

Statement of Permission to Copy

In presenting this professional paper in partial fulfillment of the requirements for an advanced degree at Montana State University, I agree that the Library shall make it freely available for inspection. I further agree that permission for extensive copying of this professional paper for scholarly purposes may be granted by my major professor, or, in his absence, by the Director of Libraries. It is understood that any copying or publication of this paper for financial gain shall not be allowed without my written permission.

Signature

J. B. Bawley, Jr.

Date

March 1, 1971

EFFECT OF HIGH SCHOOL GRADUATING CLASS SIZE IN PREDICTING
FUTURE ACADEMIC PERFORMANCE

by

JAMES BARRY SLAUGHTER

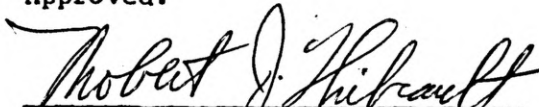
A professional paper submitted to the Graduate Faculty
in partial fulfillment of the requirements for the degree
of

MASTER OF EDUCATION

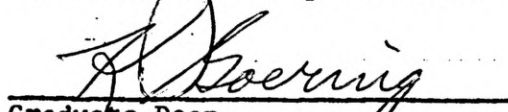
with emphasis in

Counseling

Approved:


Head, Major Department


Chairman, Examining Committee


Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

June, 1971

Acknowledgment

The completion of this study was made possible only with the assistance of several individuals. The investigator especially wishes to extend his gratitude to Dr. Eric Strohmeyer and Dr. Albert Suvak for their aid in setting up the study and their patience in explaining the statistical procedures used, and Mrs. Zeier of the computer department for her cooperation and assistance with the computer programs. The investigator also thanks the members of the freshman class of 1969, who unknowingly and anonymously provided all the data contained in the following pages.

J.B.S.

Table of Contents

Introduction	1
The Problem--Definitions, Limits and Procedures	1
The Significance	3
Summary	5
Review of Literature	5
General Knowledge	5
Out of State Studies	6
Instate Studies	9
Summary	11
Procedures	12
The Population	12
Data Collection	14
Statistical Hypotheses	15
The Prediction Equation	16
Procedures	17
Summary	20
Findings	21
Other Data	27
Summary	30
Summary, Discussion, and Conclusions	32
Discussion	33
Conclusions and Recommendations	36
References	39

List of Tables

TABLE 1. Multiple R Obtained Using Full Model	21
TABLE 2. Values of F for Full Model and Restricted Model. .	23
TABLE 3. Significance of the Reduction in R^2 Between The Full and Restricted Models	25
TABLE 4. Relationship Between College GPA And HS Class Size	26
TABLE 5. Correlation Between HS Class Size and Each Of Three Other Variables	28
TABLE 6. Effect of Grouping on Correlation Between Paired Variables	30

Abstract

The effect of high school graduating class size as a predicting variable upon future academic performance was determined by including this variable in a multiple regression equation with two other independent variables, the University Computed High School Grade Point Average, and the Ohio State University Psychological Test score, and the dependent variable, actual three quarter cumulative college grade point average. The population studied was 1509 Montana State University freshmen who had graduated in the spring of 1969 from public Montana high schools. This population was considered as a whole, and was also broken down by sex and by certain curricular groupings. An F-ratio formula was used to test the significance of the reduction occurring in the variances of the equations when the experimental variable was dropped from the equation. A t-test was applied to the simple correlations obtained between variables.

The tests of significance indicated a small reduction in variance for all groups, significant for five of the eight groups tested; and a small positive correlation between high school class size and actual college grade point average for seven groups, significant for two.

The investigator concluded that high school class size could be used to increase the efficiency of college grade prediction, for large populations, but that the relatively small increase in the multiple correlation coefficient obtained did not justify the collection of class size data on a continuing basis. The investigator also concluded that there was evidence to suggest that, on the whole, a student's academic success in college was in small part dependent upon the size of high school from which he had graduated.

Introduction

In the fall of 1969, 1509 Montana public high school graduates registered as freshmen at Montana State University (Suvak, 1969). An individual in this group may have come from a high school graduating class of several hundred students, or from one of half a dozen. His school may have had a hundred instructors teaching nearly as many courses, or it may have had less than ten instructors and offered fewer than thirty courses.

With such wide differences in quantity existing among the state's high schools, it would seem appropriate to ask if all Montana children are receiving an education of roughly equal quality, or if perhaps the graduate of one of the state's smaller schools does not suffer when compared with his counterpart from a large school. One basis for comparison would be his performance in college. Is he placed at an immediate competitive disadvantage upon enrollment? Will his grade point average at the end of his freshman year be significantly lower than that of a large school graduate in the same curriculum? A positive answer to these questions seems intuitively possible, if not even probable.

The Problem--Definitions, Limits and Procedures

This study sought to determine if the size of the student's Montana high school, when included with the variables of high school grade point average and his score on a standardized aptitude or achievement test affected positively the accuracy of prediction of his college grade point average at the end of his first year at Montana State University.

Briefly, the problem was solved by ordering the instate Fall 1969 MSU freshmen by the exact size of their respective public high school graduating classes and performing a multiple regression analysis including the high school class size as an independent variable with high school grade point average and a standardized test score. These last two variables are currently used in predicting college grades at Montana State; therefore, data regarding these variables was available in the computer banks.

