

# Predicting Academic Achievement at the University of Oregon

CLIFFORD L. CONSTANCE

ASSOCIATION • 1 LLEGiate REGISTRARS

Last spring we completed a rather comprehensive investigation of the students who enter the University of Oregon, with special reference to their academic achievements in the University, and a method of anticipating in advance the quality of those achievements. The main purpose was the determination of regression equations, to use in estimating the grades which students would earn and the length of time they would remain in the University. However, there have been many important by-products. Committees of the faculty have been studying various aspects of curricular organization and this study has been very useful to them. Also we have made an analysis of students received from individual schools, and have reported to each high school in the state concerning the graduates of that school who have enrolled here. The techniques used are so fruitful that we are continuing them as aids to deriving similar data as needed in the future.

In scope the study included all undergraduate matriculates entering during a five-year period, excepting a very few individuals of irregular status. For each matriculate a Hollerith card was prepared, summarizing data dealing with him personally and with his academic record. The card shown (Figures 1 and 2) differs from those actually used in this study but is our present best idea for the purpose. It concentrates the data used in this investigation and, on the reverse side, summarizes many additional important details of a student's academic record for regular office use (see our Chapter V of Part II in *Practical Applications of the Punched Card Method in Colleges and Universities*).'

The basic data of the study divide naturally into four periods in student chronology as indicated below.

1: *Data known before matriculation:* name; sex; school last attended, and preparatory school from which the student graduated if matriculating from another college; and decile based on preparatory school grades, or advanced grade point average and advanced credits if matriculating from another college or university.

2. *Data known shortly after matriculation:* date of matriculation;

1 Baehne, G. W., pp. 59-90.

19 3. Data known after

completion of first term in the University; term hours failed in first

term; and estimated grade point average age at matriculation;

major at matriculation; and decile based on entrance psychological

examination.

first-term grade point average; term hours passed in first term; S.:

F	dti-hip:	U.S.	Date of birth	of birth
Ad., GPA (A)	Ad., cm dit3 (B)			

Oregon - -----

Summary Hollerith Card.

referred to in the second paragraph.' Our evaluations of preparatory school records, grade point averages, actual and estimated, and a few other factors are described in our article, "The University Presents Student Facts," in the *Oregon Education Journal* for February, 1935. Since that time our grade point system has been changed; now, 4 **points are given** for each hour of A; 3, for B; 2, for C; 1, for D; and 0, for F. All statistics in this present article have been changed accordingly, but the Report Sheet in Figure 3 is based on the for-

UNIVERSITY OF OREGON Record of Students from

P "• E.B.

z  
(1929 to 1934)

t g <sub>i-D</sub> E; a r

1. 7..., Margatet R.	7 29	AA	20	7 3 L4 1.	1.3 18 78 2	6 -	-
2.. M....., Russell B.	7.29	BA	26	8 7 2.4 2.3 1.9.123 59 0 13 34			-
3. P....., 080r•e 2.	7 33	BA	-18	7 4 0.8 1.2 0.6	0 41 4	3 -	P
4. P....., John R.	F 31.	Bd	18	6 4 0.4 0.9 0.3	8 78 25	9	- PPOPD
5. 3....., Carl 2.	729	So	18	6 7 -0.2 0.7 -0.1	2 16 11	2 -	PD
6. T....., Mary 3.	F 29	AA	16	0 5 2.1 2.	1.9 120 69 0 12 33		-

FIG. 3. Sample of Report Sheet for Oregon High Schools.

mer system. Possibly none of the other data require explanation here. Our choice of factors for deriving prediction equations was highly eclectic; we attempted to follow out various promising leads, especially those discovered at this University by Dr. Howard R. Taylor, Dr. Earl M. Pallett, and Dr. Harl R.

Douglass. Many of the factors proved of little or no value in statistical prediction, although they still may be worth securing for student descriptive purposes. REPORT TO THE HIGH SCHOOLS

In this section we shall describe particularly the report made

21 the high-school principals when they received a copy of the data on their graduates who had come to the University, together with an

accompanying letter of explanation. Quotations from this letter will indicate the spirit in which the data were submitted

We are sending you a list of students from your school who have entered the University of Oregon at some time during the years 1929-30 to 1933-34; their records with us have been summarized to the beginning of the fall term of 1934-35. This is one result of a research study we have made, aiming to show what various types of students have accomplished with us, and to indicate what steps we may take to improve their educational guidance. We have thought you would be interested to receive this information for your school, both in order to know the subsequent educational success of your graduates and perhaps also to check back, using your early personal knowledge of the students, into their high-school work for possible reasons for later success or failure.

