ChE letters

EXAMINING TRENDS IN GRADING

Sir: In the period including May 1961 through May 1971, 173 students stood for the general examinations for the doctorate in chemical engineering at Princeton. During that period a number of minor procedural changes were made in the conduct of the examinations and a substantial number of changes were made in the faculty which formulated and graded these examinations. At the same time there appeared to be little change in number and quality of students applying for, and accepting admission to, the doctorate program. The numbers of foreign students had, however, increased appreciably.

Concern was expressed by some members of the faculty that the department was either grading or formulating the examinations progressively harder, or both. To test this hypothesis the grades were examined by an arbitrarily chosen empirical linear model containing as independent variables the date, based on zero time in May 1961 expressed in years, the years which a student had spent in residence before presenting himself for the examination, the fraction of foreign students in each group of common data and common experience, and an arbitrary index to indicate students who were taking the examination a second time, having failed on the first attempt. The number of students presenting themselves for each examination varied widely, from a minimum of one to a maximum of 13. It was always necessary therefore to use absolute grades since the numbers involved were insufficient for any normalized or otherwise adjusted curve.

The model chosen may be represented as

$$\hat{Y}(i) = \sum_{i}^{N} \cdot \sum_{k}^{p+1} b(k)x(ik) \qquad (1)^*$$

where

$k = 1,2,\ldots,(p+1)$, where p is the number of variables,

$i = 1,2,\ldots,N$ with n(j) the number of replicates at any one point in factor space;

$$\sum_{j} n(j) = N$$

where N is the total number of experimental points (i.e., grades available),

$x(ik) =$ the i-th value of variable k,

$b(k) =$ the coefficient estimated by a standard least squares procedure, and

$\hat{Y}(i) =$ the grade estimated by the model for the i-th student (i.e., the i-th value of the independent variable). Various powers of these variables and various interactions were included in the model, as indicated in Table I. The response was, of course, the numerical grade given. The results of this analysis are reported here in the hope that this sort of treatment may prove of interest to other departments who suspect similar or related problems.

The data were analyzed by a regression program reported by Daniel and Wood and available through SHARE or VIM*.** In order to minimize correlation between variables, the approximate average value of each variable was subtracted from each item of data. Thus the model was written in terms which were essentially deviations rather than the original variables. A number of passes were made to take advantage of the various features of this program. For example, as indicated in Table I, the second and third powers of time, the second power of experience, and the interaction of time and experience were included at various times to see whether their contributions to the sums of squares removed by the models contributed appreciably to improvement of the fit of the data by the empirical equation. The Mallows' criterion (see Daniel and Wood, **op. cit.**, pages 86-87) was used as an aid to judging the importance of these variables.

Two techniques were used to estimate whether any individual grade might not fit the general correlation or

TABLE I. Variables Investigated

X_1	— Time in Years of Residence from May 1961: 5.2 (calculated averages)
X_2	— Experience in years: 1.31 (calculated averages)
X_1^2, X_1^3, X_2^2, X_1X_2	— Combinations of variables initially thought to have possible influence and, consequently, included in the initial model
X_3	— A code indicating a second try of variable
X_4	— Fraction of foreign students

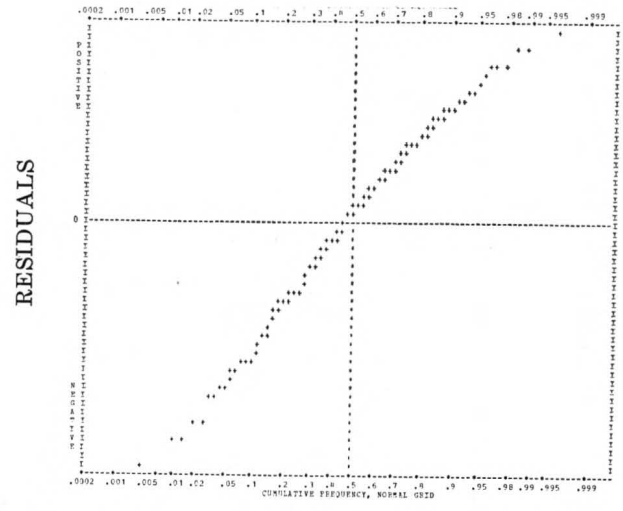

Figure 1 — Cumulative Distribution of Residuals.

* Note: Items in parentheses are to be subscripts.