The multiple regression analysis, using as the dependent variable the individual's actual freshman college grade point average, was performed not only for the population as a whole, but also for the following populations existing within this larger group:

- (1) All males within the population.
- (2) All females within the population.
- (3) All members of the population enrolled in the College of Engineering.
- (4) All members of the population enrolled in the College of Agriculture.
- (5) All members of the population enrolled in the College of Education.
- (6) All members of the population enrolled in the School of Nursing.
- (7) All members of the population enrolled in the Commerce Department.

Furthermore, since correlation was computed with the end of year (Spring 1970) cumulative grade point average, only those for whom this information was available were included in the study.

In the statement of the problem, the term "standardized aptitude or achievement test" is deliberately vague because research at Montana State (Suvak, 1966) has shown that many such tests can be used effectively to increase the correlation coefficient of grade prediction. That which was actually used in this study was the Ohio State University Psychological Test. A further discussion of this matter will be presented in Chapter Three.

The Significance

Earlier it was stated that it does seem intuitively possible that the graduate of a small high school will not perform as well in college as a graduate of equal ability from a larger high school, and at least one study (Hoyt, 1959) seems to bear this out. The present study sought, within its established limits, to verify or disprove that assumption; it does not pretend to investigate the reasons behind its findings. The statistical significance and the practical significance of these findings will be explored in later chapters, but the significance of the study itself lies largely in its attempt to provide by statistical analysis answers to questions regarding the efficiency of the small high school as an educational medium. As the review of literature in the following chapter will indicate, such studies have seldom been

performed, and those which exist have often been inconclusive or conflicting in their findings.

Summary

In a large, thinly populated state such as Montana, college students come from widely varied high school backgrounds, with one of the most obvious differences being in the size of high school. Because our unsupported reasoning, and indeed, some research, might lead us to the assumption that the quality of education may vary with the size of the school, this study was conducted to determine if the size of a student's high school graduating class could be used to increase significantly predictions of his college success. No causal relationships were explored in this study, but perhaps further research into this area will be encouraged by the study.

The following chapter will review in greater detail some of the literature already mentioned, as well as other work done in this and related areas.

Review of Literature

For several decades, but especially since the beginning of the post-World War Two boom in college enrollment, there has been a great deal of interest in college grade prediction, with the result that nearly every possible measurable variable in the student's existence has been studied by someone, somewhere. Because the general area has been covered in such breadth, this chapter will focus only upon studies that have included, at least casually, information on the specific area under investigation; i.e., the relationship between school size and subsequent student performance. Several out of state studies will be reviewed first, then some studies done at Montana State University and concerning Montana high school students will be examined. Before considering these, however, a brief discussion will be made concerning generally accepted knowledge in the area of grade prediction.

General Knowledge

Considering the sheer number of published reports dealing with grade prediction, it is perhaps inevitable that some areas of agreement would be found, despite greatly differing research procedures and criteria used in the various investigations. While these findings may not be common knowledge among educators in general, and therefore should perhaps receive at least casual mention in this review, it would be improper to attribute them to any specific investigator.

It is now generally (although not unanimously) accepted that the best predictor of college success is the high school grade point average (HSGPA). Many investigators use the HSGPA without modification,

but some, such as Suvak (1966) have obtained a slightly higher correlation by weighting various subjects differently through a system of grade averaging. Not too many years ago it was argued that high school rank (HSR) was the most accurate predictor of academic success (Aiken, 1964), and some, such as Hoyt (1959) devised elaborate weighting formulae to lessen the obvious effects on HSR of high school graduating class size.

A great many studies have also indicated that the addition of one or two standardized aptitude or achievement test scores to the formula will result in a small but significant increase in the correlation over that found when the HSGPA alone is used. Addition of still further variables, however, has usually proved to be of little value.

Finally, predictors of college level academic success from high school determined variables have generally come to limit themselves to performance at some point during the freshman year, usually the end of the first semester or the end of the academic year.

Out of State Studies

As has been implied elsewhere in this paper and shown by Suvak (1966) and many others, scores on virtually any well known and accepted standardized aptitude or achievement test will correlate positively with college grades. It would not seem irrelevant to this investigation, therefore, to review in some depth a study which compared by achievement test scores junior high students from different size schools. Street, Powell, and Hamblin (1962) reported on such an investigation of some

2,700 seventh and eighth graders from two districts in eastern Kentucky. The investigators grouped students by grade and by total school enrollment, the latter as follows: 300 or more, 100-299, and less than 100. They found differences in scores for every subject area tested, significant to the .05 level, in favor of the larger schools. From this they were led to conclude:

"The evidence of this study points toward a strong likelihood that students in larger schools ... tend to outperform students in smaller schools, in the same (or perhaps comparable) districts. The evidence does not reject the possibility that factors other than size influence the differences in levels of achievement of students. It does imply that (whatever the reasons) ... larger schools appear to produce higher achievement levels (p. 266)."

The authors also pointed out that it might not be wise to generalize their findings beyond eastern Kentucky (a mostly rural mining and farming area) and also that the largest school in their study had an enrollment of only 836, which might be considered small in some areas. Although not explicitly stated by the authors, it is also well to note that the study made no allowances for ability differences; hopefully, the relatively large sample size assured that the groups contained a fairly normal distribution.