The February number of the *Oregon Education Journal* will contain an article explaining the factors used and the purpose of this report as a whole. For your immediate information we will briefly explain the types of data presented on the enclosed list.....

In order that you may compare your own school with the schools of the entire state, we have listed below the averages for 3,658 students from 223 Oregon preparatory schools.....

Four copies of this report have been made: one is enclosed; one may be sent to the State Department of Education; two will remain in our files. We believe that rigorous investigation but discreet use of the findings will benefit both the preparatory schools and the higher educational institutions. This study is but one indication of the co-operative research projects which could be organized. Our regular facilities do not permit much extra work, and, as a matter of fact, these reports might have been impossible without considerable clerical assistance furnished us under Federal Emergency Relief appropriations. However, we hope this information will be of interest and use to you as well as to us, and are glad to offer any further information or explanations which we can provide.

One anonymous but otherwise actual sample of these reports is shown in Figure 3. The blank forms were multigraphed at slight cost, and the data typed directly from the Hollerith cards which had been sorted by schools and alphabetized.

For our own files we have prepared summary tables showing averages for all matriculates from each school, both preparatory and advanced. As we suggested to the principals, comparison of a highschool's averages with the total averages for all Oregon schools yields a standardized measure of the graduates of that school. So few students come from many of the smaller schools that it may be years before our samples are definitely representative and reliable.

## PREDICTION EQUATIONS

In this section we shall describe, first, the background and purpose and, then, the procedures of our statistical treatment. Not all of the data previously enumerated were suitable for analysis by correlation but nine or ten objective factors or variables were used in this part of the study. Only those students were included who had been graduated or who had dropped out and were not registered during the last year of the study, 1933-34; thus the active students with indeterminate University records were omitted here. Six sets of correlation coefficients were computed, for different population groups. The entire group was divided into those matriculating directly from preparatory schools (henceforth designated "Prep.") and those with intervening work at another college or university (henceforth designated "Adv."). Within each of these divisions separate determinations were made for men, for women, and for both men and women. Sex differences proved small and unreliable, however, and the separate determinations were discontinued after zero-order correlation coefficients. The populations involved are large enough that the probable error in no case exceeds 0.03. These populations are: Prep. men, 1098; Prep. women, 967; Prep. men and women, 2065; Adv. men, 524; Adv. women, 496; Adv. men and women, 1020. The final regression equations, one set for Prep. men and women and one set for Adv. men and women, are therefore based on populations of over two thousand and over one thousand respectively.

The pattern of factors included in the analysis differs slightly for Prep. and Adv. groups, and the variables are listed below with the code letter used for each:

<i>Prep.</i>	<i>Adv.</i>	<i>Independent Variables</i>
<i>a</i>		Prep. decile (based on preparatory-school grades; decile 1 is lowest tenth)
	<i>A</i>	Adv. GPA (grade point average based on transferred college work; 4.00 is equivalent to "A" grade)
	<i>B</i>	Adv. credits (term hours of transferred college work)
<i>b</i>	<i>I</i>	Age at matriculation

		term GPA (grade point average first term in University)
<i>p</i>	<i>P</i>	Hours passed first term in University
<i>q</i>	<i>Q</i>	Hours failed first term in University
<i>Dependent Variables</i>		
<i>m</i>	<i>M</i>	Final GPA (to end of undergraduate career in University)
<i>k</i>	<i>K</i>	Number of terms registered (to end of undergraduate career in University)
<i>n</i>	<i>N</i>	Per cent completion of requirements for bachelor's degree (to end of undergraduate career in University)

Our purpose in this analysis was to find the best available basis for predicting, for each student, his most probable grades and the most probable length of time he would remain in the University. To do this we studied the predictive value of each independent variable for each dependent variable, as listed above. We desired to make the most accurate predictions that are possible at each of three times in a student's career: before matriculation, shortly after matriculation, and after the completion of one term in the University.

