A STUDY IN THE PREDICTION OF SCHOLARSHIP.

C. E. ERFFMEYER, NORTH CENTRAL COLLEGE, NAPERVILLE, ILLINOIS.

Many lines of investigation in the problem of predicting college scholarship have been followed. The possibilities of estimating at the time of entrance a student's academic success in college, on the basis of physical and sensory traits, of high school grades, of entrance examinations, and of intelligence test scores have all been tried. Predictions based upon physical and sensory traits alone have proved of slight value, and those depending upon high school grades, entrance examinations, or intelligence test scores possess a large element of error. College scholarship undoubtedly is the resultant of many forces and influences, and any single factor will scarcely serve as a basis for an accurate prediction. There is the possibility, however, that a combination of measures, none of which shows a particularly high relationship with scholarship, might enable us to predict scholarship with greatly increased accuracy. The technique for making such a combination is the method of multiple correlation. It may be stated at this point that the chances for a high multiple correlation coefficient are best when each of the measures entering into the combination exhibits a fairly high correlation with the criterion, in this case college scholarship, and a low correlation with one another. Our problem, then, is to determine the factors which influence scholarship, and to make the best combination of these factors for its prediction.

The present study was made to determine the accuracy of predictions of college scholarship made on the basis of the best combination of measures which were available in the case of students of Ohio State University. It was recognized that there are other factors influencing scholarship which were not found among those available for this study. Measures in fourteen factors were secured for each of 416 students selected at random from the class of 1925 of the College of Arts of Ohio State University.

Six of these measures were in physical traits and the zero order correlation coefficient of each of these traits with college scholarship was as follows: weight, minus .12; chest expansion, minus .02; lung capacity, plus .02; height, plus .07; pulse, plus .13; vital index, plus .13. None of these coefficients indicated a close relationship with scholarship, but it was possible that measures in one or two of these traits when combined with measures in other traits might serve to increase the accuracy of the prediction appreciably. The other traits correlated with scholarship and the coefficients of correlation were as follows: age, minus .07; number of units of high school work, plus .21; the total of all high school grades, irrespective of whether the number of units was the same for all students, plus .44; average high school foreign language grade, plus .50; intelligence test score, plus .51; average high school English grade, plus .53; average high school mathematics and science grade, plus .54; average high school grade, which included all subjects, plus .57.

While several of the above correlation coefficients may be regarded as fairly high, it is quite apparent that predictions of scholarship based upon coefficients of such magnitude will in many cases prove to be considerably in error. The next step was, therefore, to determine whether these measures, when combined and weighted in the best possible manner, would permit more accurate predictions of scholarship. The multiple correlation and the multiple ratio correlation techniques, explanations of which may be found elsewhere,[1] were used in answering this question. It was noted just above that the zero order correlation coefficient between college scholarship and the average high school grade was plus .57. Beginning with this measure, which correlated highest with scholarship, which of the remaining measures when combined with this first one will yield the highest multiple correlation coefficient? The measure found to do this was the intelligence test score, the multiple correlation coefficient between scholarship and the best weight-

[1](a) Yule, G. U. An Introduction to the Theory of Statistics, p. 248. London, Charles Griffin and Company, 1922.

(b) Clem, O. M. Detailed Factors in Latin Prognosis, p. 27-28. New York, Teachers College, Columbia University Contributions to Education, No. 144, 1924.

(c) Garfield, Evelyn. The Measurement of Motor Ability, p. 18-20. New York, Archives of Psychology, No. 62, 1923.

ed combination of the average high school grade plus the intelligence test score being plus .648. After these two measures were used in combination as a basis for predicting scholarship, to what extent did the addition of a third measure to the combination serve to raise the correlation coefficient? Of the twelve remaining measures, the one which made the greatest contribution to the two already being used was the vital index. The correlation between scholarship and the best weighted combination of the average high school grade, the intelligence test score, and the vital index was plus .654. Furthermore, the addition of all the remaining measures to the combination served to raise the correlation only to plus .665. That is to say, when a prediction of college scholarship is based upon the average high school grade and the intelligence test score, its accuracy is not appreciably increased even though all of the other twelve variables included in this study are taken into consideration.

It was also desirable to ascertain the relationship between these various measures and college scholarship when all the other factors were held constant. This was done by means of partial correlation. The results indicated that the closest relationship of college scholarship was with the average high school grade and the intelligence test score when the older factors included in the study were held constant. The correlation coefficient between college scholarship and the average high school grade, where the zero order correlation was plus .57, became plus .47 when the influence of all the other factors had been eliminated. In the case of college scholarship and the intelligence test score, where the zero order correlation coefficient had been plus .51, the correlation when the other influences had been ruled out by partial correlation was plus .38. Of the remaining factors, the highest partial correlation with scholarship was found in the case of the vital index, the coefficient being plus .12. It is thus apparent that both the average high school grade and the intelligence test score are unique in their relationship with scholarship. Only these two factors each showed a fairly high correlation with college scholarship when the concomitant influence of all the other variables was eliminated. Each of these two

measures supplied a factor for the prediction of college scholarship which could not be gained from any of the other variables included in this study. Furthermore, as might be inferred from what has just been said, the data revealed that the presence or absence of either of these two measures significantly affected the accuracy of the estimation of scholarship.