While the above study produced quite striking evidence of differences in achievement between students of different size schools (at least in one instance), no effort at prediction of future performance was attempted by the authors, nor has a followup study been performed, to the knowledge of this investigator. An earlier study by Hoyt (1959),

however, did use school size as a variable in predicting performance. Despite some rather strange, and perhaps minor, inconsistencies, this study has since become one of the most widely referred to in this subject area.

Using the 1956 freshman class at Kansas State College, Hoyt grouped the students by size of high school graduating class, as follows: 25 and fewer, 26-49, 50-99, 100-250, and 251 and more (no explanation is given for the unequal grouping intervals). When he had made adjustment for individual HSR (which was his method for accounting for ability differences), Hoyt found that the groups from the larger high schools had a significantly higher (.05 level) college GPA than the group from the smallest schools. Moreover, each group GPA was higher than that of any group below it in class size; however, while these differences were consistent in all cases, they were not significant at the .05 level between adjacent groups. This study did differentiate by sex, a variable which Lavin (1965) says is too often overlooked in grade prediction, but not by college curriculum. Other limitations were that the college was on a semester system, so only two sets of grades at most made up the individual grade point averages; those who dropped out of school after one semester were included in the study with no mention of what proportion of the total of each group they comprised; and finally, the population included only those members of the freshman class for whom HSR was available (63.8% of men; 77.1% of women).

Although Hoyt's findings were not particularly spectacular, they

were verified a few years later to about the same degree in a study at the University of Minnesota (Watley, 1964). In an attempt to increase the efficiency of college grade prediction, Watley considered type, location and size of high school for 1101 freshman males enrolled over a two year period in the school's Institute of Technology. Results indicated that inclusion of these variables in a multiple regression formula would increase prediction efficiency significantly (.05 level) for those graduates from the largest (urban) high schools. Small, but not significant, differences were observed for all other groups.

In summary, a few studies linking school size and subsequent academic performance have been conducted, and there seems to be some indication that students from larger high schools achieve better than those from smaller schools, although results have often been insignificant and sometimes even conflicting, as Aiken (1964) points out. One reason for the existing inconsistencies between reports was indicated by Hoyt (1959), who commented that of the 20 investigations he had studied, some of which were as much as four decades old and several unpublished, practically no two were alike concerning the variables that were controlled or the manner in which the study was conducted. (This investigator has noted that Hoyt, himself, hardly provided a model to follow in this regard).

Instate Studies

While no work directly relating high school size and grade prediction has been performed to date at Montana State University (at least

to the knowledge of the investigator), studies relating college dropouts and school size, and others concerning the curriculum offerings of the various size schools have been made and bear brief mention. Two separate studies (Hamilton, 1962; Boyd, 1969) noted that the first year attrition rate from MSU in the early 1960s was approximately 20 percent, and Hamilton further noted, as did Aubert in 1963, that a disproportionate number of these came from the smaller schools. The Aubert study is especially interesting in that it indicates that not only do small school (0-125 total enrollment, as defined in his study) graduates have a greater statistical chance of becoming dropouts, but also a smaller percentage attend college in the first place.

The smaller schools have received considerable attention in MSU studies---mostly of the type they would likely prefer to do without. Currie, in 1961, concluded that 40 course units were needed in order for a school to provide a good "general education ... and some depth in all areas (p. 19)," and found 77 of 83 class "C" schools deficient in this respect, while only two of 13 class "A" and "AA" schools failed to meet this criterion (both offered 39). Only three class "C" schools, in fact, offered as many courses as any class "A" school studied.

Daniels (1968, p. 24) provides a good indication of the reason for the above when, in a study of schools with a total enrollment of 100 or fewer students, he found that "the average number of full time secondary teachers ... was 2.8. The average number of part time

(secondary) teachers was 4.0." One would hardly expect a school with a staff of this size to be able to provide an adequate curriculum.

Summary

While some studies of the effect of high school graduating class size on college academic performance have been conducted, findings have often been inconsistent and/or of little significant value. There does seem to be some evidence that students from larger schools may achieve better than those from smaller schools. While no Montana State University studies have been performed in this area, two investigations concerning school size have indicated that the smaller schools have a smaller percentage of their graduates entering college and a larger percentage of college dropouts, compared with the state's larger schools. No causal relationships were explored (other than by conjecture) in either of these studies.

The following chapter will provide a detailed examination of the methods and procedures used in this investigation.

Procedures

As the review of literature in the previous chapter indicates, studies concerning school size and academic performance have been inconclusive and even conflicting in their findings. This investigation considered high school size as one of three variables used to predict freshman college grades, and should, at least insofar as Montana students attending this university are concerned, provide a fairly complete examination of this issue.

This chapter will first include a description of the population and the characteristics by which it was examined, followed by a statement of the statistical hypotheses and a description of the analysis of data which was used to test these hypotheses.