Of the six or seven independent variables, two-age and hours failed first term-proved to make no significant contribution. Previous grades, high-school or college, are significant in predicting University grades but not in predicting continuance; amount of previous college work, naturally, is significant in predicting continuance. Ability as shown on the psychological examination contributes to all the predictions until its values are absorbed in the greater significance of the first-term University record. Throughout the entire set of correlations, the predictive value of the first term is strikingly strong; we recognize that statistically its inclusion in the total introduces a spurious factor in correlations between them, but this is justified and made necessary by our objective of prediction.

It proved almost impossible to predict the number of terms a student would register in the University, partly because factors favorable to keep a student in school until graduation in the normal twelve terms would operate in the reverse direction for those unable to graduate until they had spent more than twelve terms-the more able student tends to continue to graduation but he does not tend to take more than twelve terms in the process.

5 substituted, therefore, consisting of the per cent of completion of the requirements for graduation; thus the student who graduated had

achieved 100 per cent on this basis, regardless of the length of time required, and a student had achieved 50 per cent when the quality and quantity of his work were equivalent to six of the twelve terms

normally required for graduation. Fairly reliable prediction may be made in this way, but much less definitely than is possible for point average; it seemed excessively unreliable to attempt the prediction of continuance until after the first term here.

Of the possible techniques for obtaining product-moment zero-order correlation coefficients, we chose to sort our punched Hollerith cards and to record data distributions on printed correlation charts; in Figure 4 is shown a sample chart, a simplified form developed here especially for use with a calculating machine. After subdivisions of the population are recorded in the proper compartments of the chart, the routine is direct and definite for finding means, standard deviations, and zero-order correlation coefficients. In the next step of analysis, simultaneous equations were formed and solved by the Doolittle method, leading to Beta regression coefficients, coefficients of determination, and multiple correlation coefficients. At this stage it is possible to strike a balance between the contributions of additional factors and the extra labor of including them in the final calculations. Afterwards, the final regression coefficients and equations are computed, and tables are prepared from which can be read off the predictions for any individual student.

Statistics of the first part of the analysis are listed in Tables I and II which follow. Means, standard deviations, and zero-order inter-correlation coefficients are shown for all the variables included in the study (see list of variables in second paragraph of this section). In every instance they are given first for men, then (in italics) for women, and at the bottom for both men and women.

From this stage on, sex differences were ignored and the subsequent analysis based on data for men and women together. Table III shows the multiple correlation coefficients obtainable from various combinations of independent variables, as well as their zero-order correlation coefficients, in predicting the dependent variables *m*, *n*, *M*, and *N*. Variables *k* and *K* were dropped, since they yield maximum multiple correlations of only .37 and .34 respectively as compared with those of .50 and .67 with the

K'I

oil loll

~aea~Ba aeonov n

Oil long IBM

nn

In mom

` 19NN~pppp~N~NN  
go mom BoomRoo  
mall In E1~OV~~ 0v  
~BBBB

Ice ~

all all 111110 son

some comel logo molin

TABLE I. MEANS, STANDARD DEVIATIONS, AND ZERO-ORDER CORRELATION Men  
COEFFICIENTS FOR PREP. GROUP: Women

[illegible]

7 TABLE II. MEANS, STANDARD DEVIATIONS, AND ZERO-ORDER CORRELATION  
Men

COEFFICIENTS FOR ADV. GROUP: Women											
Men and Women											
Variables: A B I C D P Q M % N Means:											
	2.149	70.743	21.970	6.645	2.253	13.768	0.830	2.264	4.229	59.367	
	2.518	81.177	22.654	6.999	2.644	14.581	0.398	2.563	4.044	64.375	2.330
	75.830	22.302	6.817	2.396	14.164	0.620	2.411	4.139	61.817	Standard	
Deviations:											
	0.594	41.625	3.578	2.783	0.734	4.249	1.945	0.694	2.662	38.450	0.549
	42.453	5.448	2.594	0.674	3.786	1.547	0.647	8.426	36.592	0.602	42.354
	4.595	2.699	0.717	4.051	1.776	0.688	2.551	37.642	Correlation Coefficients:		
A.....	254										
		.094	.316	.460			.187	-.244	.474	.084	.278
		.137	.199	.379	.533		.201	-.171	.586	.162	.265
		.220	.165	.348	.521		.213	-.237	.551	.102	.276
B .....	357										
			.121	.182	.125	-.180		.236	-.187		.442
			.368	.264	.203		.181	-.104	.261	-.238	.491
			.356	.189	.212		.162	-.159	.263	-.213	.479
Z.....		-.133				.066	.018	-.067	.088	-.109	.121
			.117	.170	-.026	-.069		.209	-.204		.133
			.011	.134		.001	-.068	.165	-.160		.129
C.....	11						.300	-.258	.428	.085	.217
							.498	.240	-.197	.608	.040
							.501	.278	-.238	.466	.063
											.245
D.....			.544	-.634	.838	.219					.438
							.356	-.636	.881	.160	.366
							.467	-.600	.863	.179	.407
P.....					-.432				.533		.541
								-.369	.341		.480
								-.413	.453		.518
Q .....						-.536		-.290	-.521	-.275	
M.....							.552	.493			
TABLE III. CORRELATION											
COEFFICIENTS FROM VARIOUS COMBINATIONS OF INDEPENDENT VARIABLES											
Time of	m	Corr.	n	Corr.	M	Corr.	N	Corr.			
Prediction	from	Coef. from		Coef. from		Coef. from		Coef. from			
Prediction	from	Coef. from		Coef. from		Coef. from		Coef. from			

