealed in the following statement by Darley:

> In the design of the ordinary prediction experiment, an attempt is made to use predictors showing relatively low intercorrelations among themselves and relatively high correlations with some criterion of later success. By appropriate statistical treatment, the contribution of each separate predictor can be maximized and weighted into a multiple regression equation that gives the best prediction of the criterion measure. This is essentially an actuarial procedure by which the experimenter hopes to improve, but cannot make perfect, his selection for success in the criterion task.[4]

---

[4]John G. Darley, "The Functions of Measurement in Counseling," In E. F. Lindquist, Educational Measurement (Washington, D. C., 1951), p. 74.

After the multiple regression equations were developed, the test data for members of the validation group were substituted into the prediction equations for their respective sex and over-all grade averages for these persons were predicted. The Pearson product-moment method was used to compute the coefficients of correlation between the predicted over-all grade averages and the actual grade averages made for the two sex groups. This procedure is known as cross-validation. It is useful in establishing the validity of the prediction equations for the population sampled and thus avoiding situations like the following described by Travers and Wallace:

> The Schools of Dentistry selection battery showed some predictive value when studied for a class admitted in 1948, but practically no predictive value for the 1949 freshman class. This study illustrates the need for more careful scrutinizing of the relationship between admission tests and subsequent grades.[5]

However, when a researcher finds that relationships found between variables within one group or sample tend to hold fairly constant in a subsequent sample from the same population, he feels much more confident about his findings. If it is found that the coefficient of correlation between actual and predicted grades for the validation group does not differ significantly from the coefficient of multiple correlation between the test variables and the criterion, then we may feel that we are dealing with relationships which remain fairly stable from sample to sample within the population. If the two coefficients of correlation differ significantly, then we may assume that we are dealing with

---

[5]Robert M. Travers and W. L. Wallace, "Inconsistency in the Predictive Value of a Battery of Tests," Journal of Applied Psychology, XXXIV (May, 1950), p. 239.

relationships which vary, for reasons which may be known, suspected, or unknown, from sample to sample within the same population.

In order to isolate test groupings or factors in the DAT the test intercorrelation data were submitted to factor analysis. Since the authors of the DAT do not claim factor purity for the battery, discovery of such factors should not necessarily be surprising. The centroid method with orthogonal rotation was employed in the factor analysis of the test intercorrelation data and the resulting isolation of test groupings. Coefficients of multiple correlation were then computed between the component tests of each factor and the over-all grade average criterion.

The results of the treatment of the data which has been outlined in this section will be presented in Chapter V.

CHAPTER III

THE TEST BATTERY

This chapter will be concerned with the presentation of relevant information about the battery of tests used in this research study. An attempt will be made to show the need of guidance workers for a battery such as the DAT, to present opinions of prominent test experts on the battery, to present data on the reliability of the tests, to give evidence concerning the validity of the tests from research, and to present evidence concerning the diagnostic value of the battery.

The almost instantaneous popularity of the DAT attests the need for just such a set of measuring instruments in the guidance world. Before the DAT was developed, counselors had to rely on assembled batteries of tests of specific aptitudes for differential measurement of abilities preparatory to counseling. These batteries were rather unsatisfactory for a variety of reasons. This was brought out by Traxler as follows:

> One reason why tests of aptitude in specific fields are not highly satisfactory is that they are designed according to different patterns and standardized on widely different populations. Thus, there is no adequate basis of comparison between the results of the tests in separate areas, and it is often difficult for a counselor to decide in which of several areas an individual possesses the greatest aptitude even when a variety of test scores is available.
>
> Even if valid tests for all occupations were available, no guidance department could give more than a few of them to any individual and the problem of selecting the most appropriate ones would involve a large element of subjective judgment concerning the aptitude of the individual concerned.[1]

---

[1]Traxler, Techniques of Guidance, p. 49.

Super describes the situation similarly, and states that the DAT was developed in response to this need:

> This battery of tests was developed by Bennett, Seashore, and Wesman, in response to widespread feeling among vocational counselors that a major defect in current testing programs is the lack of a uniform base line for the various tests which are used with a given client. ... A student may be at the 65th percentile when compared to liberal arts freshmen on one test, and at the 55th on another, but actually have more ability of the type measured by the second test; the seemingly lower score being due to differences in the normative groups. It is only when the tests in a battery have been standardized on strictly comparable groups, if not the same group, that one can effectively study aptitude or interpret differences within individuals.
>
> Other needs also contributed to the development of this battery. One was the improvement of statistical procedures which made possible the construction of tests which effectively measure narrower aspects of ability than general intelligence.[2]

A recognized source of expert opinion on published tests is Buros' Fourth Mental Measurements Yearbook.[3] In this reference, test experts, working independently, review all of published tests currently in production and use in this country. A prominent test or test battery, such as the DAT, usually receives reviews by several authorities. Excerpts from the three authoritative reviews of the DAT follow:

> The test items show signs of rare ingenuity. The test items have intrinsic interest for students, and the tests are easy and convenient to administer.
>
> The most impressive feature of the DAT is the test manual. This is unquestionably the best test manual published, and its authors received recognition from the Council of Guidance and Personnel Associations in 1951 for the superior job they have done in presenting the information about the tests.

---

[2]Donald E. Super, Appraising Vocational Fitness, p. 368.

[3]Oscar K. Buros, Fourth Mental Measurements Yearbook (Highland Park, New Jersey, 1953).

There is complete information concerning the purposes of the tests, procedures to be used in administering the tests, statistics concerning the tests and their standardization, principles to be considered in interpreting test scores, and normative frames of reference.

The DAT has been carefully developed and standardized by competent authors who have done an excellent job in making information about these tests available to the public. The tests have some validity in predicting success in high school courses, and some evidence has appeared concerning their validity in predicting vocational success.[4]

One of the more valuable tools for sound vocational and educational guidance available today is represented by the DAT.

The positive correlations with school grades are consistent with the general literature on the prediction of educational success. Briefly, the evaluation techniques currently used in our high schools are such that little, if any, differential prediction can be obtained.

The reviewer would like to commend the authors of the DAT for their work to date and to recommend these tests to counselors for use in educational guidance and educational research programs. The results of further work on the problem of the criterion and on the problem of differential prediction of success in educational or vocational activities will be awaited with interest. These may indicate that prediction (differential or simple) of available criterion measures can be accomplished as well by composite scores or groupings of the eight tests as by scores on separate tests.[5]

Many validity data, mostly follow-up are reported. The publishers have presented the zero correlations along with the high ones for the schools where the data was obtained and have consolidated the information obtained from many small samples, giving the test user more stable values for the various validity coefficients.

The norms are far better than those available for most tests. The numbers are large and well distributed by grade, sex, and school.

---

[4]Ralph F. Berdie in Buros, p. 713.

[5]Harold Bechtoldt in Buros, p. 711.

> It is not obvious that with knowledge of validity coefficients of the individual tests, the counselor can reach valid decisions from the battery as a whole. Perhaps the next step is to supply regression equations for these tests where the data permit.[6]

Super comments on the uses of the battery:

> The preliminary evidence concerning the development and standardization of the DAT battery suggests that these tests measure a number of variables which have frequently been found to have vocational significance.
>
> In schools and colleges, when clinical counseling is to be done, that is, when the objective is the study of a counselee in terms of his psychological make-up and its general educational and vocational implications, the battery should prove useful.
>
> Guidance and employment centers which habitually carry on normative studies may also find it worth their while to use this battery of tests in clinical counseling, supplementing it with others which have occupational norms when such data are really needed.[7]

Thus, it appears that the DAT battery is very well accepted by authorities in the field of testing and test construction. Reviewers were especially complimentary of the standardization data, the norms, the test manual, and of the honest and meticulous methods of reporting research used by the test authors. Since much of the research with the test battery has been done by the authors and collaborators and much other research with the battery is reported through the authors, this confidence expressed by the reviewers and other experts is important.

Description of the Tests

The DAT battery consists of a series of eight different tests. These eight tests are entitled Verbal Reasoning, Numerical Ability,

---

[6] Lloyd G. Humphreys in Buros, p. 714.

[7] Super, p. 374.

Abstract Reasoning, Space Relations, Mechanical Reasoning, Clerical Speed and Accuracy, Spelling, and Sentences. A brief description of each of these tests including some of the expectations of the test authors concerning them will be given. A more complete description is found in the test manual.[8]

The Verbal Reasoning test, as its name implies, is designed as a measure of ability to understand concepts framed in words. The test is aimed not at simple vocabulary, but rather toward a measure of the student's ability to abstract and generalize in a verbal context. The analogies form of test item is thought by the test authors to be particularly appropriate to the measuring of reasoning ability. The type of analogies form used in the test is somewhat unique in test construction and is thought to be more reliable in that it reduces the chance of success by guessing to one in sixteen. The test authors state that this test may be expected to predict success in fields where understanding of complex verbal relationships is important, and they consider that academic success in most fields would fall in this category.

The Numerical Ability test is designed to measure understanding of numerical relationships and facility in handling numerical concepts. Prompted by a desire to avoid contamination of their numerical test with language elements, the authors frame the problems in the item type usually known as "arithmetic computation" rather than in what is usually called the "arithmetic reasoning" type of item. They do not believe that the measurement of reasoning is necessarily sacrificed by the use of this form of item. Some of the problems require only computational skill

---

[8] Bennett, Seashore, and Wesman, A Manual for the Differential Aptitude Tests, p. 5.

while others call for an understanding of more complex numerical relationships. By tryout the authors have found the test to be sufficiently complex to be challenging to students at all high school grade levels. They believe it to have predictive value educationally in such fields as mathematics, physical science, and engineering.

The Abstract Reasoning test is intended as a non-verbal measure of the student's reasoning ability. Each problem presents a series of changing diagrams from which the student is required to deduce the principle or rule governing the systematic change. He then gives evidence of his understanding by designating from a multiple-choice arrangement the diagram which should logically follow. The figures are clearly and visibly different so that it is an intellectual task of discernment of the principle of change that is involved and not a measure of visual acuity. The Abstract score is thought to be relevant when the curriculum or vocation requires the perception of relationships among things rather than among words or numbers.

The Space Relations test attempts to measure the ability to visualize a constructed object from a picture of a pattern, combined with the ability to visualize how an object would appear if rotated in three dimensional space. Again, as in the Abstract Reasoning test, no premium is placed on visual discrimination, but rather with judgments of how the objects would look if constructed and rotated. The intent of the Space Realtions test is to measure the ability to deal with concrete materials through visualization. This ability to manipulate things mentally is thought to be important in such fields as drafting, designing, architecture, art, and decoration.

The Mechanical Reasoning test is essentially a new form of the series of Mechanical Comprehension Tests devised previously by Bennett.[9] Each item consists of a pictorially presented mechanical situation with a simply worded question directed toward ascertaining the subject's understanding of the relationships involved. It is believed by the authors that a person who scores high on this test will find it easy to learn the principles of operation and repair of mechanical devices. The test is thought to have predictive value in the technical, manual, and physical science fields. Boys commonly score higher on this test than do girls.

The Clerical Speed and Accuracy test is intended to measure speed and accuracy of performance of a simple perceptual task. The item form used involves perception of a marked letter or number combination from a group in the test booklet and then locating this same combination in a similar group on the answer sheet and marking it. Little or no intellectual difficulty is thought to be involved. Instead, this test places a heavy premium on speed. The ability to do rapidly routine work of the sort required by this test is thought to be important in filing, coding, stock room work, and similar occupations. The authors feel that it is of relatively little importance for most educational purposes. While this latter opinion seems to be substantiated in some validation studies, some contradictory evidence is furnished by Stinson[10] and by

---

[9] G. K. Bennett, Manual, Mechanical Comprehension Test (New York, 1947).

[10] Pairlee J. Stinson, "A Statistical Analysis of the Differential Aptitude Tests for the Purpose of Predicting First Semester Grade Averages of a Freshman High School Group," Unpublished Masters Thesis, Oklahoma A. and M. College (Stillwater, Oklahoma, 1952), p. 11.

the data of this study, both of which find a rather strong relationship between Clerical scores and academic achievement.

The Spelling test of the battery is not particularly new. The words which appear on the test were selected from the lists in Gates' Spelling Difficulties in 3876 Words.[11] An attempt was made also to select those words which were more prominent in everyday vocabulary. The task involved in solving an item is the selection of the correctly spelled form of a word from among a group of common incorrect spellings of the word. The incorrect spellings included as distractors in the items were those which the research of Gates found to be the most frequent errors.

The Sentence Error test is intended to measure the student's ability to distinguish between good and bad grammar, punctuation, and word usage. The item form is that of sentences subdivided into five parts each of which may have an error in grammar, punctuation, or word usage. Since there may be any number of errors in each sentence, this item offers considerable freedom of response.

Together the Spelling and Sentences tests make up what is called the language usage part of the battery. These tests are more nearly achievement tests than any others in the battery. They were included because it was thought that they represent basic skills which are necessary in many professions and vocations. Separate scores are reported for the two tests since it was found that they were not so highly related as to make separate scores meaningless. Taken together they are thought to measure the student's ability to distinguish correct from incorrect English usage. This ability the authors deem important in stenography,

---

[11] A. I. Gates, A List of Spelling Difficulties in 3876 Words (New York, 1937).

business correspondence, journalism, and wherever the written language is a stock in trade.

The Reliability of the Battery

A rather important characteristic of any test is its reliability. This characteristic is particularly important in tests designed for individual guidance. The authors of the DAT have made rather extensive studies of the reliability of the battery. In Table IV are given the average reliability coefficients of the eight tests. These data are based upon approximately one thousand students of each sex. In all cases except for the test of Clerical Speed and Accuracy reliability coefficients of the tests were obtained by the split-half method and corrected by the Spearman-Brown formula. Since the Clerical test places a premium on speed the alternate forms method was utilized to determine its reliability. Each of the coefficients is an average of the four grade level coefficients obtained on grade level groups of similar size. The method of averaging utilized conversion to Fischer's "z" function. As will be noticed, each of the tests seems to have respectable reliability with the exception of Mechanical Reasoning for the girls.

Since the scores used in the present study were obtained at the high school freshman level, it is of interest to note the reliability coefficients reported by the authors of the test battery for this group. These are given in Table V.

Each reliability coefficient in Table V is based upon approximately two hundred fifty students. Again, the split-half method was utilized in obtaining these coefficients except for those associated with the test of Clerical Speed and Accuracy. For this test the alternate forms method

was used. All reliability coefficients appear to be fairly respectable with the exception of that of Mechanical Reasoning for the girls.