The Population

The population, as briefly described in the first chapter, was made up of the Fall 1969 freshman enrollees at Montana State University who graduated from public Montana high schools in the spring of 1969, and seven sub-groups taken from this population and considered separately. The decision to use sub-group populations as well as the whole was based upon the assumption that such a group as a college freshman class must include a wide range of individual interests and abilities; the members of each sub-group were assumed to possess certain characteristics in common, not necessarily common to the group as a whole. The eight populations for which data was gathered, the size of each, and certain other comments are as follows:

- (1) The population as a whole, defined as above. $N = 1509$.
- (2) All males in the population. $N = 860$.
- (3) All females in the population. $N = 647$. (It will be noted that the total of these two N 's = 1507. Two key punch cards not coded for sex had to be discarded in this grouping, as, to protect the confidentiality of data included in the study, no information which would identify individuals was placed on the key punch cards provided to the investigator.)
- (4) All members of the population enrolled in the College of Engineering. $N = 269$. Data was analyzed for this population because the relatively technical and specialized course content of the various curricula in this college might lead one to assume, from the review of literature (Daniels, 1968 and Currie, 1961) that the small school graduate might be placed at a particular disadvantage in competing with those from larger schools with more varied course offerings.
- (5) All members of the population enrolled in the College of Agriculture. $N = 156$. This population was included as a counterpoint to the above.
- (6) All members of the population enrolled in the College of Education. $N = 162$.
- (7) All members of the population enrolled in the School of Nursing. $N = 89$.
- (8) All members of the population enrolled in the Commerce Department. $N = 127$.

Analysis of data for five different curricular groupings allowed

the investigator to observe if the prediction equation was more or less effective for different curricula; these particular groups were selected because, in addition to other reasons given above, each provided a substantial number of subjects to allow test of significance curves approaching the normal, and each is usually thought of as differing substantially in course requirements from any of the others.

Data Collection

The following data were used in this study:

(1) Size of high school spring 1969 graduating class. The investigator obtained this information for every public high school in the state from the State Office of Public Instruction.

(2) University Computed High School Grade Point Average (UCHSGPA). This is the predictive (independent) variable currently used at Montana State University which has been shown to produce the highest overall one variable correlation with college freshman grade point average. It is basically the average of **nine** different subject grade point averages, compiled from eight semesters of high school work. Suvak (1966), who developed this system, provides a more detailed explanation.

(3) Individual scores on the Ohio State University Psychological Test (also known as the Minnesota Scholastic Aptitude Test). This test is taken by all MSU freshmen prior to admission to classes and is used by the school as a predicting variable with the UCHSGPA. Its inclusion also acts as an equalizer of ability differences among individuals.

(4) Individual freshman college grade point average (CGPA). This was a three quarter cumulative average of the student's college grades, computed on a per credit basis, 4.0 system. Those students who did not complete three quarters of study were not included, as data bank information was available only for those for whom a three quarter cumulative grade point average was available.

Statistical Hypotheses

Following is a list of the statistical hypotheses tested:

(1) That the multiple correlation coefficient (R) obtained from the prediction equation using the three independent variables of UCHSGPA, Ohio Test, and high school graduating class size (GCS), yielding a predicted CGPA, will be positive in its relationship to the dependent variable, actual CGPA. The 1966 report by Suvak would lead one to expect that values of R for the various groups considered would range between .50 and .70.

(2) That the hypothesis stated above will hold true for all eight of the populations under investigation.

(3) That each R^2 (the variance of R for each of the respective groups) will be significant.

(4) That the reduction in R^2 , when the GCS variable is dropped from the equation, will be significant.

(5) That the linear correlation between CGPA and GCS will be positive and will be significant.

The Prediction Equation

In order to analyze the data in such a manner as to test the above hypotheses, a prediction equation, also called a regression equation or model, must be set up, with the variables of the investigation producing a predicted result. This predicted result can then be mathematically compared to the actual result to produce a multiple regression coefficient (R), and further computations will reveal the contribution made to R by each variable. All of the actual computations were done by computer, and tests of significance applied to the data thus obtained.

The prediction equation with three independent variables and one dependent variable may be written as follows:

$$Y' = k + b_1 x_1 + b_2 x_2 + b_3 x_3$$

Where:

Y' = freshman CGPA, the dependent variable

k = a pure constant, mathmatically determined

b_1, b_2, b_3 = constants used as weighting factors, mathmatically determined

x_1 = UCHSGPA, an independent variable

x_2 = Ohio Test score, an independent variable

x_3 = GCS, an independent variable

To test the reduction in R^2 , the above equation will be reduced by the element x_3 and its constant, resulting in the equation

$$Y' = k + b_1 x_1 + b_2 x_2$$

known as the restricted model (while the equation with all three independent variables is known as the full model). The F-test of significance can be applied to the difference between the resulting R^2 of the full model and the R^2 of the restricted model.

A further explanation of the terms and equations used in this and the following paragraphs may be found in any basic statistics text; the investigator used Ferguson (1966).

Procedures

The actual procedures of this investigation were greatly simplified in that all computations except the testing for significance could be computer performed and most of the information needed by the computer was already available on tape. The only new information which had to be placed on key punch cards was the high school GCS coding.

Accordingly, the investigator was able to request that information pertinent to the study be transferred from tape to key punch cards, add the high school graduating class size to these same cards, and then use a multiple regression program already available, to obtain all necessary data.

Information placed on each key punch card was as follows:

(1) High school code number. This was a three digit number assigned by the state and was unique to each school. For example, Bozeman Senior High was coded 102; Wolf Point High was 267.

(2) Sex of the individual (coded either M or F).