The next step was the writing of the regression equation by means of which the college scholarship was predicted on the basis of the best combination of the average high school grade and the intelligence test score. This equation was as follows: $X_1=3.58\ X_2+.46\ X_3-268.2$. In this equation X_1 is the scholarship score which is to be estimated on the basis of X_2, the average high school grade, and X_3, the intelligence test score. It may be stated here that the scholarship score was the average number of honor points earned per year when each quarter hour of A grade work counted four such points; B, three points; C, two points; D, one point; and E, no points.

Several illustrations will serve to make clear the use of the regression equation in making the estimates or predictions of scholarship. Student A entered Ohio State University as a freshman at the same time as did the group included in this study, but he was not one of this original group. A's average high school grade was 75.5; while the mean of the average high school grades for the group included in the study was 85.4, and the standard deviation 5.2. A's intelligence test score was 92; the mean of these scores for the group was 109.6, and the standard deviation 30.1. Using the regression equation, it was found that A's predicted average number of honor points per year was 45.2. Since a student at Ohio State University carries ordinarily about sixteen hours of work each of the three quarters of the year, this means that a prediction of an average grade of approximately D would be made for this student. A's actual achievement in scholarship turned out to be 55.5 points, which was very close to the predicted score.

In the practical situation of predicting scholarship, however, the scores which would subsequently be made

would still be in the unknown future at the time of prediction. It would therefore be important to have some notion at that time as to how accurate the prediction will prove to be. This information is given by the standard error of estimate or the probable error of estimate. The probable error of estimate, when the prediction was based on the average high school grade and the intelligence test score, was found to be 21.7 points. This means that the chances are even that the actual score will be within 21.7 points of the estimated score. When estimates of scholarship were made for thirty additional students by means of the regression equation which was derived in this study, and these estimates were checked with actual achievement scores in scholarship, it was found that in seventeen of the thirty cases the error was less than the probable error of estimate.

In predicting college scholarship, how serious is an error which is equal, for example, to this probable error of estimate? Suppose that an entering freshman student has an estimated scholarship score of 96 points. Assuming that he carries sixteen hours of work each quarter, this means that his predicted average grade is C. The chances are even that he will actually earn not less than 74.3 points, nor more than 117.7 points; or to state it somewhat differently, the chances are even that his grade will not be lower than a combination of grades C and D, in which the C's predominate (for example, 27 hours of C and 21 hours of D), nor higher than a combination of grades C and B in which the C's again predominate (for example, 27 hours of C and 21 hours of B).

Another illustration may be presented which involves the regulation of Ohio State University that students failing to make an average of 1.8 scholarship points for each hour of work carried (an average slightly below C) during the first three years are to be dropped from the university at the end of the third year. Assuming a schedule of sixteen hours of work, this means that a student must make a total of 86.4 scholarship points per year. Suppose that the estimated scholarship score for a given student who will carry sixteen hours of work is 65 points. It will be noted that this is almost exactly one P. E. distance, 21.7

points, below the minimum required. What is the probability that this sudent will meet the minimum requirement, in spite of an estimated score of only 65 points? The chances are even, it has been stated, that his actual score will vary 21.7 points or more from the estimated score. However, there is as much probability that this variation will be in the direction of a lower score as in the direction of a higher score. Consequently, he has one chance in four of meeting this minimum requirement; or, of four students for each of whom a score of 65 points is estimated, one only may be expected to make a score as high as 86.4 points.

Another student who has a percentile intelligence rating of five, or an original score of 60, and a high school record better than that of sixteen percent of entering freshmen at Ohio State University, presents himself for admission. That is, his average high school grade, which is 80.2, is one standard deviation below the mean. Accordingly, his predicted scholarship score is 47.1 points. This is 39.3 points, or 1.81 times the probable error of estimate, below the minimum required of a student with a schedule of sixteen hours. He has but eleven chances out of one hundred of meeting the standard of 1.8 honor points for each hour of work carried.

The data which have been presented do not of course answer the question as to whether such a minimum requirement of an average of 1.8 honor points per hour of work, or some similar regulation, is or is not a desirable requirement. Neither do they answer the question as to whether, if it were possible to determine at the time of application for admission those students who will, to a high degree of probability, fail to meet such a requirement, such students should be denied admission. The data do, however, permit an estimate of college scholarship which may be accompanied by a statement of the probability that the actual accomplishment of the student will vary from this estimate by any given amount.

It seems evident from the data presented that admission to college should not be based alone upon an intelligence test rating. Even where the purpose of the college or university is to admit only those students who will achieve a certain degree of success in scholarship, this

object can be attained more satisfactorily if other criteria in addition to the intelligence test score are used. Certainly, to such a score some measure of high school scholarship should be added. The present study has shown that either of these two measures taken singly does not result in as accurate a prediction as the inclusion of both will give. On the other hand, if to the combination of these two factors any or all of the remaining measures included in this study are added, the accuracy of the prediction is increased scarcely at all.