TABLE IV

AVERAGE RELIABILITY COEFFICIENTS OF FORM A OF THE DAT[12]

Test	Boys	Girls
Verbal Reasoning	.90	.90
Numerical Ability	.90	.86
Abstract Reasoning	.90	.89
Space Relations	.93	.90
Mechanical Reasoning	.85	.71
Clerical Speed and Accuracy	.87	.87
Spelling	.92	.92
Sentences	.88	.87

TABLE V

RELIABILITY COEFFICIENTS OF FORM A OF THE DAT FOR GRADE NINE[13]

Test	Boys	Girls
Verbal Reasoning	.88	.89
Numerical Ability	.88	.82
Abstract Reasoning	.86	.87
Space Relations	.92	.90
Mechanical Reasoning	.85	.73
Clerical Speed and Accuracy	.83	.84
Spelling	.92	.92
Sentences	.86	.84

[12]Bennett, Seashore, and Wesman, A Manual for the Differential Aptitude Tests, p. 66.

[13]Ibid., p. 66.

Another aspect of reliability is the standard error of measurement. This statistic is a measure of the reliability to be expected of an individual score. The chances are approximately two out of three that the individual's score does not vary from his true score more than the amount indicated by the standard error of measurement. The standard errors of measurement of the eight tests for boys and girls at the ninth grade level are reported in Table VI. The test authors base these figures upon the same groups as the reliability coefficients reported in Table V. These figures appear small enough.

TABLE VI

STANDARD ERRORS OF MEASUREMENT FOR FORM A OF THE DAT FOR GRADE NINE[14]

Test	Boys	Girls
Verbal Reasoning	2.8	2.6
Numerical Ability	2.7	3.0
Abstract Reasoning	3.7	3.7
Space Relations	6.7	6.6
Mechanical Reasoning	5.0	5.4
Clerical Speed and Accuracy	4.2	3.9
Spelling	6.7	6.9
Sentences	5.2	5.4

Townsend reports a study of the reliability of the DAT battery for a group of independent school boys of junior high school age.[15] She reports finding the reliability data to be very close to the reliability coefficients reported by the authors.

---

[14] Ibid., p. 67.

[15] Agatha Townsend, "The Differential Aptitude Tests--Some Data on the Reliability and Intercorrelation on the Parts," Educational Records Bulletin, LIII (January, 1950), pp. 39-47.

The study of relationships made in the present research is between achievement criteria and tests scores obtained some four and a half years earlier. Obviously, if there were no constancy in test results over a similar intervening time little relationship could be expected between these earlier test scores and the criteria. Long term consistency of measurement, of course, is partly a matter of constancy of the trait and partly a matter of test reliability. These considerations have important implications for long term advisement of students. Wesman evaluates the DAT in the light of these considerations:

> Can career counseling make use of early aptitude results, or is it necessary to wait until the student is almost ready to leave the high school? This question involves not only the reliability of the tests but also the constancy of the aptitudes being measured. If the student's relative standing on the tests changes appreciably between his ninth and twelfth grades, short term high school counseling is possible, but long term career counseling based on the tests is not feasible. If, on the other hand, the student's performance is relatively stable, long term planning may be undertaken with greater confidence. Considerable constancy of performance was found in a study in which students who had taken the DAT in the ninth grade were retested in the twelfth grade. Despite different kinds and amounts of practice, in and out of school, in the abilities measured by the tests, the students maintained their relative ranks quite well (range of uncorrected r is .58 to .87; median r is .72). This fact speaks well for the stability with which the tests are measuring, and permits the counselor to think in long range terms as well as to consider immediate goals.[16]

The longitudinal study of the reliability of the tests mentioned above by Wesman was made by Doppelt and Bennett. Further details of their findings are shown in Table VII.

The coefficients reported in Table VII are very respectable stability indices. If corrected for attenuation, that is, if corrected for the instantaneous unreliability of the tests at both administrations, they

---

[16]Alexander G. Wesman, "The DAT, A Five Year Report," Personnel and Guidance Journal, XXXI (March, 1952), p. 168.

would be much higher, in all liklihood, thus revealing a high degree of constancy of the abilities measured.

TABLE VII

COEFFICIENTS OF STABILITY OF THE DAT[17]

Test	Boys	Girls
Verbal Reasoning (VR)	.87	.82
Numerical Ability (NA)	.75	.74
Abstract Reasoning (AR)	.62	.64
Space Relations (SR)	.59	.70
Mechanical Reasoning (MR)	.73	.63
Clerical Speed and Accuracy (CSA)	.68	.58
Spelling (Sp.)	.77	.77
Sentences (Sen.)	.75	.80
N	71	90

The Validity of the Battery

The usefulness of any test ultimately depends on the extent to which it will predict the performance of the persons tested. This means that the test scores must be found to have a high relationship with the performance to be predicted, or a high validity coefficient. The criteria against which tests intended for educational and academic guidance are usually correlated are grade averages and achievement test scores. Although some data are reported on the relationship of DAT scores with achievement test results at the high school level by Doppelt and Wesman,[18]

[17] J. E. Doppelt and G. K. Bennett, "A Longitudinal Study of the DAT," Educational and Psychological Measurement, XI (April, 1951), p. 232.

[18] J. E. Doppelt and A. G. Wesman, "The Differential Aptitude Tests as Predictors of Achievement Test Scores," Journal of Educational Psychology, XLIII (April, 1952), pp. 210-217.

of more immediate concern to this problem is the relationship of these tests with achievement as measured by grade average.

In Table VIII are given the median coefficients of correlation found between the eight tests and high school grade averages in several fields.

TABLE VIII

MEDIAN VALIDITY COEFFICIENTS BETWEEN DAT SCORES AND HIGH SCHOOL COURSE GRADES[19]

Test	English		Science		Soc. Sci.		Math.	Home Ec.
	M	F	M	F	M	F	M	F
VR	.49	.52	.54	.55	.48	.52	.33	.26
NA	.48	.48	.52	.50	.46	.50	.47	.27
AR	.32	.40	.42	.45	.32	.38	.32	.18
SR	.26	.28	.34	.39	.24	.27	.26	.22
MR	.21	.26	.40	.37	.21	.26	.19	.12
CSA	.22	.26	.24	.27	.21	.30	.16	.24
Sp.	.44	.44	.36	.36	.36	.35	.28	.11
Sen.	.50	.53	.45	.52	.43	.49	.32	.15

The individual findings of various studies are obscured in this composite report of relationships, but for the purpose of a quick survey these figures are probably more meaningful than would be the results of any selected studies. The number of different groups upon which these correlations are based vary from twenty-two to forty, the number of individuals from one thousand to four thousand, and the number of state localities from five to eight. This is with the exception of the figures for home economics which were based on some five hundred individuals in

---

[19] Bennett, Seashore, and Wesman, A Manual for the Differential Aptitude Tests, pp. 35-55.

six groups in three localities. Time intervals between testing and criterion evaluations range from less than one to three years, the shorter period of time being the more usual.

In 1952 Wesman generalized on the validity research on the DAT as follows:

> In 1947, the Differential Aptitude Tests were launched. Now, approximately five years and 4,000 validity coefficients later, it may be well to summarize the experience with these tests.
>
> In the first place, one may generalize that course grades are usually best predicted by those tests which an experienced counselor would expect to be the best predictors. Thus Sentences and Verbal Reasoning are the best predictors of grades in English, the Numerical test is most effective in predicting mathematics and bookkeeping grades, social studies have useful predictors in the Verbal, Sentences, and Numerical tests, and science is best predicted by those same three tests with Abstract Reasoning also helpful. Space Relations is effective for mechanical drawing. The Numerical test has been found to predict well in somewhat unexpected courses; it provides fairly good forecasting of grades in English, social studies, mechanical drawing, languages, and even typing.[20]

A study of the DAT for the purpose of predicting first semester grade averages was made by Stinson. The tests were administered to high school freshmen at the beginning of the school year and the criterion over-all grade averages collected at the end of the first semester. Her subjects were one hundred eighty freshmen of both sexes at Stillwater High School. Table IX reveals the relationships which were found.

Utilizing the Wherry-Doolittle test selection method for computing the multiple R, Stinson obtained the results in Table X. The tests are listed in order of strength of contribution, and the increment in R may be noted as each test is added.

---

[20]Alexander G. Wesman, "A Five Year Report," p. 167.

TABLE IX

CORRELATION OF THE DAT WITH GENERAL GRADE AVERAGE AT THE END OF THE FIRST SEMESTER[21]

Test	r
Verbal Reasoning	.45
Numerical Ability	.41
Abstract Reasoning	.38
Space Relations	.45
Mechanical Reasoning	.02
Clerical Speed and Accuracy	.51
Spelling	.48
Sentences	.52

TABLE X

MULTIPLE CORRELATION COEFFICIENTS BETWEEN THE SELECTED TESTS AND GRADE POINT CRITERION[22]

Test	Multiple R
Sentences	.5200
Clerical Speed and Accuracy	.6059
Space Relations	.6797
Spelling	.7038
Numerical Ability	.7087

Going a step further, Stinson also computed a multiple regression equation for predicting first semester grade point average from scores on these five tests.[23] This equation had a standard error of estimate

[21] Stinson, p. 11.

[22] Ibid., p. 12.

[23] Ibid., p. 13.

of .621 grade points. The point system for averaging grades was the same as that employed in the present investigation.

In a prediction study of engineering grades, Berdie[24] used the DAT, high school rank, and certain other standardized tests. The best predictors were high school rank and the Numerical Ability test of the DAT in that order. A multiple R of .62 with grade average in engineering was obtained using these two predictors. The distributions of scores on the DAT for these engineering freshmen were found to be skewed and exhibiting a marked kurtosis. Berdie interpreted this, and reasonably so, as indicating that the tests were not sufficiently difficult for use with this group. He thought his obtained test intercorrelations to be close to those of the authors.

Berdie found the test most closely related to mathematics grades was the Numerical Abilities test with a correlation of .45. Numerical ability was also most closely related to chemistry grades with a coefficient of .35. Space Relations appeared to be most closely related with drawing course grades with a coefficient of .38. In general, Berdie found most correlations with course grades to be somewhat lower than those reported in the manual for high schools.

Wolking made a rather interesting comparative study of the predictive efficiency of the verbal, numerical, and spatial tests of the DAT and the same subtests of the Primary Mental Abilities Test.[25] The following rather extensive excerpts give the essence of his findings:

---

[24]Ralph F. Berdie, "The Differential Aptitude Tests as Predictors in Engineering," Journal of Educational Psychology, XLII (March, 1951), pp. 114-123.

[25]W. D. Wolking, "Predicting Academic Achievement with the Differential Aptitude and the Primary Mental Abilities Tests," Journal of Applied Psychology, XXXIX (March, 1955), pp. 115-118.

The trend in aptitude testing has been toward the special measurement of more or less specific abilities by a battery of tests which has been standardized as a group and which makes use of a profile scoring approach. The principle underlying this type of test battery is that each measurable aptitude is usable in a number of prediction problems; hence, a standard test battery can be constructed and normed in such a way as to yield scores for predicting a number of different criteria. Various methodological approaches have been used in the construction of such a test battery. Two of these approaches have resulted in test batteries which have many similarities, and which have commanded attention in the field of aptitude testing. The older of these is the PMA Test developed by Thurstone as a practical implementation of his factorial studies of intelligence.[26] The more recent battery is the DAT, constructed by Bennett, Seashore, and Wesman through the use of the differential scores approach.

The mean DAT scores for boys and girls in this sample agree closely with the published norms. Agreement of the mean PMA scores with their norms is poor.

The critical ratios of...(Table XI)... give evidence that the Differential Aptitude Battery has higher validity in every case where there is a significant difference between the two batteries.

The data bearing on whether the subtests are most valid for predicting grades of school subjects which are assumed to require the ability measured by that particular test are surprising in some cases. Regardless of sex, all tests are most valid for predicting science grades. Industrial arts and home economics are not predicted well by any of the tests. The number test of the DAT battery is the best over-all predictor of academic success. Its correlations with all subjects except industrial arts and home economics are .55 or higher. The DAT verbal test is the second best over-all predictor of academic success. Both number tests predict English grades as well or better than either of the verbal tests. None of these tests appear to be conspicuously superior at predicting grades in courses generally assumed to require the ability measured by that particular test for success, although there is a tendency for spatial tests to have higher validities for science, geometry, and algebra, whereas the verbal tests have higher validities in general than the spatial in general and are relatively more valid predictors of English grades.

---

[26] L. L. Thurstone and T. G. Thurstone, Chicago Tests of Primary Mental Abilities, Manual of Instructions (Chicago, 1943).

TABLE XI

CORRELATION OF THE DAT AND PMA TEST SCORES WITH SCHOOL GRADES AND SIGNIFICANCE TESTS OF THE DIFFERENCES BETWEEN THE CORRESPONDING SUBTESTS WITH PARTICULAR GRADES[27]

Subjects	Sex	N	DAT	PMA	CR	Diff. favors
Verbal Tests						
English	M	128	.45	.49	---	---
English	F	139	.54	.55	---	---
Algebra	M	126	.41	.31	1.69	DAT
Algebra	F	138	.43	.41	---	---
Geometry	M	73	.66	.53	2.13	DAT
Geometry	F	77	.53	.50	---	---
Ind. Arts	M	115	.05	.11	---	---
Home Ec.	F	132	.38	.30	1.37	DAT
Science	M	127	.61	.55	1.21	DAT
Science	F	137	.66	.66	---	---
Number Tests						
English	M	128	.55	.47	---	---
English	F	139	.58	.44	2.11*	DAT
Algebra	M	126	.56	.32	3.04**	DAT
Algebra	F	138	.58	.44	2.10*	DAT
Geometry	M	73	.62	.30	3.15**	DAT
Geometry	F	77	.67	.23	4.93**	DAT
Ind. Arts	M	115	.18	-.06	2.83**	DAT
Home Ec.	F	132	.46	.27	2.49*	DAT
Science	M	127	.69	.46	3.86**	DAT
Science	F	137	.69	.45	2.93**	DAT
Spatial Tests						
English	M	128	.37	.17	2.58**	DAT
English	F	139	.30	.25	---	---
Algebra	M	126	.40	.22	2.51*	DAT
Algebra	F	138	.36	.35	---	---
Geometry	M	73	.39	.27	1.18	DAT
Geometry	F	77	.39	.26	1.39	DAT
Ind. Arts	M	115	.26	.01	2.84**	DAT
Home Ec.	F	132	.36	.34	---	---
Science	M	127	.52	.24	3.84**	DAT
Science	F	137	.47	.45	---	---

*Significant at .05 level of confidence.
**Significant at .01 level of confidence.