(3) University computed high school grade point average, based upon eight semesters of high school work. This was a standard three digit GPA based upon the 4.0 system (for example, 2.78).

(4) Actual score received by the individual on the Ohio Test.

(5) Predicted freshman college GPA, based upon a multiple regression analysis of the above two variables.

(6) Actual freshman college GPA, used as the dependent variable.

(7) College curriculum in which the individual was enrolled. This was a two digit code, such as "22" for Elementary Education.

(8) Various other test score information not used by the investigator in this study.

To this information the investigator added, on each card, a four digit code to identify the student's high school graduating class size. The first three digits represented the actual graduating class size, and the last was a grouping code which gathered all class sizes into one of seven size groups. This group code allowed the investigator to observe the amount of information, or degree of accuracy, lost by grouping, by comparing the simple correlation obtained between actual CGPA and actual GCS with the correlation between CGPA and grouped GCS.

Once the above data was placed on the key punch cards, a multiple regression program was added to the card deck and computer runs were performed to yield the following information:

(1) A simple correlation pairing all variables.

(2) A multiple correlation (R) and variance (R^2) between the three independent variables and the dependent variable (the full model).

(3) A multiple R and R^2 between UCHSGPA and Ohio Test only, and the dependent variable, freshman CGPA (the restricted model).

This information allowed the investigator to test the significance of the relationship between high school GCS and freshman CGPA with both the F-test and t-test, as described below.

The F-test indicates whether there is a significant reduction in R^2 between the full model and the restricted model (see p. 16). The formula for the F-test may be written as follows:

$$F = \frac{(RSQ_f - RSQ_r) / df_1}{(1 - RSQ_f) / df_2}$$

Where:

$RSQ_f = R^2$ obtained using all three independent variables

$RSQ_r = R^2$ obtained using only UCHSGPA and Ohio Test as independent variables

$df_1 =$ degrees of freedom $k-3$; in this study, 1

$df_2 =$ degrees of freedom $N - K$, where $N =$ total members in group and $K =$ total number of variables

The t-test was used to test the significance of the simple correlation between CGPA, the dependent variable, and GCS, the independent variable. The equation for this test may be written as follows:

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$

Where:

r = correlation between CGPA and GCS

N = total in population

The F-test was also applied to the value of R^2 in the full model, in order to test the significance of the multiple R. No computations were necessary on the part of the investigator, however, as a value for F was supplied by the computer printout for the multiple R.

Summary

The actual conduct of the investigation was greatly facilitated by the use of computer performed computations. All variables and other information used except the variable of high school graduating class size was already available on a computer tape, and all computations except some tests of significance could be computer performed with available programs.

Procedural operations consisted of performing a multiple regression analysis using the independent variables of high school GPA, a standardized test score, and size of high school graduating class with the dependent variable of actual freshman CGPA, on a population consisting of all instate public school Fall 1969 freshman enrollees at Montana State University, and seven sub-group populations taken from the whole. The primary hypothesis investigated was that the third variable, high school class size, affects college academic performance. Results will be reported by written analysis and illustrated with various tables in the following chapter.

Findings

The first and second statistical hypotheses listed on page 15 may be taken together and summarized as follows: "The multiple correlation coefficients obtained with the full model of three independent variables and one dependent variable will be positive for all eight of the populations in the study." The investigator expected, from prior research (Suvak, 1966) using the variables which formed the restricted model in the present study, values of R ranging between .50 and .70 for the various populations. This expectation was met. The values actually obtained for R are listed in Table 1 below.

TABLE 1
Multiple R Obtained Using Full Model

Group	N	Multiple R
Freshman Class	1509	.6566
Males	860	.6419
Females	647	.6730
Coll. of Engineering	269	.5912
Coll. of Education	162	.7170
Coll. of Agriculture	156	.5406
Commerce Dept.	127	.6871
School of Nursing	89	.6854

No test of significance was applied to the differences in R between groups, but it is interesting to note the rather wide range in the values of R obtained for the various curricula. The two male-dominated colleges of Engineering and Agriculture have the two lowest values, while the curricula drawing the majority of their students from the female group have the higher values; however, the investigator would hesitate to infer that the fair sex is the more predictable, based upon this data alone.

The third statistical hypothesis stated that the variance (R^2) for each of the groups would be significant; the .05 level of the F-test was chosen as the test and level of significance to be met. The values for F given in Table 2 were supplied by the computer, so no computation of F was necessary on the part of the investigator. Degrees of freedom for the lesser mean square equals $N - k$, where k equals the number of variables; and degrees of freedom for the greater mean square equals $k - 1$. This test was performed for both the full model and the restricted model for each of the eight groups; $k - 1 = 3$ for the full model, and 2 for the restricted model, where the variable of high school graduating class size (GCS) has been dropped from the equation (p. 16).

Table 2 is presented on the following page.