mated scores should be comparable to those of actually earned scores. The "expanded" regression equations are as follows:

Before matriculation:  $m = 0.24 \pm 0.88$  ( $\pm 0.70$ )

$$M=1.04 \text{ A}+0.004 \text{ B}-0.33 \text{ } (\pm 0.67)$$

After matriculation:  $m = -0.16$   $a = 0.12$   $c = 0.65$  ( $\pm 0.59$ )

$$M = 0.76A + 0.003B + 0.12C - 0.40 (\pm 0.57)$$

After first term in  $m = 0.03 \text{ a} + 0.86 \text{ d} + 0.11 (\pm 0.24)$

University:

$$n = 22.48 \quad d+4.84 \quad p - 74.85 \quad (\pm 40)$$

M=0.18 A+0.87 D-0.10 ( $\pm 0.26$ ) N=0.51

B+11.33D+5.41 P-80.99( $\pm 28$ )

As an illustration of the increasing accuracy of successive equations, we may follow a matriculating freshman who comes directly from high school (Prep.), with (Prep. decile) a= 5, (Psy. decile) c=4, (first term GPA) d =1.7, (hours passed first term) p =16. The predictions of (final GPA) m at various times, and of (per cent completion for graduation) n are: .

Before matriculation:  $m = 2.08$ , with 50 per cent probability it lies between 1.38 and 2.78

After matriculation:  $m=1.93$ , with 50 per cent probability it lies between 1.34 and 2.52

(value of c lower than that of a reduces probability of high value of m, and addition of second variable restricts range of error) After first term in  $m = 1.72$ , with 50 per cent probability it University: lies between 1.48 and 1.96

(lower value of d than expected lowers predicted value and still further reduces range of error)

n=41, with 50 per cent probability it lies between 1 and 81

(prediction that student will complete 41 per cent of work for graduation, with wide limits of error but strong improbability that he will continue to graduation)

The predictive machinery resulting from this study should be invaluable for students and their advisers. We could construct prediction tables for each high school in Oregon, by which the principal could read his pupils' prospects in the University from high-school grades alone. Similar tables including also results of the entrance psychological examination could be (and have been)

advisers for new students entering the University. After the first term their prospects would be so well established that significant deviations below predicted norms would indicate to students, ad

visers, deans, and faculty committees the intrusion of serious obstacles to successful work. Psychological and social, as well as scholastic, difficulties might appear in their effect on academic work before coming to attention directly.

The whole predictive procedure attempts to combine the informa

tion from all factors of known significance for ready application to any student. We cannot anticipate all the occasions when such concentrated data will be needed, but the main concern of the registrar's office must be the achievements and records of scholarship, and these data bring also pertinent information from the personnel field. In fact, this study operates to correlate and evaluate the information obtained and needed by all the persons concerned with student achievement in the University.

Of course this type of investigation is not original with us. Perhaps our contribution is the Hollerith card which fulfills so many purposes. Our Hollerith machine installation consists only of punch and sorter, and a larger set of equipment would modify the techniques and introduce automatic procedures still further to decrease clerical work. Really *the* important point is that personnel and academic records can be expertly and objectively summarized in ways of unforeseen importance, and that vital uses will appear on every hand when the techniques have been provided. Whatever the methods may be which are best adapted to each school, this type of service will earn the registrar more recognition as a researcher as well as recorder, and as an