[27] Wolking, p. 117.

Wolking's study was made with a group of 266 juniors in high school. His findings are interesting in that they seem to show the DAT battery relatively superior to another battery of its kind in predictive efficiency. It might be that this difference is due in part to the greater length of the tests of the DAT compared with the subtests of the Primary Mental Abilities Test, thereby making for increased reliability, and, in turn, increased validity.

In general, then, the DAT battery has been shown to have validity in the prediction of course grades at the high school level. Some tests seem to have a high relationship with grades in many courses, and some course grades appear to be predicted best by certain related test scores. In combination, tests from the battery have been shown to have a high multiple correlation with over-all grade average (.71).

The Intercorrelations of the Tests

An essential characteristic of tests used in combination is independence of one another. This is especially true of a guidance battery. If tests correlate highly, it is likely that an individual's score on one will not vary far from his standing on the others and important ability and trait differences will be missed. Bennett and Doppelt comment:

> In the process of educational or vocational guidance, it is the usual practice to employ several tests which are administered to the client for the purpose of revealing his strengths and weaknesses so that he may decide upon a career which will utilize his superior capacities and avoid, so far as possible, dependence upon those characteristics in which he is least able. Consequently, it is desirable to have tests which do not correlate very highly with each other, thereby measuring relatively independent traits. Another characteristic of importance is high reliability so that differences in scores

shall be, so far as possible, true differences rather than accidental ones.[28]

The intercorrelations of the tests reported by the authors of the battery are shown in Table XII. These are average coefficients obtained by utilizing Fisher's "z" function to average the coefficients obtained separately for each of the five grade levels. The total number of students upon which these figures are based number approximately one thousand of each sex. In general, the authors feel the coefficients indicate that abilities measured by the separate tests are sufficiently different to warrant the inclusion of all of the tests in a series.

TABLE XII

AVERAGE INTERCORRELATION COEFFICIENTS OF THE DAT[29]

	VR	NA	AR	SR	MR	CSA	Sp
			Males				
NA	.58						
AR	.56	.54					
SR	.52	.47	.56				
MR	.53	.39	.51	.59			
CSA	.12	.20	.16	.13	.06		
Sp	.51	.41	.28	.19	.17	.13	
Se	.62	.50	.44	.36	.36	.14	.62
			Females				
NA	.58						
AR	.61	.52					
SR	.49	.37	.57				
MR	.54	.41	.53	.54			
CSA	.16	.22	.18	.12	.13		
Sp	.52	.43	.33	.21	.25	.15	
Se	.67	.55	.46	.34	.37	.16	.64

[28] G. K. Bennett and J. E. Doppelt, "The Evaluation of Pairs of Tests for Guidance Use," Educational and Psychological Measurement, VIII (1948), p. 319.

[29] Bennett, Seashore, and Wesman, A Manual for the Differential Aptitude Tests, p. 69.

Counseling based upon the results of a test battery yielding measures of different and relatively independent abilities is done on the basis of strengths and weaknesses which are assumed from differences in scores on the different tests in the battery. For long range advisement, it is important that such differences between test scores exhibit some stability. Relative strengths and weaknesses revealed in present measurement should be expected to appear similarly at some future measurement. The following data concerning the three year stability of differences between scores on the DAT come from a study by Doppelt and Bennett:

TABLE XIII

INTERCORRELATIONS AMONG GRADE NINE AND GRADE TWELVE DIFFERENCES BETWEEN PAIRS OF THE DAT[30]

	NA	AR	SR	MR	CSA	Sp	Se
			Males	(N = 140)			
VR	.45	.38	.40	.62	.71	.45	.21
NA		.20	.47	.61	.53	.51	.29
AR			.33	.50	.48	.51	.19
SR				.34	.55	.72	.11
MR					.66	.74	.62
CSA						.64	.60
Sp							.48
			Females	(N = 183)			
VR	.53	.47	.60	.50	.64	.61	.38
NA		.27	.45	.43	.48	.48	.46
AR			.38	.26	.55	.58	.44
SR				.33	.66	.69	.67
MR					.56	.55	.51
CSA						.50	.57
Sp							.54

[30]Doppelt and Bennett, "A Longitudinal Study of the Differential Aptitude Tests," p. 234.

This table is read in the following manner: For 140 boys, the correlation between the difference of Verbal Reasoning and Numerical Ability test scores in grade nine and the difference between the same test scores in grade twelve was .45. The coefficients, for the most part, are fairly high. In evaluating these coefficients, it must be kept in mind that they are depressed by any instantaneous unreliability present in each of the pair of tests at both administrations. Consideration of these depressing influences lends a great deal more respectability to these coefficients than would be given at first glance. These results afford evidence of the stability of the differential measures of ability made by the DAT. They give the counselor some confidence that those traits which appear strong in the individual at the present will be likely to appear strong in the future.

Summary

The DAT appeared on the guidance scene in response to a felt need by guidance workers for a series of measures of different abilities based upon a common normative population. In the short time in which the battery has been available it has attained a high level of popularity among test users. Test experts, in general, feel that the tests have good possibilities. These experts feel that much research needs to be done with the battery, and commend the authors for their encouragement and reporting of this research.

The tests appear to be reliable with the exception of the Mechanical Reasoning test with girls. Considerable long range stability of scores also appears evident. A great number of validity coefficients between test scores and high school grades have been obtained. While some of these are high and others are low, average coefficients while high

enough to be of predictive value have not been as high as might be desired for accurate individual prediction. The tests have been shown to have definite diagnostic value, in that differences in scores between tests appear to be fairly stable. While correlated with one another, tests do not overlap sufficiently to warrant the exclusion of any from the battery.

CHAPTER IV

REVIEW OF RESEARCH IN PREDICTING COLLEGE SUCCESS

A great deal of research has been done in the area of prediction of academic success both at the secondary school level and at the college level. Almost every type of measurable variable which has the appearance of being at all relevant has been studied for its bearing upon school success. A great many of these variables have been shown to have some relationship to academic success, but very few have been demonstrated to have sufficient relationship to be of predictive value.

Illustrative of the number and kinds of relevant variables found are the results of a study by Carter and McGennis.[1] They selected the one hundred freshmen having the lowest grade average and the one hundred freshmen having the highest grade average at Western Michigan College of Education in 1949. The differences between the two groups on each of forty-six different factors were examined in an effort to ascertain the effectiveness of each factor in differentiating between good and poor students. Of the "t" values obtained in these comparisons, fourteen were significant at the .01 level of confidence. The factors yielding these significant "t" values, in order of magnitude, were: high school grade average, general scholastic aptitude test score, standardized reading test score, standardized grammar test score, estimate of high

---

[1] H. L. Carter and Dorothy J. McGennis, "Some Factors Which Differentiate College Freshmen Having Lowest and Highest Point-Hour Ratios," Journal of Educational Research, XLVI (November, 1952), pp. 219-226.

school principal, standardized vocabulary test score, number of books reported read, terms of acceptance by the college, re-entrance for the second semester, number of units failed in high school, number of periodicals reported read, definite vocational choice, non-preference for physical education curriculum, and units of high school credit in mathematics.

A review of one hundred ninety-four studies related to scholastic success at the college level was made by Garrett.[2] He found that the measures having the greatest predictive value were, in descending order of correlation, high school scholarship, general achievement tests, general college aptitude tests, and special aptitude tests. He found that multiple correlation coefficients of two factors usually resulted in somewhat higher correlation with the criterion than did the factors singly; however, the addition of a third factor was found to add very little.

Certain variables have thus been found to be more pertinent than others for prediction of academic success in college. The remainder of this chapter will be devoted to a review of the research with these major predictors. Consideration will be given high school record, general achievement tests, group intelligence tests, and tests of specific aptitude as predictors of college grades. Consideration will also be given to multiple relationships of more than one of these variables to academic success. Some attention will also be given to research in long-range prediction of college success, and studies in differential prediction will be considered.

---

[2]Harley F. Garrett, "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Sciences and Teachers Colleges," Journal of Experimental Education, XVIII (February, 1949), pp. 91-138.

High School Record as a Predictor

Most investigators report that high school marks generally provide a more accurate basis for the prediction of college scholarship than do intelligence tests and other measures of ability. In 1934, Segel's summary of studies of correlation between college scholarship and average high school marks gave a median coefficient of .55, which was .11 higher than the median coefficient between intelligence test scores and college scholarship.[3]

In 1954, Stone reported the results of a predictive study made at Brigham Young University.[4] He found high school grades to be a better predictor than either the ACE Psychological Examination or the Cooperative General Culture Test.[5]

Travers relates an opinion as to why high school marks are generally found to be superior predictors:

> High school and college averages are more closely correlated than is either with test scores. For the prediction of second semester grades, the first semester grades are by far the best criterion. The value of high school grades for predictive purposes is undoubtedly a result of the fact that they represent a combination of ability and motivational factors operating in much the same way as they will operate in college. The advantages of these circumstances seem to outweigh the factors that tend to reduce the validity of high school grades.[6]

---

[3]David Segel, Prediction of College Success, U. S. Office of Education Bulletin No. 15 (Washington, D. C., 1934), p. 90.

[4]Joics B. Stone, "Differential Prediction of Academic Success at Brigham Young University," Journal of Applied Psychology, XXXVIII (March, 1954), pp. 109-110.

[5]Published by the Educational Testing Service.

[6]Robert M. Travers, "Significant Research on the Prediction of Academic Success," In Wilma T. Donahue, The Measurement of Student Adjustment and Achievement (Ann Arbor, Michigan, 1949), p. 154.

Garrett's extensive review of the research related to college success and its prediction contains some thirty-two studies of the relationship of high school grade average to achievement in college. The coefficients of correlation reported in these respective studies are shown in Table XIV. As will be noted, the range of coefficients of correlation is from .29 to .83 and the median correlation is .56. Reference to individual studies is made by name only in the table, but the full citation may be found in the bibliography at the close of this thesis.

Garrett's review also covers some twenty-six studies relating rank in the high school graduating class to college success. The correlation coefficients reported in these respective studies are reported in Table XV. As will be noted, the range of correlation coefficients is from .26 to .72 with a median correlation of .55. Again, reference to individual studies is made by name only while the full citation is given in the bibliography.

While high school grade average and rank in graduating class have been shown to be effective predictors of academic success in college, other aspects of the high school record have been shown to be ineffective predictors. One of these is the pattern of courses taken in high school. In decrying the formerly common practice of conditioning college admission upon the completion of certain prescribed units in high school, Douglass states:

> Not only do the results of this study indicate that such entrance requirements contribute practically nothing to the differentiation of good from poor student risks in college, but all other studies of a similar nature agree in this respect. In fact, no record can be found by any thorough-going objective investigation ever conducted which affords any rational support for the practice of conditioning

TABLE XIV

STUDIES IN THE CORRELATION OF HIGH SCHOOL AVERAGE WITH COLLEGE AVERAGE (AFTER GARRETT[7])

Date	Reference	Institution	N	r
1921	Scates	Chicago Univ.	1707	.61
1921	Cocking & Holy	Ohio Univ.	266	.55
1922	Beatley	Harvard Univ.	423	.56
1922	Henderson	Columbia Univ.		.45
1922	"	Carnegie Inst.		.29
1922	"	Ohio Univ.		.38
1923	Anderson & Spencer	Yale Univ.	402	.44
1922	Symonds	Hawaii Univ.		.55
1923	Sommers	Colorado State	212	.83
1924	Proctor	Standford Univ.	473	.45
1924	Odell	Illinois Univ.	2000	.55
1924	"	" "	2000	.54
1927	Douglass	Oregon Univ.	385	.56
1927	Jones	Indiana State	325	.56
1927	"	" "	325	.60
1928	Pierson	various coll.	50	.52
1928	Hartson	Oberlin Coll.		.55
1928	"	" "		.46
1929	Whitney	Colorado State	899	.50
1930	Prosser	Iowa Univ.	280	.51
1931	Crawford & Burnham	Yale Univ.	3277	.57
1931	Patterson	Minnesota Univ.	309	.31
1931	"	" "	229	.45
1932	Hartson	Oberlin Coll.	150w.	.47
1932	"	" "	120m.	.45
1934	Finch & Nemzek	Minnesota Univ.	90	.79
1935	Byrns & Henmon	Minnesota Univ.	250	.72
1936	Read	Wichita Univ.	400	.63
1936	"	" "	400	.64
1936	"	" "	400	.67
1935	Garrett	52 Colleges	200	.67

Number of coefficients 32
Range .29 to .83
Median correlation .56

[7]Harley F. Garrett, p. 98.

TABLE XV

STUDIES MADE OF STUDENTS' RANK IN HIGH SCHOOL GRADUATING CLASS AND THEIR GRADES IN COLLEGE (AFTER GARRETT[8])

Date	Reference	Institution	N	r
1920	Goldthorpe	Northwestern Univ.	136	.62
1921	Ellefson	California Univ.	143	.56
1921	Johnson	Minnesota Univ.		.63
1923	"	" "		.72
1923	"	" "		.68
1928	Fricken	Macalester Coll.	126	.69
"	"	" "	132	.69
"	"	" "		.68
"	"	" "		.67
1929	Crawford	Yale Univ.	501	.61
1931	"	" "	3277	.57
1926	Williamson	Minnesota Univ.	300m.	.68
1926	"	" "	207w.	.63
1931	Segal	Long Beach J.C.	90w.	.42
1931	Patterson	Minnesota Univ.	300	.27
1931	"	" "	229	.49
1927	Douglass	Oregon Univ.		.48
1933	Williamson & Freeman	Minnesota Univ.	379	.47
1934	" "	" "	951	.45
1934	" "	" "	283	.52
1934	Douglass & Lovegren	" "	190	.43
1935	" "	" "	827	.54
1936	Martin	Trenton State Coll.	228	.26
1936	"	" " "	123	.30
1937	Butsch	Marquette Univ.	750	.56
1939	Seyler	Illinois Univ.	3006	.60

Number of coefficients	26
Range	.26 to .72
Median coefficient	.55

[8] Ibid., p. 101.

general admission to the university upon the completion of prescribed units of certain favored fields.[9]

The Eight-Year Study of the Progressive Education Association indicates that the specific high school subject matter patterns required by a large number of colleges as prerequisites to admission are not necessarily essential for the academic success of students.[10] The study included 1475 pairs of students, with the control group conforming to conventional subject matter patterns and the experimental group taking a program based on student needs. Follow-up in college showed that the experimental group, among other things, had earned a slightly higher total grade point average and a higher percentage of non-academic honors. There is, on the basis of these studies, little evidence for patterns of courses as predictors of college success.