TABLE 2

Values of F for Full Model and Restricted Model

Group	Full Model			Restricted Model		
	N-k	k-1	F	N-k	k-1	F
Freshman Class	1505	3	380.16	1505	2	558.19
Males	856	3	199.96	852	2	292.03
Females	643	3	177.50	631	2	252.29
Coll. of Engr.	265	3	47.45	269	2	70.16
Coll. of Educ.	158	3	55.74	158	2	82.91
Coll. of Ag.	152	3	20.92	152	2	29.96
Commerce Dept.	123	3	36.66	123	2	52.67
Sch. of Nurs.	85	3	25.10	85	2	37.38

Critical values for F at .05 level, where $N-k=80$ & $k-1=3$: 2.72
Critical value for F at .05 level, where $N-k=80$ & $k-1=2$: 3.11

Since the critical value of F decreases as sample size increases, the table was simplified by noting only the critical value of F for a population slightly smaller than the smallest of those studied; if F is significant for an $N - k$ of 80, it will also be significant for an $N - k$ of 85, or of 1505. The critical value of F was exceeded in all instances, and the hypothesis is therefore accepted.

The data in Table 2 indicates a minor fault in the study which could conceivably have had some effect on the analysis of data

regarding the fourth statistical hypothesis. Key punch cards mutilated during some of the computer runs resulted in different N's being reported for the full and restricted models for three of the groups studied. As the differences were small--less than two percent even in the instance of greatest deviation--the investigator has assumed them to be inconsequential to the measure of the reduction in R^2 . For the sake of consistency, the N reported for the full model was used in the F-test formula for this reduction.

The fourth statistical hypothesis states that the reduction in R^2 , when the variable of high school graduating class size is dropped from the regression equation, will be significant. The test of this hypothesis is the central test of the study, in that it indicates whether or not GCS is in fact a significant variable in the prediction of college academic performance. If the reduction in R^2 is significant when any variable is dropped from the equation, then that variable may be said to be a significant contributor to the multiple R obtained when the variable is included in the equation. The F-test, .05 level, was chosen to test this hypothesis.

Table 3, on the following page, presents the data pertaining to this hypothesis.

TABLE 3
Significance of the Reduction in R^2
Between the Full and Restricted Models

Group	N	RSQ _f	RSQ _r	F	Critical Value of F, .05 level
Freshman Class	1509	.4311	.4257	14.21*	2.61
Males	860	.4120	.4064	8.12*	2.62
Females	647	.4530	.4439	10.71*	2.62
Coll. of Engr.	269	.3495	.3420	3.00*	2.65
Coll. of Educ.	162	.5142	.5105	1.19	2.67
Coll. of Ag.	156	.2922	.2814	2.30	2.67
Commerce Dept.	127	.4721	.4593	2.98*	2.68
Sch. of Nurs.	89	.4698	.4650	.77	2.72

As Table 3 indicates, the reduction in R^2 was significant for five of the eight groups studied. It should be noted, however, that the F-ratio is not independent of sample size, and that four of the five groups for which F was significant were the four largest groups. More will be said later regarding the fifth group, for which some rather interesting data has yet to be presented.

Regarding the fourth hypothesis, the null is rejected, and the statistical hypothesis accepted, for the population as a whole, for the male and female sub-groups of the population, and for two of the five curricular sub-groups taken from the population.

The final statistical hypothesis states that the linear correlation between college GPA and high school graduating class size will be positive. The test of significance chosen by the investigator was the t-test, one tailed, at the .05 level.

Although the data gathered to test this hypothesis cannot indicate the effect of GCS in a multiple regression equation, it is presented to give an indication of the simple relationship present in this study between the two variables. This data is contained in Table 4.

TABLE 4
Relationship Between College GPA and HS Class Size

Group	N	r	t	Critical Value of t, .05 level
Freshman Class	1509	.0579	2.25*	1.66
Males	860	.0612	1.80*	1.66
Females	647	.0581	1.48	1.66
Coll. of Engr..	269	.0560	.92	1.66
Coll. of Educ.	162	.0253	.32	1.66
Coll. of Ag.	156	.0186	.23	1.66
Commerce Dept.	127	-.0181	(-).20	1.66
Sch. of Nurs.	89	.0691	.64	1.67

While the reduction in R^2 showed GCS to be a significant variable in the multiple regression equation for five of the eight groups, the data in Table 4 showed a significant correlation between GCS and CGPA for only the two largest groups. Five of the remainder correlated positively, but did not meet the test of significance, and one curricular group showed a small negative correlation, not statistically significant.

The null hypothesis, then, is rejected for the group as a whole and for the males, but is retained for the female group and for all five curricular groups.

Other Data

In the course of testing the statistical hypotheses, the investigator collected certain data not directly concerned with the relationship between high school graduating class size and college grade point average. Because this data is interesting in itself, and because its inclusion in this report may add meaning to the findings presented earlier, the investigator has chosen to include it here.

The multiple regression program used by the investigator provided, in addition to the information already presented, a two variable correlation pairing each of the four total variables with every other. While the correlations obtained from pairs including only the three variables of UCHSGPA, Ohio Test, and CGPA are not the concern of this study, it may be of interest to note the correlations obtained when the variable of GCS is paired with each of the other three. This data is presented

in Table 5. (For the convenience of the reader, the correlation between GCS and CGPA, presented in Table 4, is repeated.)