** The numbers under which these programs are registered are: SHARE, No. 360D-13.6.008; VIM, No. G2-CAL-LINWOOD. Daniel C. and Wood, F.S., Fitting Equations to Data, Wiley Interscience, 1971.

might have undue influence on the values of the coefficients estimated. One was an examination of the residuals, representative plots of which are shown in Figures 1 and 2. The highest and lowest points which

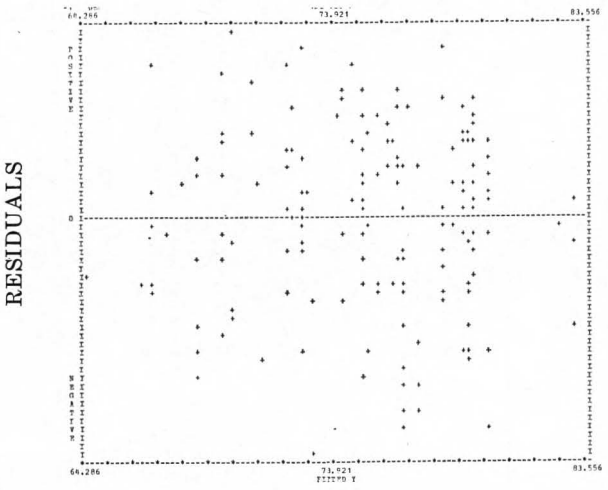

Figure 2 — Residuals vs. Fitted Y.

seemed to fall off the normal line were omitted and the data rerun. No appreciable influence of these points was observed, and they were returned to the data deck. The other technique is an examination of the relative influence which any point might have in establishing the estimate of the b(k) for any variable x(k). Three suspicious points (each being an exceptionally high grade) were detected by this technique and are therefore not included in the final analysis.

TABLE II. Study of Grades on General Examinations for Last Ten Years

Date is years from May 1961; DEVDAT is deviation from average for variable.

Exp'ce is years experience prior to first submission for exam.

DEVEXP is as with DEVDAT

DVD**2 and **3 are squares and cubes of DEVDAT

Model is linear combination of variables retained.

Ind. Var (I)	Name	Coef. B(I)	S. E. Coef.	T-Value
0		7.78096D 01		
1	DEVDAT	−1.33827D 00	2.17D−01	6.2
2	DEVEXP	−6.08825D 00	1.51D 00	4.0
3	DVD**2	−2.47099D−01	7.19D−02	3.4

No. of observations	170
No. of Ind. Variables	3
Residual Degrees of Freedom	166
F-Value	16.6
Residual Root Mean Square	7.33415476
Residual Mean Square	53.78982609
Residual Sum of Squares	8929.11113141
Total Sum of Squares	11606.40677941
Mult. Correl. Coef. Squared	.2307

The summary of the regression data is given by the computer output reproduced in Table II. It will be noted that only three parameters were needed to provide the best fit but that only about 23 percent of the original variance is accounted for by the regression model (see $R^2 = 0.2307$).

The failure of the R^2 statistic to act as a discriminating criterion of success for regression models is, of course, well known. In this instance it is very misleading since there are many replicates whose sum of squares should be removed from the remainder after accounting for regression (marked RESIDUAL SUM OF SQUARES in Table II) in order to leave a sum of squares estimating the lack of fit. The program will not perform this calculation. It does have a technique of searching "nearest neighbors" and converging on a number which should relate closely to the square root of the replication (i.e., "error") variance. In this case 6.2 to 6.3 appears to be a reasonable approximation of this standard deviation, suggesting that the error variance should be 38 to 40. Calculated independently from the truly replicated values (i.e., grades taken at the same time by students with the same months in residence), the error variance is 47.04. The program thus implies that the empirical model provides an excellent fit. In the experience of one of the authors the Daniel-Wood program tends to underestimate error variance when true replicates are available. It is not possible to decide whether this underestimate is a characteristic of the method and equally true when no true replicates exist. It is certainly a helpful estimate to provide some indication of the adequacy of the model if no true replicates exist.