General Achievement Tests As Predictors

General achievement tests have usually been next to high school scholarship as a satisfactory basis for predicting college scholarship. Segel, grouping the results of many studies, reported a median correlation of .54 between general achievement test results and college scholarship.[11] Durflinger's summary of some twenty studies showed a median correlation of .48.[12] From a study of three samples of public

---

[9]Harl R. Douglass, "Relation of Pattern of High School Credits to Scholastic Success in College," North Central Association Quarterly, VI (May, 1931), p. 296.

[10]W. M. Aiken, The Story of Eight Year Study (New York, 1942).

[11]Segel, p. 47.

[12]G. W. Durflinger, "The Prediction of College Success," Journal of the American Association of College Registrars, XIX (June, 1943), pp. 68-78.

school graduates at Harvard, Chauncey and Fredericksen report correlation coefficients of .55, .55, and .56 between average scores on the achievement tests of the College Entrance Examination Board and freshman grades.[13]

Garrett's review of the research related to the prediction of college success covers twenty-four studies of the relationship of general achievement. The coefficients of correlation reported in these respective studies are shown in Table XVI. The range of coefficients obtained is from .23 to .85 with a median correlation of .49. Citations of the individual studies contained in this tabular summary of Garrett may be found in the bibliography.

A summary of several summaries of correlational studies of the predictive power of general achievement tests was compiled by Durflinger and is shown in Table XVII. There appears to be a great deal of agreement among the summaries.

Thus, general achievement tests appear to be effective predictors of college achievement as demonstrated by the median correlations of .55, .54, .55, .48, and .49 reported by different authors who have summarized the research.

Intelligence Tests as Predictors

Few problems in education or psychology have received as much attention as that of the relation of intelligence or scholastic aptitude to college achievement. A summary by Segel showed a median correlation

---

[13]Henry Chauncey and Norman Fredericksen, "The Function of Measurement in Educational Placement," In E. F. Lindquist, Educational Measurement (Washington, D. C., 1951), p. 92.

TABLE XVI

STUDIES IN THE USE OF GENERAL ACHIEVEMENT TESTS IN PREDICTING GRADE AVERAGE IN COLLEGE
(AFTER GARRETT[14])

Date	Reference	Institution	Test	N	r
1922	Beatley	Harvard Univ.	CEEB Tests	423	.50
1922	"	" "	" "	423	.46
1923	Arlitt	Bryn Mawr	Col. Ent. Ex.	103	.27
1923	"	" "	" " "	104	.29
1926	Condit	Colorado State	Local Achv.	559	.46
1926	"	" "	" "	559	.49
1927	Jones	Indiana State	Ind. Comp. Achv.	325	.63
1928	"	" "	" " "	325	.85
1928	Reeves	Oberlin College	Scones-Harry Ach.	750	.59
1929	Crawford	Yale Univ.	CEEB Test	591	.64
1930	Prosser	Iowa Univ.	Qualifying Ex.	280w.	.65
1931	Segal	Long Beach J.C.	Iowa H.S. Cont.	90	.23
1936	Reed	Wichita Univ.	Iowa H.S. Cont.	400	.44
1936	"	" "	" " "	400	.48
1936	"	" "	" " "	400	.43
1941	Weber	Wells College	CEEB Test	59	.47
1941	"	" "	" "	48	.55
1941	Butsch	Marquette Univ.	Iowa H.S. Cont.	750	.47
1941	Weaver	Mount Holyoke	CEEB (Essay)	96	.57
1942	"	" "	CEEB (Achiev.)	261	.52
1941	Weber	Wells College	CEEB Old Type		.46
1941	"	" "	Regents' Exam.		.51
1941	"	" "	CEEB New Apr. Ex.		.41

Number of coefficients	24
Range of coefficients	.23 to .85
Median coefficient	.49

[14] Harley F. Garrett, p. 111.

TABLE XVII

MEDIAN COEFFICIENTS OF CORRELATION BETWEEN ACHIEVEMENT EXAMINATIONS AND COLLEGE SCHOLARSHIP FROM SUMMARIES MADE BY DIFFERENT AUTHORS (AFTER DURFLINGER[15])

Author	Date	Number of Studies	Median r
Harl R. Douglass	1931	67	.55
David Segel	1934	13	.54
Mazie E. Wagner	1934	88	.55
G. W. Durflinger	1942	20	.48

of .44 between intelligence as measured by group tests and college scholarship.[16] Durflinger's summary in 1943 indicated a median correlation of .52 between group intelligence test scores and college marks.[17]

Garrett's summary of twenty-five studies of the relationship of group intelligence test scores to college grades is shown in Table XVIII. The range of coefficients obtained in these studies is from .21 to .67 with a median coefficient of .47. Complete citations of tabular references may be found in the bibliography.

Durflinger's report of several summaries made by different authors of the relationship of measured intelligence to college success is shown in Table XIX.

---

[15]Durflinger, p. 74.

[16]Segel, p. 62.

[17]Durflinger, p. 75.

TABLE XVIII

STUDIES IN THE CORRELATION OF INTELLIGENCE WITH COLLEGE GRADE AVERAGE (AFTER GARRETT[18])

Date	Reference	Institution	Test	r
1929	Gerberich	Arkansas Univ.	A.C.E. Psych.	.45
1930	"	" "	" "	.58
1931	"	" "	" "	.55
1930	Fritz	Kans. S.T.C., Pitts.	A.C.E. Psych.	.53
1930	Guiler	Northwestern U.	Army Alpha	.43
1931	Segal	Long Beach J.C.	Thurstone	.48
1931	Douglass	Oregon Univ.	A.C.E. Psych.	.45
1932	Harston	Oberlin Coll.	Ohio St. Psy. Exam.	.55
1932	Nelson	Iowa S.T.C.	A.C.E. Psych.	.67
1932	Fleming	"Colleges"	Thorndike	.37
1935	"	"	A.C.E. Psych.	.50
1935	"	"	Ohio St. Psy. Exam.	.46
1933	Edds & McCall	Milligan Coll.	Otis Group	.50
1934	Douglass	Minnesota Univ.	A.C.E. Psych.	.50
1936	Read	Wichita Univ.	Ohio St. Psy. Exam.	.42
1937	Butsch	Marquette Univ.	A.C.E. Psych.	.53
1939	Prescott	Phoenix J.C.	Otis Self-Admin.	.21
1939	Dubois	New Mexico U.	A.C.E. Psych.	.44
1940	Attender	New Jersey S.T.C.	Henmon-Nelson	.35
1940	Garrett	52 Colleges	Ohio St. Psy. Exam.	.61
1941	Votaw	S.W. Texas State	A.C.E. Psych.	.53
1944	Weber	Wells College	A.C.E. Psych.	.45
1945	Smith	Fresno State	A.C.E. Psych.	.42
1945	"	" "	" "	.38
1947	Bent	Arkansas Univ.	Otis Self-Admin.	.63

Number of coefficients	25
Range of coefficients	.21 to . 67
Median coefficient	.47

[18] Harley F. Garrett, p. 113.

TABLE XIX

SUMMARY AVERAGES OF COEFFICIENTS OF CORRELATION REPORTED BETWEEN INTELLIGENCE TESTS AND COLLEGE GRADES (AFTER DURFLINGER[19])

Author	Date	Number of Studies	Median r
Harl R. Douglass	1931	160	.45
L. B. Kinney	1932	442	.45
David Segel	1934	100	.44
Mazie E. Wagner	1934	39	.45
G. W. Durflinger	1942	23	.52

Median correlations between intelligence test scores and college success of .45, .45, .44, .45, .52, and .47 have been reported in summaries by different authors. These figures confirm intelligence test scores as effective predictors of academic achievement in college.

Intelligence Tests as Longitudinal Predictors

A few attempts have been made to relate intelligence test scores obtained early in the school life of a student with success in college. A study by Adams, reported in 1940, revealed that a greater number of those elementary school children with intelligence quotients above 100 will enter college than those who are below this figure.[20] Byrns and Henmon found the correlation between intelligence quotients derived from

---

[19]Durflinger, p. 76.

[20]F. J. Adams, "College Degrees and Elementary School Intelligence," Journal of Educational Psychology, XXXI (June, 1940), pp. 360-368.

group tests given in grade four to eight and scores on psychological tests given at college entrance to be .81, or nearly the same as the relationship between scores on two group different intelligence tests given at the same time.[21] Thomson administered the A.C.E. Psychological Examination to high school pupils and retested these same pupils two years later upon their entrance to college as freshmen.[22] Correlations of the results of the two administrations with college marks were .57 and .58, respectively, indicating that the test could have been administered in high school or college with similar results.

Rosenfeld and Nemzek found a correlation of .21 between intelligence quotients obtained in the first grade and honor point ratio in college.[23] Keys found a correlation of .35 between college grades and intelligence quotients measured by group tests before the age of fifteen.[24] Bilhartz found similar results.[25] Finch and Nemzek found a correlation of .48 between intelligence quotients obtained on tests administered before the ninth grade and subsequent honor point averages.[26] Similarly, Byrns and

---

[21] Ruth Byrns, "Long Range Prediction of College Achievement," School and Society, XLI (June, 1935), pp. 877-880.

[22] W. A. Thomson, "Note on Retest Results on the A.C.E. Psychological Examination for College Freshmen," Journal of Educational Psychology, XXXI (June, 1940), pp. 229-233.

[23] M. A. Rosenfeld and C. L. Nemzek, "Long Range Prediction of College Marks," School and Society, XLVII (March, 1938), pp. 127-128.

[24] N. Keys, "The Value of Group Test I.Q.'s for Prediction of Progress Beyond High School," Journal of Psychology, XXXI (February, 1940), pp. 81-93.

[25] W. H. Bilhartz, "Determining College Ability During Junior High School Years," School and Society, LIII (November, 1941), pp. 547-552.

[26] F. H. Finch and C. L. Nemzek, "Prediction of College Success from Data Collected during the Secondary School Period," Journal of Applied Psychology, XVIII (September, 1934), pp. 454-460.

Nemzek studied a group of college students who had been given a group intelligence test at some time between the third and eighth grade and found a correlation of .45 with first semester grades, and of .37 with four-year average grades at the University of Wisconsin.[27] These results are summarized in Table XX.

TABLE XX

SUMMARY OF STUDIES BY DIFFERENT AUTHORS OF THE LONGITUDINAL RELATIONSHIP OF INTELLIGENCE TEST SCORES WITH COLLEGE SUCCESS

Author	Date	r
Finch and Nemzek	1934	.48
Byrns and Nemzek	1935	.45
Rosenfeld and Nemzek	1938	.21
Keys	1940	.35
Bilhartz	1941	.35

The range of correlation coefficients obtained in these longitudinal studies of the relationship of intelligence test scores with college success is from .21 to .48 with a median correlation of .35. Generalizing from these results, it may be said that group intelligence tests administered at the upper grade or secondary school level do have some predictive relationship with college grades, but not as much as do those group intelligence tests administered on college entrance. In general, however, it appears that the later such tests are administered in the

[27] Ruth K. Byrns, "Long Range Prediction of College Achievement," School and Society, XLI (June, 1935), pp. 877-880.

secondary school life of the student, the smaller becomes the difference in predictive power. Differences appear to be rather minute between the predictive power of intelligence tests administered in the final two grades of high school and those administered at college entrance.

Multiple Variables as Predictors

Numerous studies have demonstrated that a combination of several factors may be considerably more valuable in predicting general scholarship than any factor alone. Edds and McCall, employing a combination of average high school marks, Otis Group Intelligence Test, and Cross English Test scores, obtained a multiple correlation of .81 with general scholarship.[28] A study indicating that non-verbal tests contribute to some extent to the prediction of grade point is reported by Heston who used seventeen performance tests along with the Ohio State Psychological Examination.[29] For the combination of scores he found a multiple correlation of .65 with grade point average at the end of the first quarter of college work.

A study by Fredericksen and Schrader involved the calculation of validity coefficients for high school rank and the A.C.E. for 4030 freshmen in twelve colleges.[30] In each case the criterion was the first year college grade average. Findings were that the median correlation

---

[28] J. H. Edds and W. M. McCall, "Predicting the Scholastic Success of College Freshmen," Journal of Educational Research, XXVII (October, 1933), pp. 127-130.

[29] J. C. Heston, "The Use of Non-verbal Tests in Prediction of Academic Success," Journal of Educational Psychology, XXXIII (May, 1942), pp. 308-314.

[30] Norman Fredericksen and W. B. Schrader, "The A.C.E. Psychological Examination and High School Standing as Predictors of College Success," Journal of Applied Psychology, XXXVI (May, 1952), pp. 261-265.

of the A.C.E. with grades was .47, and the median correlation of high school standing with the same criterion was .57. The use of a weighted composite of A.C.E. score and high school standing provided a useful prediction of freshman grade average; the median multiple correlation being .60 for veteran students and .68 for non-veterans.

Webb and McCall, at Emory University in Atlanta, found a multiple R of .753 with freshman grade average using the A.C.E., high school average, and a local mathematics placement test.[31] Wallace administered a battery of three commercially published tests and three locally constructed tests to University of Michigan freshmen.[32] The commercial tests were the A.C.E., the Cooperative English Test, and the Iowa Foreign Language Aptitude Test; the local tests included a mathematics placement test, a visualization test, and a vocabulary test. The multiple R between average grades for the first semester and the combination of test variables was .554.

Hartson, using an N of over 500, obtained a multiple R of .71 with college success using high school grades, intelligence test scores, and principal's ratings as factors.[33] Williamson, using an N of 122, obtained a multiple R of .71 between the factors of high school grades, intelligence test scores, and personality ratings by teachers and

---

[31]Sam C. Webb and J. M. McCall, "Predictors of Freshmen Grades in a Southern University," Educational and Psychological Measurement, XIII (October, 1953), pp. 660-663.

[32]W. L. Wallace, "The Prediction of Grades in Specific Courses," Journal of Educational Research, XLIV (September, 1951), pp. 587-597.