TABLE 5

Correlation Between HS Size and Each of Three Other Variables

Group	N	GCS Correlated With:		
		CGPA	UCHSGPA	Ohio Test
Freshman Class	1509	.0579*	-.0720*	.0943
Males	860	.0612*	-.0505	.0782*
Females	647	.0581	-.1040*	.1180*
Coll. of Engr.	269	.0560	-.0868	.0330
Coll. of Educ.	162	.0233	-.0884	.1248
Coll. of Ag.	156	.0186	-.1742*	-.0262
Commerce Dept.	127	-.0181	-.2204*	-.0208
Sch. of Nurs.	89	.0691	-.0852	.1762*

*Significant at .10 level, t-test, two tailed

The most striking finding illustrated in Table 5 is the negative correlation between high school grade point average and GCS, consistent in all groups and large enough to pass the test of significance in four. Especially worthy of note is the relatively large negative correlation between UCHSGPA and GCS for the Commerce curriculum group. It will be

recalled that the reduction in R^2 (Table 3) was significant for this group even though the simple correlation between CGPA and GCS was negative, at a level which did not meet the requirements for significance.

Correlation coefficients between the Ohio Test and GCS were positive in six cases and negative in two. In all four instances where the test of significance was met, however, the correlation was positive.

One final piece of data collected by the investigator indicates the effect of grouping on the correlation coefficient. Most previous studies in this area, as mentioned in the review of literature, have grouped GCS in some manner rather than using the exact size as did the investigator. The investigator wished to observe the amount of information lost by one particular method of grouping, and accordingly divided class sizes into seven groups, ranging from graduating classes of size 0-20 for the first group to 450+ for the seventh. Increments were selected so as to provide a comparable number of subjects in each group. The correlation coefficients obtained are presented in Table 6.

TABLE 6

Effect of Grouping on Correlation Between Paired Variables

N = 1506	
Paired Variables	r
CGPA with GCS (exact)	.0525
CGPA with GCS (grouped)	.0420
GCS (exact) with GCS (grouped)	.9221

Although no tests of significance were performed on the data in Table 6, grouping did produce an observable loss of information. It is perhaps noteworthy that the same population, grouped, correlated only .92 with itself, ungrouped.

Summary

Data collected to test the five statistical hypotheses presented in the preceding chapter yielded the following information:

(1) Hypotheses 1 and 2: That the Multiple R obtained from the full regression model will be positive for all eight of the groups studied. These hypotheses were accepted.

(2) Hypothesis 3: That the variance for each group will be significant, in both the restricted model and the full model. This hypothesis was accepted.

(3) Hypothesis 4: That the reduction in variance (R^2) between the full and restricted models will be significant for each group.

The reduction in R^2 was found to be significant for five of the eight groups.

(4) Hypothesis 5: That the linear correlation between CGPA and GCS will be positive and significant.. The relationship was found to be positive for seven of the eight groups, but was significant for only the two largest.

Other data presented showed a negative correlation consistent for all groups and significant for four between GCS and UCHSGPA. The Ohio Test was found to correlate positively with GCS for six of the eight groups, four significantly, and negatively for two groups, but not significantly at the level chosen.

One method of grouping investigated was shown to cause a substantial loss of information, but no test of significance was applied to this data.

The final chapter will present a summary and the conclusions of the investigator.

Summary, Discussion, and Conclusions

The purpose of this study was to determine if the size of a student's high school graduating class could be a significant variable in the prediction of his college academic performance, as reflected by his cumulative three quarter grade point average. Five statistical hypotheses were set up and tested in order to analyze the effect of this variable as a predictor for those members of the 1969 Montana State University freshman class who had graduated from public high schools in the state the previous spring. The hypotheses were tested for the group both as a whole and divided by sex, and for members of the group enrolled in one of five selected curricula.

The study was conducted with the aid of a computer program used by the university to predict college grades in various subject areas. Key punch cards which contained all information pertinent to the study were supplied to the investigator, who added a code number representing the exact size of the individual's high school graduating class and performed a multiple regression analysis to determine the effect of the added variable on the resulting correlation coefficient. As a means of identifying the individual's high school was available, it was not necessary for the investigator to gain access to any information which would have enabled him to identify individuals themselves in any manner.

The study found that high school graduating class size contributed significantly to the prediction of freshman college grades for five of

the eight groups for which data was collected. These five groups were the population as a whole, the male population, the female population, the population of College of Engineering students, and the population of students enrolled in the Commerce Department. Small contributions were also found to exist for the other three groups--College of Education, College of Agriculture, and School of Nursing--but these contributions did not meet the test of significance, which was the .05 level of the F-test.

The study also found that high school class size and high school grade point average (as reflected by the University Computed High School GPA, an averaging formula) correlated negatively for all groups, although the correlation was significant for only half--the group as a whole, females, College of Agriculture, and Commerce Department. The Ohio State University Psychological Test correlated positively, and to a significant degree, with class size for the population as a whole, both sex groups, and the enrollees in the School of Nursing.

Discussion

Perhaps the most important point of discussion raised by this study is the distinction which must be made between the significance of high school class size as a predictor of academic performance and its practical value as a predicting variable. Although class size has been shown to contribute significantly to a multiple regression equation, for some curricula and for large sample sizes, is this

contribution great enough to justify, on a continuing basis, the effort necessary to collect the data? It is the conclusion of the investigator that it is not, and the following comments are offered to support this conclusion.

The contribution made by the GCS variable, even where the reduction in R^2 was significant, was very slight. Values of R obtained with the restricted model ranged between .53 and .71; with the GCS variable included in the full model, these R 's were raised by values between .0035 and .0101. This is hardly sufficient to allow a practical increase in the accuracy of prediction.

