Success in Freshman Chemistry
A Predictive Analysis of Chemistry 115 at Purdue with a Single Class of 1100 Students in 1962

ROBERT EARL DAVIS[1], Purdue University

The Problem—Nearly all entering students at Purdue enroll in some freshman chemistry course. The enrollment has risen to the point that it severely taxed our crowded freshman laboratory facilities.

A faculty committee[2] on Graduate Student Teaching Duties was set up to examine the future needs and current problems. Laboratory work is most dear on time, facilities and cost. Therefore, the committee asked the question: can we predict success in freshman chemistry? If we could predict the failures before they enroll, we could prevent refluxing (a chemical condition in which students fail, are boiled up, condense and repeat the course next time around).

Previous Studies

The results of several other studies have been reported. Scofield (11) suggested that high performance in high school chemistry, high school physics and high school math usually predicated good performance in freshman chemistry at Syracuse University. However, in the group of students ranked "high" in high school work, numerous failures and nonhonor grades (D or lower) in chemistry were reported. Likewise in the "low" group, 14.5% of the students received A or B grades.

Herrmann (4) observed as a result of his statistical studies that high school chemistry is advantageous to success in college chemistry. He also noted that perhaps the students taking high school chemistry were more intelligent, more industrious, more apt in chemistry or even more "college-oriented". Steiner (12) added that the study of high school chemistry plays a significant part in the development of chemistry majors.

Martin (8) at Purdue confirmed the results of F. E. Brown: "those who have trouble with chemistry are generally poor in reading, writing and arithmetic". Martin commented on hypothyroidism but did not mention quantity data on the blood levels of protein-bound iodine on each student.

Willard (9) attempted an analysis of success in freshman chemistry using several criteria: high school chemistry, size of the home town, and the achievement tests given at the University of Wisconsin.

Willard confirmed that students with high school chemistry are as a group intellectually superior to the students not taking the high school course.

He also noted that while there is a direct positive relationship between grades and entrance achievement scores, extreme caution must be used when predicting a result for an individual student.

---

1. Alfred P. Sloan Fellow, 1962-1966.
2. Professor G. Urry, Chairman, R. E. Davis, W. F. Edgell, N. Kornblum, R. L. Livingston, and H. L. Pardue, members. Department of Chemistry, Purdue University.

Hadley (3) discussed the work of Scofield, Herrmann, Steiner and Clark showing that high school chemistry is advantageous and the work of West (13) (who concluded that general intelligence is more important than specific high school training). On the basis of the new study, Hadley concluded that high school math, physics and chemistry aided the most to success. He also concluded that most common grade for a student having all three courses was a B.

Carlin (2) determined the critical ratios (a statistical measure to determine if there is a difference between two averages) between six groups of students. He concluded that high school chemistry is an asset in college chemistry. He found that high school chemistry *and* high school physics contribute even more to success. But he concluded a course only in high school physics does not aid the performance in college chemistry at the freshman level.

Hovey and Krohn (5) noted that they started their studies because of the great increase in quantity (but not quality) of their students. They cited the increasing trend to use television (10) to decrease the staff requirements and the replacement of the first semester work by demonstrations (7). Hovey and Krohn concluded that they could predict 89% of the failures (F and drops) from the group of students making less than a 30% score on the Iowa Chemistry Aptitude Test, Form M (ICAT). The correlation coefficient was 0.51 for the ICAT test. The additional use of the Toledo Achievement test improved the correlation coefficient to 0.61. Such a number indicates a degree of success. But some students (about 11%) would be placed in a remedial, non-laboratory course when in fact they would pass the higher course with honors.

Kunhart, Olsen and Gammons (6) examined the placement of students to various courses using the high school chemistry grade, the high school algebra grade, the American Council of Education Psychological Examination linguistic score, the quantitative score and the total score. The extent of predictive success was low as reflected in the low positive values of correlation: the high school chemistry grade (0.263), algebra (0.205), A.C.E. quantitative score (0.178), A.C.E. total (0.122) and using *only* the A.C.E. linguistic score, the correlation was *only 0.071.*

Combining all five (as a sum) the multiple coefficient rose to 0.397. This number is not high enough to eliminate considerable scatter and false predictions of individual behavior. They admit that maturation of the students from high school to college age and the nonuniformity of grading in the various high schools probably cause the low correlations.