[33]L. D. Hartson, "Further Validation of Rating Scales Used with Candidates for College Admission," School and Society, XLVI (September, 1937), pp. 155-160.

college success.[34] Starrak found a multiple coefficient of .87 between four year college marks and the factors of high school grades, intelligence test scores, and freshman grades.[35] Byrns, with an N of 250, found a multiple correlation of .67 using high school sophomore grades and intelligence test scores obtained in elementary school.[36]

Gray, using an N of 756, found a multiple R of .71 between the factors of high school grade average and intelligence test scores and college achievement.[37] Messenger, using an N of 583, obtained an R of .66 between the factors of high school grades and placement test scores and freshman success at a state teachers college.[38] Byrns and Henmon, using an N of 1825, found a multiple R of .63 between freshman success in college and the factors of high school grades, intelligence test scores, and rank in high school graduating class.[39] Crawford and Burnham, using an N of 3277 Yale students, found a multiple correlation coefficient of .69 between freshman grade average and the combined factors of high school grades, scholastic aptitude, mathematics aptitude, College Entrance Examination Board test average, and age.[40]

---

[34] E. G. Williamson, "An Analysis of Scholastic Aptitude of Freshmen," School and Society, XXXIV (November, 1931), pp. 674-680.

[35] J. A. Starrak, "Matching Ability with Achievement," Journal of Higher Education, VIII (June, 1937), pp. 315-320.

[36] Ruth K. Byrns, "Scholastic Aptitude and Freshman Achievement," School and Society, XXV (May, 1932), pp. 713-718.

[37] W. S. Gray, Provision for the Individual in College Education (Chicago, 1932), p. 57.

[38] H. R. Messenger, "The Probability-Table, A Possible Means for Early Elimination of Poor Students from Teachers' Colleges," Northern Illinois State Teachers College Quarterly, VII (May, 1930), pp. 23-29.

[39] R. K. Byrns and V. A. C. Henmon, "Entrance Requirements and College Success," School and Society, XLI (February, 1935), pp. 101-104.

[40] A. B. Crawford and P. S. Burnham, "Entrance Examinations and College Achievement," School and Society, XXXVI (June, 1932), pp. 344-352; 378-384.

The range of multiple coefficients obtained in the studies which have just been reviewed is from .55 to .87 with a median coefficient of .68. Travers, in reviewing a number of similar studies, obtained a median multiple R of .69.[41] Garrett summarized the results of fifty-nine studies using two variables to obtain multiple correlations with college grades.[42] He found a range of coefficients from .47 to .71 with a median R of .58. These findings seem to indicate that combinations of factors are much better predictors of college success than are their components taken singly.

Studies in Differential Prediction

The accuracy of predicting success in specific subjects has been considered in several studies. Several types of measures have been used. Those measures occurring most frequently in reported studies are general mental tests, general achievement tests, and tests of specific traits, aptitudes, or achievements. Segel, in summarizing the data bearing on this problem, reported that the correlations between general mental tests or general achievement tests and scholarship in specific subjects are lower than those between these same tests and general college scholarship.[43] Correlations between specific traits, aptitudes, and achievement tests and success in specific subjects are approximately the same as those between mental tests and general scholarship. Segel reports median coefficients of correlation between tests of specific aptitude

---

[41]Robert M. Travers, "Significant Research on the Prediction of Academic Success," In Wilma T. Donahue, p. 158.

[42]Harley F. Garrett, p. 117.

[43]David Segel, p. 86.

and scholarship in these same fields. The median coefficient between English grades and aptitude is .42; mathematics, .42; science, .45; and social science, .44.

Carlin compared cumulative scholastic averages at graduation with percentile rank obtained as freshmen on the following tests, using a five by five analysis: A.C.E. Psychological Examination, Cooperative English Test, California Test of Personality, and Cooperative Achievement Tests when applicable.[44] The A.C.E. L-score was found to be a better predictor of achievement in English and literature than the Q-score, but the Cooperative English Test is better than either. The Q-score was found to be a better predictor of success in mathematics courses than the L-score, but this does not apply to science courses.

The 1947 edition of the A.C.E. was administered to 350 freshman students at the University of Wisconsin and test scores correlated by Hoerres with first year grade averages.[45] Correlations of grades with Q-scores ranged from .13 in art to .37 in physics and chemistry, with L-scores from .16 in art to .45 in English, and total scores from .06 in art to .54 in physics.

Brown correlated Q-scores, L-scores, and total scores on the A.C.E. with grade point averages in quantitative subjects (mathematics and science), linguistic subjects (English, social sciences, and language), and also with all subjects.[46] The highest correlation obtained was

---

[44]L. C. Carlin, "A Longitudinal Comparison of Freshman-Senior Standing," Journal of Educational Research, XLVII (May, 1953), pp. 285-290.

[45]Mary A. Hoerres and J. D. O'Dea, "Predictive Value of the A.C.E.," Journal of Higher Education, XXV (March, 1954), pp. 97-99.

[46]Hugh S. Brown, "Differential Prediction by the A.C.E.," Journal of Educational Research, XLIV (March, 1950), pp. 116-121.

between L-score and linguistic subjects (.54). The total or T-score proved to be a better predictor of grades in quantitative subjects (.40) than did the Q-score (.30).

In 1944 Goodman reported a study in the prediction of success of freshman engineers at Pennsylvania State College using the Thurstone P.M.A. tests.[47] The correlation coefficients obtained, based on 170 subjects, are shown in Table XXI.

TABLE XXI

CORRELATIONS AND MULTIPLE CORRELATIONS OF THE P.M.A. TESTS WITH COLLEGE COURSE GRADES[48]

Ability	Sem. Average	Chemistry	English	Mathematics
Perceptual	.04	.07	.05	.04
Number	.32	.27	.26	.27
Verbal	.33	.32	.44	.16
Space	.23	.19	.11	.25
Memory	.10	.04	.23	-.05
Induction	.34	.23	.21	.29
Deduction	.38	.41	.21	.44
NVSID	.51			
NVID		.49		
NVMID			.49	
NSID				.49

The highest single correlations obtained were .44 for Deduction with mathematics and .44 for Verbal with English. The best predictors of semester average were Deduction, Induction, Verbal, Number, and Space,

---

[47] C. H. Goodman, "Prediction of College Success by Means of the Thurstone Primary Mental Abilities Test," Educational and Psychological Measurement, IV (February, 1944), pp. 125-140.

[48] Ibid., p. 132.

in that order. The best predictors of chemistry grades were Deduction, Verbal, Number, and Induction, in that order. The best predictors of English grades were Verbal, Number, Memory, Induction, and Deduction, in that order. The best predictors of mathematics grades were Deduction, Induction, Number, and Space, in that order. With a few exceptions, scores on single tests were not effective predictors of grades in these specific courses. In combinations, however, multiple coefficients of .49 or better were obtained between selected tests and grades in each of the courses.

Taylor administered the General Aptitude Test Battery of the United States Employment Service to 479 seniors at the University of Utah graduating with majors in various course areas.[49] He correlated the results of the tests in the battery with grade averages according to major curricula. His findings are shown in Table XXII

TABLE XXII

VALIDITY COEFFICIENTS OF GENERAL APTITUDE TEST BATTERY SCORES FOR SENIORS MAJORING IN VARIOUS UNIVERSITY AREAS[50]

Test	Bi. Sci.	Bus.	Educ.	Engr.	Soc. Sci.
G - Intelligence	.31	.51	.37	.52	.54
V - Verbal	.41	.61	.34	.35	.53
N - Numerical	.22	.37	.35	.35	.45
S - Space	.15	.20	.16	.25	.20
P - Form Perception	.26	.09	.18	.30	.21
Q - Clerical Perception	.36	.31	.35	.33	.54
A - Aiming	.26	.09	.03	.08	.25
T - Motor Speed	.26	.13	.16	.04	.30
N	52	90	123	92	85

[49]Calvin W. Taylor, "G.A.T.B. Patterns for College Success," Occupations, XXIX (April, 1951), pp. 518-526.

[50]Ibid., p. 522.

Generally speaking, the general intelligence, verbal, and numerical factors appeared to have good validity in each of the reported college areas of study. Clerical perception also appeared to have good validity in each of the areas. The space factor seemed to have low validity in all areas, with the greatest relationship being a correlation of .25 with grades of engineering graduates. Form perception appeared to be of some, but not great, importance in biological science and engineering. Aiming or eye-hand coordination appeared of some, but not great, importance in biological sciences and social sciences.

For biological science the best predictors were the verbal, clerical perception, and general intelligence factors, in that order. For business the best predictors were the verbal, general intelligence, numerical, and clerical perception factors, in that order. For education the best predictors were the general intelligence, numerical, clerical perception, and verbal factors, in that order. For engineering the best predictors were the general intelligence, numerical, verbal, clerical perception, form perception, and space factors, in that order. For social sciences the best predictors were the general intelligence, clerical perception, verbal, and numerical factors, in that order.

While the findings of these studies concerning the relationship of ability factors to success in different subject matter fields tend to indicate that certain factors are good predictors in all fields, there appears to be enough differentiation in the patterning of abilities important in the different areas of study to make differential prediction possible. Although no multiple relationships between combinations of factors and success in different course areas were reported by Taylor, it is entirely conceivable that relatively high multiple correlation

coefficients might have been obtained between various optimal combinations of factors and success in the different major curricula.

In summary, the following generalizations may be made: (1) Certain specific aptitudes, such as verbal and numerical, appear to be good predictors in most areas of study. (2) Certain specific aptitudes appear to be of different importance as predictors in different areas of study. (3) In course areas where similar specific aptitudes appear to be predictors, the relative importance of these aptitudes as predictors varies. Differential prediction in this case becomes a matter of differential weighting of the factors. (4) Combinations of aptitude factors are better predictors than are the factors taken singly.

Summary

Research indicates that high school scholarship is the best single predictor of college scholarship followed, in order of importance, by general achievement tests, tests of general intelligence, and tests of specific aptitude. Achievement test scores tend to correlate higher with college success in earlier studies than in later studies. For longitudinal prediction, intelligence tests assume a role of greater importance and appear to be of definite predictive value. At the present, factor scores on tests of differential abilities appear to be inferior to the above mentioned measures in the prediction of general academic success and somewhat inferior to general intelligence tests as predictors of success in specific areas of study. However, there has, as yet, been little research with the newer tests of this type, and it may well be that as research appears dealing with patterning and composite weighting problems, these tests may be shown to have real value

in both general and differential prediction. This appears to be the trend in the research on the prediction of academic success, and the present study is of this nature.

CHAPTER V

PRESENTATION AND ANALYSIS OF RESULTS

The purpose of this chapter is to present the data obtained and to analyze the results of the present study. The test intercorrelation matrices will be presented, and the results of factor analysis of these matrices will be shown. The correlation coefficients obtained between scores on each of the Differential Aptitude Tests and the over-all college grade average criterion will be presented, as will the multiple correlations with the same criterion of the test groupings revealed by factor analysis. The multiple correlation coefficients between the optimum combination of tests and over-all grade point average will be noted, and the regression equations for prediction of this criterion from known test scores will be given. In addition, the coefficients of correlation between scores on each of the tests and grade averages in the fields of language, science, social science, mathematics, and home economics will be presented and the data examined for patterns.

The Test Intercorrelation Matrices

After the test scores and grade averages of the members of the study group were obtained as discussed in Chapter II, the next step in the procedure was to have these data punched on International Business Machine cards. The decision to use International Business Machine procedures was based on the premise that the calculations could be made much more rapidly and that there could be little doubt about their accuracy.

The first step in obtaining the intercorrelation matrices was to compute (by I. B. M.) the summations, the summations of the squares, and the summations of the cross products for each of the tests involved in the study. Utilizing this information, zero order correlations were computed between all tests of the battery. The intercorrelation data for the male and female study groups are presented in Table XXIII and Table XXIV respectively.

TABLE XXIII

INTERCORRELATIONS OF THE DAT FOR THE MALE STUDY GROUP

Test	1	2	3	4	5	6	7	8
Means	25.04	20.64	34.16	56.91	45.05	51.53	34.89	36.91
Sigmas	7.92	8.12	8.01	18.77	7.81	11.58	24.48	15.37
1. Verbal		.402	.432	.490	.465	.391	.694	.685
2. Numerical			.561	.307	.427	.416	.518	.428
3. Abstract				.316	.425	.382	.392	.417
4. Space					.493	.143	.287	.393
5. Mechanical						.067	.376	.423
6. Clerical							.346	.231
7. Spelling								.425
8. Sentences								

N = 55

The highest intercorrelations for the male group appear to be between scores on the Verbal and Spelling tests and between scores on the Verbal and Sentences tests. The size of these coefficients (.694 and .685, respectively) indicates that these tests are tapping a common ability or

abilities which are highly related. These coefficients of intercorrelation are slightly higher than those of the test authors reported in Chapter III of the current study. As was mentioned earlier, the Spelling and Sentences tests are in reality achievement tests, and, since these tests have such a high relationship with the verbal test of the aptitude series, it is presumed that achievement in spelling and gramatical construction may rest somewhat on verbal aptitude as measured by the Verbal Reasoning test.

TABLE XXIV

INTERCORRELATIONS OF THE DAT FOR THE FEMALE STUDY GROUP

Test	1	2	3	4	5	6	7	8
Means	24.32	18.79	32.00	49.00	28.55	59.98	48.25	42.98
Sigman	8.57	6.21	8.12	17.77	8.07	9.52	22.98	13.13
1. Verbal		.441	.442	.315	.343	.097	.301	.711
2. Numerical			.398	.256	.226	.092	.304	.425
3. Abstract				.676	.423	.204	.068	.423
4. Space					.448	.216	.040	.370
5. Mechanical						.193	-.087	.293
6. Clerical							-.115	.047
7. Spelling								.543
8. Sentences								

N = 53

The lowest intercorrelations for the male group are between the Clerical and the Mechanical tests and between the Clerical and Space Relations tests. The size of these coefficients (.067 and .143,

respectively) indicate that the tests are tapping abilities which have little relationship. Other test intercorrelations are in the moderate range, indicating that the abilities measured, while related, are still relatively independent of one another. Generally speaking, these obtained test intercorrelations are somewhat higher than those reported by the test authors which were previously cited in Chapter III of the present study.

The intercorrelations of greatest magnitude for the female group are those between the Verbal and Sentences tests and between the Space and Abstract Reasoning tests. However, these coefficients are only slightly higher than those of the test authors. Similar to the findings of the test authors are the rather small correlation coefficients obtained between the Clerical test and all others. Very low correlation was also found between the Spelling test and the Space, Mechanical, Clerical, and Abstract Reasoning tests. These latter coefficients are somewhat lower than those of the test authors. As a group, however, the coefficients of intercorrelation obtained in the present study are fairly similar to those of the test authors cited earlier.