TABLE III. Analysis of Variance for Lack of Fit

	Source of Variance	Sum of Squares	d. f.	Mean Square
	Total	11606.41	170	
	Due to regression	2677.30	4	
	Total from regression	8929.11	166	
*	Due to replication	6089.81	130	46.85
	Due to lack of fit	2839.30	36	78.87

$F(36,130) = 78.87/46.85 = 1.68$

R^2 after removing sum of squares due to replication

$$= 1.0 - \frac{2839.3}{11606.4} = 0.757$$

* Calculated independently

In the present case an exact technique for lack of fit can be applied, as shown in Table III; the sum of squares for replication ("error") is removed from those remaining from regression and the resulting sum of squares provides a mean square which can be tested by the usual techniques of analysis of variance for lack-of-fit. As noted in Table III, the F-statistic is in the 90-95 percent region for this distribution, indicating a 5-10 percent chance that the hypothesis of zero lack-of-fit is correct. While these odds are poor by absolute standards, they are excellent for purely empirical models.

From these calculations, a suitable model for the grades in the period in question is

(Continued on page 193)

. . . We do not have the coursework system with "homework" having to be handed in and marked.

FINANCIAL SUPPORT

Most of the financial support for graduate work in the U.K. comes from the Government via the Science Research Council, which is roughly analogous to the N.S.F. This body provides studentships both for graduate courses and for Ph.D. work. It also awards research contracts, although students are not normally supported in this way. Another source of support for students is research contracts from industry. At Loughborough over the past few years about half our Ph.D. students have been supported in this way. The research grants are nicely calculated to cover the student's bare living costs and pay his fees.

THE U.S. GRADUATE

Through frequent visits to the United States and also through our exchange with Georgia Institute of Technology, we have learnt of some of the problems which British graduates have when they go to do graduate work in the United States and of those which American graduates encounter here. The biggest problem our students find across the water is not the difficulty but the sheer volume of the work which they are expected to do. It is our impression, gained both from first-hand experience and from talking to students, that the quantity of work set in the average Master's degree in the United States is so great as to make it rather difficult for the student to take time off to pursue subjects on his own and to appreciate intelligently just what he is doing in an overall sense. On the other hand, since we do not have the course work system with "homework" having to be regularly handed in and marked, American students who come here find themselves very much at a loss for the first week or two. They are not used to our system which assumes that the student knows how to work for long periods on his own and which only covers in lectures a relatively few important topics. However, we have found that those students from the United States who have come to us have settled down quickly and progressed well. Both systems evidently have their merits. We prefer our own, but often find the results of the American system impressive. □

TRENDS: (Continued from page 149)

$$\hat{Y} = 77.81 - 1.338[x(1) - 5.2] - 0.2471[x(1)-5.2]^2 - 6.088[x(2)-1.31] \quad (2)$$

where

$\hat{Y}$ = average grade, estimated by Eq. (2),

$x(1)$ = time in years based on zero time in May 1961, and

$x(2)$ = time in residence, in years, prior to taking the examination.

The line represented by Equation (2), at an average experience such that $x(2) = 1.31$, is shown in Figure 3. Of the 35 data points available, only those for which $x = 1.3$ are included.

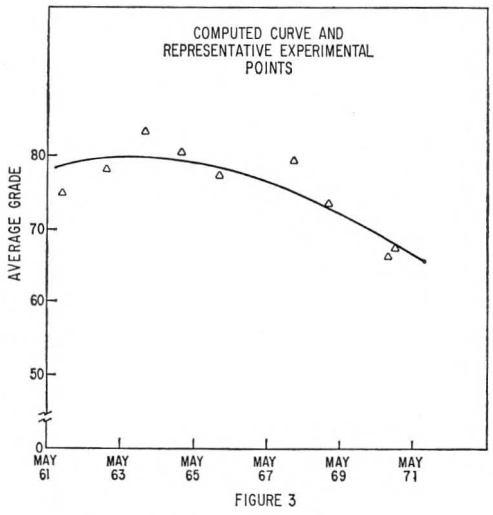

FIGURE 3

Figure 3 — Final Computed Average Grade vs. Date and Experimental Points.

On the basis of these results, the department has concluded that, in the period from May 1961 through May 1971:

1. There was a definite tendency for the grades to decrease and that this tendency accelerated in the later part of the period.
2. There was an apparent disadvantage in prolonging the time in residence before standing for the examination, the reason for which is not clear but the evidence therefore being incontrovertible from this analysis of the data.
3. None of the lowering of the grades in the later years of the period studied can be attributed to an increase in the proportion of foreign students, with potentially concurrent language and communication problems.
4. A correction should be made for the obviously increased difficulties which the faculty had suspected were progressively being built into the examinations; suitable action was taken in October 1971 with a gratifying improvement in the average grade of the eleven students who presented themselves for the examination.

J. C. Whitwell
L. Lapidus
Princeton University