Brasted (1) used his data on 2500 students to analyze the teaching of high school chemistry. He further recommended that the college-bound student take math, physics and chemistry in high school.

The Present Study

The previous studies make it clear that exams with uniform grading given at the college would probably be better than using only high school grades. It is clear that no attempt was made to obtain the maximum amount of correlation using the data available on each students.

It was also the belief of the committee[2] that

"Success in Freshman Chemistry
is Success in Freshman Chemistry"
$r \equiv 1.00$

Therefore, a "hypothetical course" was constructed. Lecture and recitation work would run for eight weeks. During this time four or five quizs would be given and two hourly exams. Only the hour exams would be used. The quizs would be used only to prepare the students for the exams and let them know what types of chemistry questions might appear on the exam.

At the end of eight weeks, the course would be split into the upper section and the lower section. The upper section would receive lab work (two labs a week for the remaining eight weeks). The lower section would continue as a remedical section.

We can ask: How carefully can we place the students into the upper section?

We populated this hypothetical course with 1100 real students from Chemistry 115 at Purdue[3] in the fall of 1962.

The Data

At the end of the eighth week, the professors had the results of two hourly exams (Exam I and Exam II). When the students entered Purdue, we had scores on rather standard orientation exams given at Purdue in English, Math, Chemistry, and the percentage rank in high school (88 meaning 12th man in his class of 100).

We then attempted a correlation of these factors with the real class of almost 1100 students in Chem. 115 in the 1962 fall term. The final maximum score in Chem. 115 was 1000 points. The actual grade ranges were 1000-850 A, 849-726 B, 725-641 C, 640-576 D, and below 576, F.

In Table 1 the data are listed. The student's total score at the end of the course (sum of four hour exams, eight short quizs [best 8 of 10],

3. Chemistry 115 is described in the Purdue University Bulletin as follows:

115. GENERAL CHEMISTRY. Sem. 1 and 2. SS. Class 3, Lab. 3, cr. 4. Required of students majoring in chemistry, physics, and engineering who do not take CHM 117-126.

Laws and principles of chemistry, with special emphasis on topics of importance in engineering. Numerical problems and relationships are introduced whenever quantitative treatment is possible.

In 1962 the text book was H. H. Sisler, C. A. Vander Werf and A. W. Davidson, **College Chemistry**, 2nd edition, The MacMillan Co., New York, N. Y., 1961. The laboratory manual was D. W. Margerum, F. D. Martin and R. E. Davis, **Laboratory Manual for General Chemistry**, Tri-State Offset Co., Cincinnati, Ohio, 1962.

TABLE 1. Data Available on Each Student.

Primary Variable	Maximum Value	Label
Exam I	150	a
Exam II	150	b
Chemistry Orientation	100	c
English Orientation	100	d
Math Orientation	100	e
High School Rank	100	f
Secondary Variables		
$(a+b)$ $(c+d)$ $(c+e)$ $(c+f)$ $(c+d+e)$ $(c+d+f)$ $(c+e+f)$ $(d+e+f)$ (a^2) (b^2) (c^2) (d^2) (e^2) (f^2) (ab) (ac) (ad) (ae) (af) $[(a+b)c]$ $[(a+b)d]$ $[(a+b)e]$ $[(a+b)f]$ $(a+b)^2$ $(c+d+e)^2$ $(c+d+f)^2$ $(c+e+f)^2$ $(d+e+f)^2$ $(c+d+e+f)^2$ a^3 b^3 c^3 d^3 e^3 f^3 $(a+b)^3$ abc abd abe abf bcd bce bcf cde cdf def		

a recitation home work score, lab scores, and an instructor evaluation [maximum of 10% of the grade] to reward effort) and the values of his six primary variables were entered on the IBM data cards.