In order to isolate test groupings or factors present in the DAT battery the intercorrelation matrices shown in Table XXIII and XXIV were submitted to factor analysis.[1] Results are shown separately for the sexes. In both cases the centroid method of factoring with orthogonal rotation was employed. In the case of the data for the male study group, orthogonal rotation was followed by oblique rotation thus yielding the factor intercorrelations.

---

[1]The author is indebted to Dr. Pairlee J. Stinson, Research Consultant at St. Louis University, for assistance in the factor analysis of the data.

The data in Table XXV indicate the presence of four factors in the DAT for males. Factor I and Factor III appear to be language factors while Factors II and IV are apparently non-language factors. The first language factor, Factor I, is composed of the Verbal Reasoning and Spelling tests, while the second language factor, Factor III, is formed by the Verbal Reasoning and Sentences tests. The Space Relations and Mechanical Reasoning tests form Factor II. Factor IV is composed of the Numerical Ability, Abstract Reasoning, and Clerical Speed and Accuracy tests.

TABLE XXV

FACTOR LOADINGS OF THE DAT FOR MALES
AFTER ORTHOGONAL AND OBLIQUE ROTATION

Tests	Factors			
	I	II	III	IV
Verbal Reasoning	.54*	.24	.49*	.10
Numerical Ability	.16	.21	-.01	.67*
Abstract Reasoning	.02	.27	.07	.62*
Space Relations	-.10	.57*	.12	.14
Mechanical Reasoning	-.24	.64*	.09	.27
Clerical Speed and Accuracy	.30	-.16	.20	.46*
Spelling	.61*	.14	.15	.21
Sentences	.11	.35	.43*	.19

*Denotes tests with highest factor loading.

Since oblique rotation was employed with the data for the males, factor intercorrelations were obtained. These are presented in Table XXVI. It appears that the factors are relatively pure since the correlations are either negligible or negative.

TABLE XXVI

INTERCORRELATION OF DAT FACTORS FOR MALES

Factors	I	II	III	IV
I		-.52	.07	-.10
II			-.31	-.02
III				-.08

The results of factor analysis of the test intercorrelation matrix for the female study group are given in Table XXVII. The centroid method of factor analysis with orthogonal rotation was utilized in obtaining these factor loadings.

TABLE XXVII

FACTOR LOADINGS OF THE DAT FOR FEMALES
AFTER ORTHOGONAL ROTATION

Tests	Factors I	II	III	IV
Verbal Reasoning	.69*	.37	-.09	-.29
Numerical Ability	.50	.31	.24*	-.04
Abstract Reasoning	.24	.77*	.19*	.07
Space Relations	.12	.78*	-.04	.24*
Mechanical Reasoning	.07	.60*	-.21	-.14
Clerical Speed and Accuracy	-.07	.35	.00	-.15
Spelling	.63*	-.10	.01	.30*
Sentences	.83*	.30	-.25	.07

*Denotes tests with highest factor loading.

The data of Table XXVII indicate the presence of four factors in the DAT for females. Factor I appears to be a language factor, Factors II and III appear to be non-language factors, and Factor IV seems to be a composite of two rather different appearing types of ability. The language factor, Factor I, is composed of the Verbal Reasoning, Spelling, and Sentences tests. The first non-language factor, Factor II, is made up of the Abstract Reasoning, Space Relations, and Mechanical Reasoning tests, while the second non-language factor, Factor III, is formed by the Numerical Ability and Abstract Reasoning tests. Factor IV is composed of the Space Relations and Spelling tests.

These findings indicate the presence of four rather distinct factors in the DAT for each sex group. Further attention will be given to these factors in the following section when the multiple correlations between the component tests of each factor and the over-all grade average criterion are considered. Use of the test intercorrelations which have been presented in the present section was made in computation of the highest multiple correlation coefficient between the optimum combination of tests as selected by the Wherry-Doolittle method and over-all grade average. These results will also be considered in a later section of the present chapter.

Correlation of Tests and Factors with Over-all Grade Average

Zero order coefficients of correlation between scores made by members of the male and female study groups on each test of the DAT battery and the subsequent grade averages made by members of the same groups during their freshman year in college were computed. It is recalled that the tests were administered while these subjects were

freshmen in high school. The relationships found between test scores and the over-all grade average criterion are shown in Table XXVIII.

TABLE XXVIII

LONGITUDINAL CORRELATIONS OF SCORES ON THE DAT OBTAINED IN GRADE NINE WITH OVER-ALL GRADE AVERAGE AS COLLEGE FRESHMEN

Test	Males r	Females r
Verbal Reasoning	.498	.451
Numerical Ability	.473	.468
Abstract Reasoning	.473	.094
Space Relations	.338	.123
Mechanical Reasoning	.325	.053
Clerical Speed and Accuracy	.524	.160
Spelling	.481	.359
Sentences	.336	.348

In the case of the males, all tests appear to have longitudinal predictive value. Relationships are fairly high between scores on the Clerical, Verbal, Numerical, Abstract, and Sentences tests and the over-all grade average criterion. Relationships are lower, but still considerable, between scores on the other tests, Space, Mechanical, and Sentences, and the criterion, with the lowest coefficient being .325.

In the case of the females, correlations are fairly high between the Verbal and Numerical tests and over-all grade average. Correlations are moderate between the Spelling and Sentences tests and the criterion.

Relationships between the remainder of the tests and over-all grade average appear insignificant.

It is interesting to compare the longitudinal coefficients of correlation presented here with the findings of a study by Stinson of the relationships of DAT scores with over-all grade average obtained at the end of the semester during which the tests were administered. These findings were reported in Table IX. The coefficients of correlation for males derived in the present study compare very favorably with those of the short range study, while certain correlations for the females in the present study compare favorably and others do not. In the short range study, sex groups were not separated. Therefore, it cannot be known just how well the relationships found in the present longitudinal study would compare with a short range study of a similar nature in which sex groups were treated separately.

The validity coefficients obtained for the Verbal and Numerical tests for both sexes in the present study compare favorably with the longitudinal validity coefficients of general intelligence scales which were reported earlier in Table XX. Also comparing favorably with the data for intelligence tests were the Clerical, Abstract, and Spelling tests, in the case of the males. The coefficients obtained between these tests which have been mentioned actually are very similar to the median correlation coefficients between intelligence test scores obtained on college entrance and college grade average which were reported earlier in Table XIX. These median coefficients approximate .45 and all of the above mentioned tests had validity coefficients of .45 or slightly greater.

Multiple correlation coefficients were computed between the test components of each factor and the over-all grade average criterion for both sexes. In cases where more than two tests compose a factor, the two with the higher factor loading were chosen. Thus, these coefficients, which may be looked upon in the same manner as a simple correlation between factor and criterion, are actually multiple correlations between two tests found to have high loading of a common factor and the over-all grade average criterion. The coefficients thus obtained are given in Table **XXIX**.

For the male subjects, all factors of the DAT correlate highly with college achievement with the lowest coefficient being for one of the non-language factors (.384). Factor I, a language factor, and Factor IV, a non-language factor, have highly similar validity coefficients (.533 and .536). The second language factor, Factor III, is only slightly lower with a validity coefficient of .498.

For the female subjects, Factors I and III appear to have strong correlation with college success. Factor I, a language factor, has a validity coefficient of .453, while Factor III, a non-language factor, has a validity coefficient of .479. Another non-language factor, Factor II, appears to have low validity with a coefficient of .124. The composite factor, Factor IV, has moderate validity with a coefficient of .375.

It will be noted that the increment in multiple R over the zero order "r" of the component test having the highest relationship with the criterion in each factor is very small. This is an indication that the component tests of the factors are measuring abilities which are the same or highly related, and this tendency is to be expected.

TABLE XXIX

LONGITUDINAL CORRELATIONS OF DAT FACTORS WITH OVER-ALL GRADE AVERAGE

Factors	Component Tests Zero Order r	Multiple R
Males	(N = 55)	
Factor I		.533
Verbal	.498	
Spelling	.481	
Factor II		.384
Space	.338	
Mechanical	.325	
Factor III		.498
Verbal	.498	
Sentences	.336	
Factor IV		.536
Numerical	.473	
Abstract	.473	
Females	(N = 53)	
Factor I		.453
Verbal	.451	
Sentences	.348	
Factor II		.124
Abstract	.094	
Space	.123	
Factor III		.479
Numerical	.468	
Abstract	.094	
Factor IV		.375
Space	.123	
Spelling	.359	

There appears to be little advantage, if any, in favor of either the language or the non-language factors in the prediction of over-all college grade average, although the non-language factors involving the Numerical and Abstract tests are better predictors than those involving the Space and Mechanical tests.

Prediction of the Over-all Grade Average Criterion

When the Wherry-Doolittle Test Selection Method for computing highest multiple R was applied to the intercorrelations presented in Tables XXIII and XXIV, the data in Table XXX were obtained.[2] The Wherry-Doolittle method selects those tests which yield a maximum R with the criterion and discards the rest. The process actually consists of determining the multiple R after the addition of each test and stopping when no appreciable increment in R is observed.

For the male subjects, Table XXX shows that the Clerical Speed and Accuracy test (test 6) correlates .524 with the criterion. When the Verbal Reasoning test (test 1) is added, the multiple is raised to .603; when the Abstract Reasoning test (test 3) is added, the multiple is raised to .628. Adding the Numerical Ability test (test 2) raises the multiple very little and at this point little increase in predictive efficiency is obtained by adding additional tests. This combination of four tests accounts for approximately 43% of the variation in the criterion.

For the female subjects, Table XXX shows that the Numerical Ability test (test 2) correlates .468 with the criterion. When the Verbal Reasoning test (test 1) is added, the multiple is raised to .529; when the Abstract Reasoning test (test 3) is added, the multiple is raised to .554. Adding the Spelling test (test 7) raises the multiple

---

[2] H. E. Garrett, Statistics in Psychology and Education (New York, 1953), pp. 404-421.

very little and the increase in predictive efficiency is negligible. This combination of four tests accounts for approximately 36% of the variation in the criterion.

TABLE XXX

MULTIPLE CORRELATION COEFFICIENTS BETWEEN THE SELECTED TESTS AND THE OVER-ALL GRADE AVERAGE CRITERION

Tests	Multiple R
Males (N = 55)	
Clerical Speed and Accuracy (6)	.524
Verbal Reasoning (1)	.603
Abstract Reasoning (3)	.628
Numerical Ability (2)	.631
Females (N = 53)	
Numerical Ability (2)	.468
Verbal Reasoning (1)	.529
Abstract Reasoning (3)	.554
Spelling (7)	.563

One of the major reasons for computing the multiple R is to develop a regression equation from which the criterion can be predicted with the highest precision of which the tests are capable. The regression equations in raw score form for the male study group and the female study group are as follows:

$$X_c \text{ (males)} = .021\,X_6 + .025\,X_1 + .017\,X_3 + .016\,X_2 - .309$$

$$X_c \text{ (females)} = .030\,X_2 + .021\,X_1 - .013\,X_3 + .004\,X_7 + 1.958$$

The values .021, .025, etc., are the weights by which the scores of test 6, test 1, etc., are multiplied. These products and the constant at the end of the equations are summed algebraicly giving X_c, the predicted over-all grade average.

The accuracy with which it is possible to predict criterion scores using the regression equations presented above may be assessed by means of the standard error of estimate. These values may be obtained through use of the appropriate formula.[3] The standard error of estimate associated with the regression equation derived for males is .63 grade points, while the standard error of estimate associated with the regression equation derived for females is .44 grade points. The probability is that approximately sixty-eight times in one hundred the actual grade average will fall within the interval of the predicted grade average plus or minus the standard error of estimate.

Confidence intervals for estimates made with the above regression equations may be computed through use of the tabulated "t" value for the appropriate degrees of freedom. The 95% confidence interval for estimates made with the regression equation for males is the predicted grade average plus or minus 1.27 grade points, while that for the females is the predicted average plus or minus .88 grade points. The greater margin of error associated with predictions for the male group is due to the greater variability of this group on the criterion measure rather than to a difference in the multiple R for the two sexes. The standard deviation in over-all grade point average was .81 grade points for the male group while that for the female group was .53. The means were 2.31 and 2.81, respectively.

---

[3] Ibid., p. 417.

In general, the accuracy of estimates made by use of the above regression equations appears to be satisfactory for the purpose of long range advisement of students by the high school counselor in the school from which the subjects of the present study came. These equations may or may not be useful in other schools. However, as a part of the present study, these equations were applied to the data for a subsequent sample from the same school population as a check upon their validity and usefulness. This sample was the validation group described in Chapter II, consisting of thirty males and twenty-six females.

The test data for members of the male validation group were substituted into the regression equation for males and over-all grade averages predicted. The predicted averages were correlated with the actual grade averages made by these subjects and an "r" of .56 was obtained. This value is slightly lower than the multiple R of .63 obtained with the male study group. However, through conversion to Fisher's "z" function this difference was tested and found not to be statistically significant.

The test data for members of the female validation group were substituted into the regression equation for females and over-all grade averages predicted. These predicted averages were correlated with the actual grade averages made by these subjects and an "r" of .53 was obtained. This value is slightly lower than the multiple R of .63 obtained with the female study group. However, through conversion to Fisher's "z" function this difference was tested and found not to be statistically significant.

Therefore, it may be concluded that the regression equations which have been presented here are valid and useful for prediction of college

freshman over-all grade average from known Differential Aptitude Test scores for individuals within the population from which the present study has sampled.

The indices of longitudinal relationship (multiple R's of .63 and .56) between combinations of Differential Aptitude Tests and college success compare favorably with the median multiple R of .68 for the studies reviewed in Chapter IV. When it is considered that these studies were not longitudinal and that many utilized combinations of test scores and high school grade average, which is known to be a good predictor, the multiple coefficients derived in the present study have a definite respectability.

Correlation of the Tests with Grade Averages in Course Areas

It was of importance to the present study to ascertain the relative merits of the different tests of the DAT battery in predicting success in different course areas of the college curriculum. The areas chosen for investigation were those of language, social science, science, mathematics, and home economics. The size of the groups available for study in these areas vary, since the number of students in the study group who chose to take courses in the different fields was, of course, not the same. The coefficients of correlation between scores on the tests of the battery and the grade averages made in the subject areas are shown in Tables XXXI and XXXII.