The data which comprised the GCS variable in this study would have to be gathered every year, as of course the exact size of high school classes vary from year to year. Any practical application of a class size variable would probably necessitate the use of some grouping system. The investigator found, however, that one system, which divided class sizes into seven groups, resulted in a substantial loss of information, which would further lessen the value of class size as a predicting variable.

Although one might well question the practicability of including high school class in a prediction equation, it is difficult to avoid the conclusion that on the whole, and taken as a group, students' college grades depend in part upon the size of high school from which they graduated. The findings were consistent, if not always significant

at the level specified. Such a conclusion leads quite naturally to speculations regarding possible reasons why this may be the case. Some possibilities occurring to the investigator are listed below; perhaps the findings of this study may encourage investigation into these or other areas.

(1) The student from the small school may have an unrealistic concept of his own abilities and may become severely discouraged when his college grades fall short of his high school performance. The graduate of a large class may well have a much better idea of where he stands intellectually in comparison with others, and this knowledge may work to his advantage. Some support for this speculation appeared in the study. While GCS correlated positively with college GPA for seven of the eight groups, it correlated negatively with the UCHSGPA for all, significantly so for four. This would tend to suggest that students from smaller schools perform better in high school but poorer in college than do those from larger schools.

(2) The student from the large school may find the environment (class sizes, dormitory living, etc.) less threatening.

(3) The quality of instruction, in general, may be better in larger schools. Of many possible reasons this could be true, two of the less subjective are the increased teacher opportunity for subject specialization and a smaller annual turnover rate. Some support for this speculation can be found in the literature. Daniels (1968, p. 38)

found that for schools of less than 100 total enrollment, "the average turnover rate for the instructional staffs ... was thirty-seven percent." He concluded that this rate was such "that it would tend to hinder the effectiveness ... of the instructional programs ... at these ... schools."

(4) The small school graduate may be placed at a disadvantage due to the comparative lack of variety in the courses offered at his school. Daniels (p,37) found that the majority of the small schools he studied "provided the student with the minimum requirements for graduation and little more." Regarding the present study, if one assumes that this disadvantage, if any, might appear in the College of Engineering group, as this is usually considered to be a highly technical course of study, evidence is mixed. The reduction in R^2 for the GCS variable was significant, indicating that the variable made a significant contribution to the regression equation, but the simple correlation between GCS and CGPA was not significant.

Conclusions and Recommendations

It was concluded that high school graduating class size could contribute significantly to the prediction of freshman college grade point average, when included as an independent variable in a prediction equation with other independent variables such as high school grade point average and a standardized test score. This conclusion applies only to very large populations, and the investigator recommends that

it not be generalized to apply to other than Montana public high school graduates attending Montana State University.

The practical value of high school class size in a prediction equation was concluded to be so slight as not to justify the collection of data necessary for its use on a continuing basis.

It was concluded that for very large sample sizes there is a positive relationship between college freshman GPA and high school graduating class size, and a negative relationship between high school class size and high school GPA. Again, it is recommended that these results not be generalized beyond Montana public schools and this university.

Two final recommendations for future studies in the area of grade prediction are as follows:

(1) That future studies include the differentiation by sex and some sort of differentiation by curricula, as did this study. The investigator's research into the literature indicated that such differentiations were often overlooked, and his findings led him to conclude that they were of value.

(2) That future investigators concern themselves with such variables as personality characteristics (maturity, independence, judgment, etc.) or, perhaps, study habits. Despite the obvious difficulty of trying to objectify such data, the investigator feels that this area holds real promise. Perhaps a standardized personality test could yield some "maturity index" which could readily be applied in a multiple regression

formula and might be of practical value in grade prediction. When a multiple R can account for less than fifty percent of the variance in college grades, there must still be some room for refinement.

References

- Aiken, L. R. Jr. Rank in high school graduating classes of various sizes as a predictor of college grades. The Journal of Educational Research, 1964, 58, 56-60.
- Aubert, W. A summary and interpretation of a study of 1958 Montana high school graduates' attendance in institutions of higher learning. Unpublished manuscript, Montana State University, 1963.
- Boyd, J. J. A followup study of freshman dropouts from Montana State University in 1963. Unpublished manuscript, Montana State University, 1969.
- Currie, E. D. Evaluation of subject and activity offerings in Montana high schools. Unpublished manuscript, Montana State University, 1961.
- Daniels, W. H. An investigation of scheduling practices and curriculum offerings at secondary schools of enrollments of 100 or less. Unpublished manuscript, Montana State University, 1968.
- Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1966.
- Hamilton, J. J. A study of dropouts in the 1960-61 freshman class of Montana State College. Unpublished manuscript, Montana State University, 1962.
- Hoyt, D. P. Size of high school and college grades. Personnel and Guidance Journal, 1959, 37, 568-573.
- Lavin, D. E. The prediction of academic performance. New York: Russell Sage Foundation, 1965.
- Street, P., Powell, J. H., & Hamblin, J. W. Achievement of students and size of school. The Journal of Educational Research, 1962, 55, 261-266.
- Watley, D. J. Type, location and size of high school and prediction of achievement in an institute of technology. Educational and Psychological Measurement, 1964, 24, 331-338.