The secondary variables (as the product of the Math score and the English score) were computed by our program. While it may be difficult to visualize the meaning of such a product, it introduces the skewness needed to improve the goodness of fit.

Primary Variables. Scores of students.

Secondary Variables. Terms computed by the 7090 computer from the primary variables entered for each student.

Method-Multiple Regression

We assume that there is some functional relationship between the final score, Y, and the variables, X_i (either primary or secondary variables) and some constants, g, etc.

$$Y = g X_1 + h X_2 + \dots \qquad (1)$$

The question is then asked as to which variable alone (say X_j) does the best in predicting Y.

$$Y = k X_j \qquad (2)$$

The measure of the degree of fitting equation (2) can then be expressed by r, the correlation coefficient. The coefficient is computed from the standard error of estimate, s.

$$s = \sqrt{\frac{\Sigma (y - y_c)^2}{n}} \qquad (3)$$

y = true value of Y for one value of X

y_c = value calculated for Y using the equation (as 2)

n = number of points

$$r = \sqrt{1 - \left(\frac{s}{\sigma}\right)^2} \qquad (4)$$

where

$$\sigma = \sqrt{\frac{\Sigma Y^2}{n} - \bar{Y}^2} \qquad (5)$$

Thus if $r = +1.00$, all the points n fall exactly on the line (2). If r is 0.00, then there is absolutely no linear correlation at all.

The program is set up so that the second best term is then added to equation (2) to get a still better estimate of Y.

$$Y = k'X_j + 1_pX_m \qquad (3)$$

The calculation continues until the terms entering the equation no longer cause any significant improvement (as determined by the F test).

All calculations were made on the IBM 7090 computer. This program is available from us at Purdue.

Results

In Table 2 the results are listed after computation cycles 1, 2, 3 and 29. The variables are listed in order from the most significant (Exam I + Exam II) to the least significant. The coefficient for each term is listed with the standard error of the coefficient. On the 30th cycle our control card (card SØLØNG) terminated further calculations.

In Table 3 the results of some of the predictions are listed.

Of the F grades actually given, we correctly predicted 45.3% of their failures. One predicted (predicted 558) F student receive a high C grade (719 score). Only sixteen other predicted F students received C or low C grades.

The other predicted F students received D grades. No predicted F students received A or B grades.

No predicted A students failed and none received D grades. Some received C grades and many receive B grades.

As a general conclusion the grade can be rather accurately predicted to ± two grade letters.

Thus our study (which uses part of the college chemistry course) has a higher degree of predicability than those based only on high school or entrance scores.

TABLE 2

Cycle	Term	Coefficient	Standard Error of Coefficient	Standard Error of Estimate
1	X_1 = Sum Exams I and II	2.6802	0.07788	91.3
2	X_1 = Sum Exams I and II	2.46917	0.07804	
	X_2 = Chem. Orient. + High School Rank	0.09441	0.00996	87.7
3	X_1	3.7077	0.2736	
	X_2	0.0952	0.00986	
	$X_3 = (X_1)^3$	—11.176	2.369	86.8
29	X_1 (a + b)	—0.1491	2.5871	
	X_2 (c + f)	0.0437	0.2507	
	X_3	—22.966	24.757	
	(Chem. Orient.) ((Exam I)	—2.848	1.282	
	$(\text{Exam II})^3$	—47.954	10.979	
	(Chem. Orient.) (X_1)	1.736	0.788	
	$(\text{High School})^3$	0.738	0.217	
	$(\text{Chem. Orient.})^3$	0.513	0.206	
	$(\text{Exam I})^3$	251.583	110.21	
	$(\text{High School})^2$	—1.192	0.389	
	$(\text{Chem. Orient.})^2$	—0.673	0.366	
	Exam I	8.196	2.912	
	$(\text{Exam I})^2$	—82.850	30.516	