For males, all of the tests appeared to have predictive value for course grades in the language area. In order of predictive value, they were Abstract Reasoning, Spelling, Verbal Reasoning, Numerical Ability, Sentences, Clerical Speed and Accuracy, Space Relations, and Mechanical Reasoning. For the females, only four of the tests appeared of predictive

TABLE XXXI

CORRELATION OF THE DAT WITH GRADE AVERAGE IN COURSE AREAS FOR MALES

Course Area	Test VR	NA	AR	SR	MR	CSA	Sp.	Sen.	N
Language	.52	.48	.61	.37	.32	.46	.54	.46	52
Soc. Science	.46	.53	.53	.59	.33	.45	.42	.37	46
Science	.55	.43	.43	.35	.37	.51	.57	.34	39
Mathematics	.26	.42	.31	.16	.33	.61	.39	.16	41

TABLE XXXII

CORRELATION OF THE DAT WITH GRADE AVERAGE IN COURSE AREAS FOR FEMALES

Course Area	Test VR	NA	AR	SR	MR	CSA	Sp.	Sen.	N
Language	.37	.49	.14	.18	-.01	.15	.30	.48	50
Soc. Science	.31	.21	.02	.02	.15	.19	.20	.18	32
Science	.81	.51	.24	.15	-.10	.17	.46	.58	25
Home Econ.	.58	.41	.35	.25	.09	.55	.26	.36	28

value for grades in courses in the language area. These were the Numerical Ability, Sentences, Verbal Reasoning, and Spelling tests, in order of strength. From these results, it does not appear that tests of verbal ability have any advantage over certain tests of abilities classified as non-verbal in the prediction of college grades in the language area, although all tests of a verbal nature were found to have

strong correlations with achievement in this area while all tests of a non-verbal nature did not show this tendency. Therefore, verbal skills appear to be important to success in the language area while the importance of non-verbal abilities varies.

For the males, all of the tests appeared to have predictive value for course grades in the social science area. In order of predictive value, they were Space Relations, Abstract Reasoning, Numerical Ability, Verbal Reasoning, Clerical Speed and Accuracy, Spelling, Sentences, and Mechanical Reasoning. The highest six validity coefficients of the above tests were within the range from .42 to .57, indicating a certain degree of homogeneity of importance among the tests in the battery for prediction in the social science area. Thus, on the basis of these results, it seems that for males a generally high test profile rather than a patterning of abilities is important for success in social sciences.

In the case of females, the only test which appeared to be of even moderate predictive value in the social science area was the Verbal Reasoning test with a correlation .31. The Sentences, Spelling, and Numerical Ability tests appeared to have some relationship to grades in this area, but not enough to be of predictive value. Thus, for females, success in social sciences appears to be related to strength in verbal skills, but prediction on this basis would be subject to much inaccuracy.

Again, all of the tests appeared to have predictive value for course grades in science for the males. In order of importance, the tests were Spelling, Verbal Reasoning, Clerical Speed and Accuracy, Numerical Ability, Abstract Reasoning, Mechanical Reasoning, Space Relations, and Sentences. The highest five validity coefficients of the above tests were within the range from .43 to .57, indicating a similarity of importance among

these tests for prediction in the science area. Both verbal and non-verbal tests are among this group of better predictors. Therefore, as in the case of the social sciences, a generally high test profile rather than a given pattern of abilities seems to be important for male success in the science area.

The Verbal Reasoning test appeared an unusually good predictor of science grades for females, having a validity coefficient of .81. Following this test, in order of importance, were the Sentences, Numerical Ability, and Spelling tests. Other tests were not important predictively. Thus, verbal skills, in combination with numerical ability, seem of predominant importance in prediction of female success in the science area.

For prediction of male success in mathematics, important tests were found to be Clerical Speed and Accuracy, Numerical Ability, Spelling, Mechanical Reasoning, Abstract Reasoning, and Verbal Reasoning, in that order. The test of clerical perception and the numerical test appeared to be of first order importance, while the other tests were of secondary importance. Therefore, it seems that for male success in the mathematics area perceptual speed and accuracy and numerical ability are of primary importance, but verbal and reasoning skills are also of some significance.

For prediction of female success in home economics, important tests were found to be Verbal Reasoning, Clerical Speed and Accuracy, Numerical Ability, Sentences, and Abstract Reasoning, in that order. The Space Relations test assumed a more important role in its relation to home economics grades than in any other area, but with a correlation as low as .25 its predictive importance in this area is not great when used alone. It could be of importance to a pattern, however. From the

evidence, then, the pattern of abilities needed for success in the home economics area appears to involve perceptual speed and accuracy, verbal skills, numerical ability, and spatial visualization. Verbal and numerical abilities have appeared regularly in patterns important to other fields; the uniqueness of this pattern is in the significance of the perceptual and visualization tests.

In summary, it appears that for prediction of male success in the areas of science and social science a generally high test profile is important. Generally high scores also appear of importance in the language area, with some relative favor toward the verbal abilities tests in the battery. Perceptual speed and accuracy and numerical ability are of primary importance to male success in mathematics, with verbal and reasoning skills also of considerable significance.

For the prediction of female success in the areas of language and science, verbal skills seem to be of predominant importance with numerical ability of some significance. No tests emerged as good predictors of female success in the social science area, but there was some tendency for verbal tests to be of relatively more importance than others. In home economics, the pattern included verbal and numerical abilities, but the perceptual and spatial visualization tests were important also.

While the structuring of the patterns of abilities important to success in the different college curricula which have been presented here is not as discrete as might be desired, it is believed that there is enough differentiation in patterning for profiles drawn from scores on the Differential Aptitude Tests to be of considerable value to the high school counselor in assisting students to select appropriate long range educational goals.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this chapter is to review the objectives of the study, briefly state the major findings, draw conclusions, and make certain recommendations concerning the use of these findings and for further research.

Review of Objectives

The problem of the present study was to identify the relationship between scores made on the Differential Aptitude Tests by a sample of Stillwater High School freshmen and the subsequent grade averages made by the same students as freshmen at Oklahoma A. and M. College. In the course of the study the following specific information was obtained for each sex group:

A. The correlation coefficient between scores on each Differential Aptitude Test and college freshman over-all grade average.

B. The coefficients of correlation between scores on each of the Differential Aptitude Tests and college freshman grade averages in the fields of language, social science, science, mathematics, and home economics.

C. The coefficient of multiple correlation between the optimum combination of tests of the DAT battery and college freshman over-all grade average.

D. The multiple regression equation for prediction of college freshman over-all grade average from known scores on the DAT.

E. The coefficient of correlation between over-all grade averages predicted by the obtained multiple regression equation for members of a validation group from the same population and the actual over-all grade averages made by this validation group as college freshmen.

F. Test groupings or factors by factor analysis of the test inter-correlation matrix, and multiple coefficients of correlation between the component tests of each factor and the over-all grade average criterion.

Summary of Results

A. For male group, the correlation coefficients between scores on each Differential Aptitude Test and college freshman over-all grade average were as follows: Verbal Reasoning, .498; Numerical Ability, .473; Abstract Reasoning, .473; Space Relations, .338; Mechanical Reasoning, .325; Clerical Speed and Accuracy, .524; Spelling, .481; and, Sentences, .336. Similar correlations for the female group were: Verbal Reasoning, .451; Numerical Ability, .468; Abstract Reasoning, .094; Space Relations, .123; Mechanical Reasoning, .053; Clerical Speed and Accuracy, .160; Spelling, .359; and, Sentences, .348.

B. For the male group, the correlation coefficients between each of the tests and grade averages in the fields of language, social science, science, and mathematics, respectively, were: Verbal Reasoning, .52, .46, .55, .26; Numerical Ability, .48, .53, .43, .42; Abstract Reasoning, .61, .53, .43, .31; Space Relations, .37, .59, .35, .16; Mechanical Reasoning, .32, .33, .37, .33; Clerical Speed and Accuracy, .46, .45, .51, .61; Spelling, .54, .42, .57, .39; and, Sentences, .46, .37, .34, .16. For the female group, correlations between each test and grade averages in the fields of language, social science, science, and home economics, respectively, were: Verbal Reasoning, .37, .31, .81, .58; Numerical

Ability, .49, .21, .51, .41; Abstract Reasoning, .14, .02, .24, .35; Space Relations, .18, .02, .15, .25; Mechanical Reasoning, -.01, .15, -.10, .09; Clerical Speed and Accuracy, .15, .19, .17, .55; Spelling, .30, .20, .46, .26; and, Sentences, .48, .18, .58, .36.

C. For male group, the optimum combination of Differential Aptitude Tests, as selected by the Wherry-Doolittle method, included the Clerical Speed and Accuracy, Verbal Reasoning, Abstract Reasoning, and Numerical Ability tests; this combination yielded a multiple R of .631 with the over-all grade average criterion. For female group, the optimum combination of tests, selected similarly, included the Numerical Ability, Verbal Reasoning, Abstract Reasoning, and Spelling tests; this combination yielded a multiple R of .563 with the over-all grade average criterion.

D. Regression equations for the purpose of predicting the over-all grade average criterion from known test scores were developed for the two sex groups. The standard error of estimate of the equation for boys was .63 grade points while that for girls was .44 grade points.

E. The coefficient of correlation between actual and predicted grade averages for the male validation group was .56, and that for the female validation group was .53. These values were not significantly different from the multiple R's derived from the data on the study groups, thus indicating that the regression equations are valid for the population sampled.

F. Four factors were found in the DAT for males. Two factors appeared to be language factors and two appeared to be non-language factors. Four factors were also found in the DAT for females. There appeared to be a language factor, two non-language factors, and a composite factor. For the males, the multiple R's between the component

tests of the two language factors and the over-all grade average criterion were .533 and .498, while those for the two non-language factors were .536 and .384. For the females, the multiple R between the language factor and the over-all grade average criterion was .453, those for the two non-language factors were .479 and .124, and that of the composite factor was .375.

Conclusions

The following conclusions and generalizations appear to be warranted on the basis of the evidence from the present study. These statements are recognizedly valid only within the population from which this study has sampled, and caution must be exercised in applying or interpreting them broadly.

A. Scores on the Differential Aptitude Tests obtained at the high school freshman level offer the high school counselor helpful information for use in assessing the advisability of higher education for students.

B. Certain of the Differential Aptitude Tests, particularly those involving verbal and numerical skills, are better predictors of general college success than others.

C. Combinations of Differential Aptitude Tests are better predictors of general college success than are tests taken singly, and these combinations will predict college success with a fair degree of accuracy.

D. The Differential Aptitude Tests tend to group themselves into four rather distinct factors for both sexes, but the groupings are different for each sex.

E. Language factors are good predictors of general college success, but are no better predictors than non-language factors involving numerical and abstract reasoning abilities.

F. While structuring of ability patterns important to success in different college curricula is not as discrete as might be desired, there appears to be enough differentiation for the counselor to find DAT profiles helpful in advising students concerning the choice of a major field.

Recommendations

The present research is somewhat unique in that it approaches the problem of prediction of general college success and success in different college curricula longitudinally, utilizing a relatively recent battery of guidance tests purporting to measure differential abilities. Findings have been rather favorable in general.

It is recommended that generalization on the basis of the findings of the present study be tentative until further research, done in other situations, is forthcoming. When further research of a similar nature has been done, generalizations based on the totality of such research may be more confidently made.

The Dean of Boys at Stillwater High School, who is in charge of guidance activities, has requested to be informed of the findings of this research. It is recommended that the prediction equations developed in this study be put to use in this high school in the long range educational counseling of students. It would be well that occasional validity checks of these equations be made with the passage of time.

More research of the same nature as that of the present study is definitely needed and recommended. The findings of the present study

indicate that test batteries measuring differential abilities have definite long range guidance possibilities, yet further research is needed to establish more definitely the patterning of these abilities important in different fields of college endeavor. More research is also needed in order to establish more definitely the pattern of these measured abilities important to college success in general.

The tendency shown by the present research for the eight Differential Aptitude Tests to group themselves into four factors opens a new area for research with this battery. If this tendency toward grouping is found in data other than that of the present study, then new research directed toward establishing factor patterns important for success in different college areas might prove very fruitful.

BIBLIOGRAPHY

Abelson, Robert P. "Sex Differences in Predictability of College Grades." Educational and Psychological Measurement, XII (October, 1952), 638-644.

Adams, F. J. "Predicting High School and College Records from Elementary School Test Data." Journal of Educational Psychology, XXXIX (February, 1938), 56-65.

------. "College Degrees and Elementary School Intelligence." Journal of Educational Psychology, XXXI (May, 1940), 360-368.

Aiken, W. M. The Story of the Eight-Year Study. New York: Harper and Brothers, 1942.

American Council on Education. Psychological Examination. Princeton, New Jersey: Educational Testing Service.

Anderson, J. E., and L. T. Spenser. "The Predictive Value of the Yale Classification Tests." School and Society (September, 1926), 305-312.

Arlitt, Ada H., and Margaret Hall. "Intelligence Tests Versus Entrance Examinations as a Means of Predicting Success in College." Journal of Applied Psychology, VII (March, 1923), 330-338.

Attender, Louise E. "The Value of Interest, Personality, and Vocational Interest Tests in a Guidance Program." Journal of Educational Psychology, IX (September, 1940), 449-452.

Beatley, B. "The Relative Standing of Students in Secondary School in Comprehensive Entrance Examinations and in College." School Review, XXX (February, 1922), 142-147.

Bendig, A. W. "The Reliability of Letter Grades." Educational and Psychological Measurement, XIII (June, 1953), 311-321.

Bennett, G. K. Manual, Mechanical Comprehension Test. New York: Psychological Corporation, 1947.

Bennett, G. K., and J. E. Doppelt. "The Evaluation of Pairs of Tests for Guidance Use." Educational and Psychological Measurement, VIII (June, 1948), 319-325.

Bennett, G. K., H. G. Seashore, and A. G. Wesman. A Manual for the Differential Aptitude Tests. New York: Psychological Corporation, 1952.

Bennett, G. K., and A. G. Wesman. "The D. A. T.: Some Comments by the Authors." Occupations, XXVII (October, 1948), 20-22.

Bent, R. K. "Scholastic Records of Non-High School Graduates Entering the University of Arkansas." Journal of Educational Research, XL (October, 1946), 108-115.

Berdie, Ralph F. "The D. A. T. as Predictors in Engineering Training." Journal of Educational Psychology, XLII (March, 1951), 114-123.