$(X_1)^2$	15.958	13.598	
(Eng.) (Math) (High School)	0.0495	0.212	
$(\text{Exam II})^2$	0.0213	0.0193	
$(\text{Exam II})^3$	—0.264	0.285	
$(\text{Math})^3$	0.887	0.274	
$(\text{Math})^2$	—1.339	0.506	
$(\text{Eng.} + \text{Math} + \text{Chem.} + \text{High School})^2$	—0.143	0.153	
(Eng. + Math + High School)	0.549	0.254	
$(\text{Math} + \text{Eng.})^2$	—0.0126	0.0506	
(Math) (Chem.) (High School) X_1	—1.393	1.032	
$(\text{Math} + \text{Chem.} + \text{High School})^2$	0.225	0.146	
(Math) (Chem.) (High School) (Exam I)	1.182	1.625	
(Math) (Chem.) (Eng.) (High School)	—0.419	0.312	
(Chem.) (Eng.)	0.274	0.232	
(Math) (Chem.) (High School)	0.616	0.682	83.9

TABLE 3. Random Selection of Predicted Values of the Total Score Computed Using the 29-Term Equation (from Table II)

Student	Actual Score	Score Predicted	Deviation
1	984	790.7	193.3
6	934	812.1	121.9
13	917	904.1	12.9
31	880	777.5	102.5
56[a]	852	875.4	—23.4
76	835	689.1	145.9
87	829	689.0	140.0
124	810	860.9	—50.9
159	795	855.3	—60.3
191	783	648.3	134.7
213	771	811.6	—40.6
272	744	566.5	177.5
295	738	578.3	159.7
326	728	569.1	158.9
440	695	860.4	—165.4
548	665	740.2	—75.2
566	655	519.1	135.9
686	617	390.4	226.6
687	617	628.9	—11.9
754	598	748.0	—150.0
788	585	590.2	—5.2
802	581	699.6	—118.6
872	551	589.9	—38.9
918	520	676.6	—156.6
974	471	464.4	6.6
995	440	692.6	—252.6
1033	349	349.8	—0.8
1056	126	722.2	—596.2
1057	76	314.9	—238.9

a) First occurrence of a minus deviation.

Literature Cited

1. BRASTED, ROBERT C. 1957. Achievement in first year college chemistry related to high school preparation. J. Chem. Ed. **34**:562-565.

2. CARLIN, JOHN J. 1957. Do courses in chemistry and physics at the high school level contribute to success in beginning college chemistry? J. Chem. Ed. **34**:25-26.

3. HADLEY, E. H., R. A. SCOTT, and K. A. VAN LENTE. 1953. The relationship of high school preparation to college chemistry grades. J. Chem. Ed. **30**: 311-313.

4. HERRMANN, GEORGE A. 1931. An analysis of freshman college chemistry grades wtih reference to previous study of chemistry. J. Chem. Ed. **8**:1376-1385.

5. HOVEY, NELSON W., and ALBERTINE KROHN. 1958. Predicting failures in general chemistry. J. Chem. Ed. **35**:507-509.

6. KUNHART, WILLIAM E., LIONEL R. OLSEN, and R. S. GAMMONS. 1958. Predicting success of junior college students in introductory chemistry. J. Chem. Ed. **35**:391.

7. MARQUARDT, D. N. 1957. Laboratory versus demonstration for the first semester of general chemistry. Div. of Chem. Ed., Amer. Chem. Soc., 132nd meeting, New York.

8. MARTIN, F. D. 1942. A diagnostic and remedial study of failures in freshman chemistry. J. Chem. Ed. **19**:274-277.

9. MCQUARY, J. P., H. V. WILLIAMS, and J. E. WILLARD. 1952. What factors determine student achievement in first-year college chemistry? J. Chem. Ed. **29**:460-464.

10. MORGAN, L. O., W. H. R. SHAW, and P. D. GARDNER. 1957. Laboratory instruction by closed circuit television. Div. of Chem. Ed., Amer. Chem. Soc., 131st meeting, Miami.

11. SCOFIELD, MAUDE B. 1930. Further studies on sectioning in general chemistry. J. Chem. Ed. **7**:117-126.

12. STEINER, L. E. 1932. Contribution of high school chemistry toward success in the college course. J. Chem. Ed. **9**:530-537.

13. WEST, G. A. 1932. School Sci. and Math. **32**:911-913.