Billhartz, W. H. and P. W. Hutson. "Determining College Ability during Junior High School Years." School and Society, LIII (November, 1941), 547-552.

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

Black, Donald B. "Prediction of Academic Success in the University of Washington," Unpublished Doctoral Dissertation. Seattle, Washington: University of Washington, 1951.

Bowles, Frank H. "Crystal Balls and College Admissions." Measurement and Evaluation in the Improvement of Education, 15th Educational Conference. Washington, D. C.: American Council on Education, 1951.

Brown, Hugh S. "Differential Prediction by the A. C. E." Journal of Educational Research, XLIV (March, 1950), 116-121.

Buros, Oscar K. Fourth Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon Press, 1953.

Butsch, R. L. C. "Improving the Prediction of Academic Success Through Differential Weighting." Journal of Educational Psychology, XXX (September, 1939), 401-420.

Byrns, Ruth K. "Scholastic Aptitude and Achievement." School and Society, XXXV (May, 1932), 713-718.

Byrns, Ruth K., and V. A. C. Henmon. "Entrance Requirements and College Success." School and Society, XLI (February, 1935), 101-104.

Byrns, Ruth K., and V. A. C. Henmon. "Long Range Prediction of College Achievement." School and Society, XLI (June, 1935), 877-880.

Carlin, Leslie C. "A Longitudinal Comparison of Freshman-Senior Standing." Journal of Educational Research, XLVII (May, 1953), 285-290.

Carter, H. L., and Dorothy McGennis. "Some Factors which Differentiate College Freshmen Having Lowest and Highest Point-Hour Ratios." Journal of Educational Research, XLVI (November, 1952), 219-226.

Coleman, William. "An Economical Test Battery for Predicting Freshman Engineering Course Grades." Journal of Applied Psychology, XXXVII (September, 1953), 465-467.

Cottle, William C. "The D. A. T.: Some Comments." Occupations, XXVI (March, 1948), 344-345.

Crawford, A. B. and P. S. Burnham. "Entrance Requirements and College Achievement." School and Society, XXXVI (June, 1932), 344-352; 378-384.

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Brothers, 1949.

Darley, John G. Testing and Counseling in the High School Program. Chicago: Science Research Associates, 1943.

DeRidder, Lawrence M. "Relation between Gross Scores on the A. C. E. and Academic Success." Journal of Educational Research, XLVI (June, 1953), 353-358.

Donahue, Wilma T. The Measurement of Student Adjustment and Achievement. Ann Arbor, Michigan: University of Michigan Press, 1949.

Doppelt, J. E., and G. K. Bennett. "A Longitudinal Study of the D. A. T." Educational and Psychological Measurement, XI (April, 1951), 228-237.

Doppelt, J. E., and A. G. Wesman. "The Differential Aptitude Tests as Predictors of Achievement Test Scores." Journal of Educational Psychology, XLIII (April, 1952), 210-217.

Douglass, Harl R. "Relation of Pattern of High School Credits to Scholastic Success in College." North Central Association Quarterly, VI (May, 1931), 283-297.

------. Relation of High School Preparation and Certain Other Factors to Academic Success at the University of Oregon. University of Oregon Publications, Vol. 3, No. 1. Eugene, Oregon: Oregon State System of Higher Education, 1931.

------. "Different Levels and Patterns of Ability Necessary for Success in College." Occupations, XXII (December, 1943), 182-186.

Dubois, Philip H. "Achievement Ratios of College Students." Journal of Educational Psychology, XXX (December, 1939), 699-702.

Durflinger, G. W. "Prediction of College Success." Journal of the American Association of College Registrars, XIX (June, 1943), 68-78.

Edds, H. J., and W. M. McCall. "Predicting Scholastic Success of College Freshmen." Journal of Educational Research, XXVII (October, 1933), 127-131.

Finch, F. H., and C. L. Nemzek. "Prediction of College Achievement from Data Collected During the Secondary School Period." Journal of Applied Psychology, XVIII (September, 1934), 454-460.

Fredericksen, Norman O. "Making Test Scores More Useful For Prediction." Educational and Psychological Measurement, XI (Winter, 1951), 783-787.

Fredericksen, Norman O., and W. B. Schrader. "The A. C. E. Psychological Examination and High School Standing as Predictors of College Success." Journal of Applied Psychology, XXXVI (May, 1952), 261-265.

Fricken, C. E. "Predicting Achievement in the Liberal Arts College." School and Society, XLII (October, 1935), 518-520.

Garrett, Harley F. "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Sciences and Teachers Colleges." Journal of Experimental Education, XVIII (February, 1949), 91-138.

Garrett, Henry E. Statistics in Psychology and Education. New York: Longman, Green, and Company, 1953.

Garrett, W. S. "Ohio State Psychological Examination as an Instrument for Predicting Success in College." Occupations, XXII (May, 1944), 489-495.

Gates, A. I. A List of Spelling Difficulties in 3876 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1937.

Goldthorpe, J. H. "Relative Rank in High School and the First Two Years in the University." School and Society, XXX (July, 1929), 130-134.

Goodman, C. H. "Prediction of College Success by Means of the Thurstone Primary Mental Abilities Test." Educational and Psychological Measurement, IV (February, 1944), 125-140.

Gray, W. S. Provision for the Individual in College Education. Chicago: University of Chicago Press, 1932.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1950.

Hake, Dorothy T., and C. H. Ruedisili. "Predicting Subject Grades of Freshmen with the Kuder Preference Record." Journal of Applied Psychology, XXXIII (October, 1949), 553-558.

Hartsen, L. D. "The Validation of the Rating of Twenty-Three Tests for Predicting Freshman Scholarship." School and Society, XXXVI (September, 1932), 413-416.

------. "Further Validation of the Rating Scales Used with Candidates for College Admission." School and Society, XLVI (April, 1937), 155-60.

Heston, J. C. "The Use of Non-verbal Tests in the Prediction of Academic Success." Journal of Educational Psychology, XXXIII (May, 1942), 308-314.

Hoerres, Mary A. and J. D. O'Dea. "Predictive Value of the A. C. E." Journal of Higher Education, XXV (March, 1954), 97-99.

Jones, Arthur J. Principles of Guidance. New York: McGraw-Hill Company, 1945.

Keys, N. "The Value of Group Test I.Q.'s for Prediction of Progress Beyond High School." Journal of Psychology, XXI (February, 1940), 81-93.

Kinney, L. B. "The Use of Aptitude Tests for Subjects Taught at the College Level." Unpublished Paper. Iowa City, Iowa: University of Iowa, 1944.

Knapp, Robert H. Practical Guidance Methods. New York: McGraw-Hill Company, 1953.

Lindquist, E. F. Educational Measurement. Washington, D. C.: American Council on Education, 1951.

Martin, L. O. The Prediction of Success for Students in Teacher Education. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Messenger, H. R. "The Probability-Table, A Possible Means for Early Elimination of Poor Students from Teachers' College." Northern Illinois State Teachers College Quarterly, VII (May, 1930), 87-93.

Morris, Louis H. "The Relationship between Certain Factors and Academic Success in College Mathematics." Unpublished Doctoral Thesis. Stillwater, Oklahoma: Oklahoma A. and M. College, 1954.

Nelson, M. J. "Some Data from Freshman Tests." School and Society, XXXI (June, 1930), 772-774.

Osborne, R. T. "The Differential Prediction of College Marks by A. C. E. Scores." Journal of Educational Research, XLIV (March, 1950), 107-115.

Patterson, Dale O. "The Relationship between Certain Factors and Scholastic Success." Bulletin of the American Association of Collegiate Registrars, XII (April, 1937), 191-201.

Pierson, C. D. "A Study of High School Seniors to Determine Who Shall Be Admitted to College." School and Society, XXVIII (September, 1928), 215-216.

Prescott, A. C. "Teacher Estimates of Success in College." School Review, XLVIII (April, 1940), 278-284.

Read, C. B. "Prediction of Scholastic Success in a Municipal University." School and Society, XLVIII (April, 1938), 187-188.

Rosenfeld, M. A., and C. L. Nemzek. "Long Range Prediction of College Marks." School and Society, XLVIII (March, 1938), 127-128.

Ross, C. C. Measurement in Today's Schools. New York: Prentice-Hall, Inc., 1947.

Scates, D. E. "A Study of High School and First Year University Grades." School Review, XXXII (March, 1924), 182-192.

Seashore, H. G. "Tenth Grade Tests as Predictors of Twelfth Grade Scholarship and College Entrance Status." Journal of Counseling Psychology, I (March, 1954), 106-115.

Segel, David. "Differential Prediction of Scholastic Success." School and Society, XXXIX (January, 1934), 91-96.

------. Prediction of College Success. U. S. Office of Education Bulletin No. 15. Washington, D. C.: U. S. Government Printing Office, 1934.

Seyler, E. C. "The Value of Rank in High School Graduating Class for Predicting Freshman Scholarship." Journal of American Association of College Registrars, XV (October, 1939), 5-22.

Shaw, Duane C. "A Study of the Relationships between Thurstone P. M. A. Scores and High School Achievement." Journal of Educational Psychology, XL (May, 1949), 239-249.

Smith, Francis F. "The Use of Previous Record in Estimating College Success." Journal of Educational Psychology, XXXVI (March, 1945), 167-176.

Starrak, J. A. "Matching Ability with Achievement." Journal of Higher Education, VIII (June, 1937), 315-320.

Stinson, Pairlee J. "A Statistical Analysis of the Differential Aptitude Tests for the Purpose of Predicting First Semester Grade Averages of a Freshman High School Group." Unpublished Masters Thesis. Stillwater, Oklahoma: Oklahoma A. and M. College, 1952.

Stone, Joics B. "Differential Prediction of Academic Success at Brigham Young University." Journal of Applied Psychology, XXXVIII (March, 1954), 109-110.

Super, Donald E. Appraising Vocational Fitness. New York: Harper and Brothers, 1949.

Symonds, Percival M. Measurement in Secondary Education. New York: MacMillan Company, 1929.

Taylor, Calvin W. "General Aptitude Test Battery Patterns for College Areas." Occupations, XXIX (April, 1951), 518-526.

Thomson, W. A. "Note on Retest Results on the A. C. E. Psychological Examination for College Freshmen." Journal of Educational Psychology, XXXI (June, 1940), 229-233.

Thurstone, L. L., and T. G. Thurstone. Chicago Tests of Primary Mental Abilities, Manual of Instructions. Chicago: Science Research Associates, 1943.

Townsend, Agatha. "The Differential Aptitude Tests, Some Data on the Reliability and Intercorrelations of the Parts." Educational Records Bulletin, LIII (January, 1950), 39-47.

Travers, Robert M. and W. L. Wallace. "Inconsistency in the Predictive Value of a Battery of Tests." Journal of Applied Psychology, XXXIV (May, 1950), 237-239.

Traxler, Arthur E. Techniques of Guidance. New York: Harper and Brothers, 1945.

------. "The Status of Measurement and Appraisal Programs in City High Schools." Educational Records Bulletin, LXI (February, 1953), 75-76.

Votaw, David F. "A Comparison of Test Scores of Entering College Freshmen for Prediction of Scholarship." Journal of Educational Research, XL (November, 1946), 215-218.

Wallace, W. L. "Differential Predictive Value of the A. C. E. Psychological Examination." School and Society, LXX (July, 1949), 23-25.

------. "The Prediction of Grades in Specific College Courses." Journal of Educational Research, XLIV (September, 1951), 587-597.

Weaver, L. J. "New CEEB Examinations." Journal of Higher Education, XVI (January, 1945), 37-38.

Webb, Sam C., and J. N. McCall. "Predictors of Freshmen Grades in a Southern University." Educational and Psychological Measurement, XIII (October, 1953), 660-663.

Weber, C. O. "Comparative Values of College Entrance Examinations." School and Society, LV (February, 1942), 247-251.

------. "Old and New College Board Scores and Grades of College Freshmen." Journal of American Association of College Registrars, XX (October, 1944), 70-75.

Wesman, A. G. "Separation of Sex Groups in Test Reporting." Journal of Educational Psychology, XL (April, 1949), 223-229.

------. "Problems in Differential Prediction." Educational and Psychological Measurement, XI (May, 1951), 265-272.

------. "The D. A. T., A Five Year Report." Personnel and Guidance Journal, XXXI (March, 1952), 167-170.

Whitney, F. L. "College Success as Related to Intelligence." Educational Administration and Supervision, XVI (November, 1930), 668-672.

Williamson, E. G. The University of Minnesota Studies in Predicting Scholastic Achievement. Part K. Minneapolis: University of Minnesota Press, 1942.

Wolking, W. D. "Predicting Academic Achievement with Differential Aptitude and the Primary Mental Abilities Tests." Journal of Applied Psychology, XXXIX (March, 1955), 115-118.

Wood, Ben D. "Results of the Space Relations, Mechanical Reasoning, and Clerical Tests of the Differential Aptitude Tests in Six Public Schools." Educational Records Bulletin, LVIII (February, 1952), 79-84.

VITA

Edwin Earle Vineyard
candidate for the degree of
Doctor of Education

Thesis: A LONGITUDINAL STUDY OF THE RELATIONSHIP OF DIFFERENTIAL APTITUDE TEST SCORES WITH COLLEGE SUCCESS

Major: Psychology

Biographical and Other Items:

Born: August 30, 1926, near Wright City, Oklahoma.
Undergraduate Study: Eastern Oklahoma Agricultural and Mechanical College, Wilburton, Oklahoma, 1946-1947; Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1947-1949; Completed the Bachelor of Science Degree in 1949.
Graduate Study: Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1949-1951; Completed Master of Science Degree in 1951; Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1952-1955.
Experience: Taught and served as principal in the public schools of Oklahoma from 1949 through 1954; Served as a teaching fellow in the Psychology Department, Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1954-1955.

Member of Phi Delta Kappa, Kappa Delta Pi, Phi Kappa Phi, and Psi Chi.

Date of Final Examination: July, 1955.

THESIS TITLE: A LONGITUDINAL STUDY OF THE RELATIONSHIP OF DIFFERENTIAL APTITUDE TEST SCORES WITH COLLEGE SUCCESS

AUTHOR: Edwin Earle Vineyard

THESIS ADVISER: Dr. Harry K. Brobst

The content and form have been checked and approved by the author and thesis adviser. The Graduate School Office assumes no responsibility for errors either in form or content. The copies are sent to the bindery just as they are approved by the author and faculty adviser.

TYPIST: Mrs. Gordon Culver