DOCUMENT RESUME

ED 336 045	HE 024 867

AUTHOR	Hollings, Ann
TITLE	A Method for Assessing the Impact of Individual Course Marks on Overall Freshman Student Performance. AIR 1991 Annual Forum Paper.
PUB DATE	May 91
NOTE	27p.; Paper presented at the Annual Forum of the Association for Institutional Research (31st, San Francisco, CA, May 26-29, 1991).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; Chemistry; College Freshmen; *Difficulty Level; *Grade Point Average; Grading; Higher Education; Introductory Courses; Profiles; Science Instruction; *Statistical Analysis
IDENTIFIERS	*AIR Forum; *University of Guelph (Ontario)

ABSTRACT

Academic records of first year students enrolled in Introductory Chemistry at the University of Guelph, Canada, were used to develop a method to examine the effects of individual courses on freshman academic performance. At the center of the technique is a new test statistic useful in creating individual course profiles. Because grade point averages for describing academic performance can be misleading, Introductory Chemistry was selected as a test case because its reputation as a difficult course might readily produce effects on grade averages. The study used data from 11 years (1980-1990) of grade records for full-time first year students (N=10,184). Initial analysis demonstrated that having Introductory Chemistry in a timetable will on average decrease a student's average by .6 to 1.4 percent. However, that was found to be a relatively useless fact. Consequently a new formula was developed for determining how much the chemistry grade deviated from the average of the student's other grades. The resulting simple mathematical formula produced a statistic which was found to describe the effects of individual course grades on overall student performance while being intuitively sensible and easy to use. The paper includes 3 tables, 3 figures and 12 references. (JB)

A METHOD FOR ASSESSING THE IMPACT OF INDIVIDUAL COURSE MARKS ON OVERALL FRESHMAN STUDENT PERFORMANCE

Ann Hollings

System and Analysis Coordinator

Student-Environment Study Group

University of Guelph

Guelph, Ontario, Canada N1G 2W1

Telephone: (519) 824-4120, Ext. 2377

Acknowledgements

The author thanks Dr. O.B. Allen, Department of Mathematics and Statistics, Dr. Fred Evers, Department of Sociology, Dr. Alex Harrington, Paragon Engineering, and Dr. Michael Benjamin, Student-Environment Study Group for their careful reviews of this paper. Thanks are also due to Dave Montgomery, Mathematics and Statistics, for his input in the development process and Caryl Durst, Student-Environment Study Group for the literature search. Finally, thanks to Joe Consolo, Mathematics and Statistics for waiting all this time for his data.

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-First Annual Forum of the Association for Institutional Research held at The Westin St. Francis, San Francisco, California, May 26-29, 1991. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Chair and Editor
Forum Publications Editorial
Advisory Committee

A METHOD FOR ASSESSING THE IMPACT OF INDIVIDUAL COURSE MARKS ON OVERALL FRESHMAN STUDENT PERFORMANCE

Abstract

Academic records of first year students enrolled in Introductory Chemistry over an eleven year period were used in the development of a method to examine effects of individual courses on freshman population academic performance. At the centre of the technique is a new test statistic useful in creating individual course profiles, which could be an effective tool in assessment procedures. Possible applications of the technique are discussed.

A METHOD FOR ASSESSING THE IMPACT OF INDIVIDUAL COURSE MARKS ON OVERALL FRESHMAN STUDENT PERFORMANCE

Some time ago, a faculty member approached us in our role as an information management group at the University of Guelph, with a fairly simple request for data. During the ensuing discussion, it transpired that the professor had been recently appointed to a committee whose function was to implement new programs to help first year students in the transition to university. Bringing his own experience as a mathematics professor to bear on issues addressed largely by administrators, the professor talked at some length about the difficulties students often have in effectively handling particular first year courses. Students who had excelled in high school, he said, were often hit hardest by their failure to excel at university, and were particularly concerned by their unanticipated difficulty in courses such as Calculus and the legendary "Killer Chemistry".

This conversation stimulated some thought about whether such courses really do have significant effects on student performance - significant both statistically, and in the sense that they made a real difference to a given student population.

Consider hypothetical students A and B, both carrying five courses and both receiving identical lowest marks, 58%, in Introductory Chemistry. Student A's other marks are 59%, 60%, 61%, and 64%; Student B's other marks are 78%, 79%, 81% and 84%. Student A's average with Chemistry is 60.4%, without Chemistry it is 61%. Student B's average with Chemistry is 76%, without, it is 80.5%. While both students perform equally badly in Chemistry, the better student suffers much more than the poor student for a mediocre performance in a single course.

In addition, the ability of mark averages to describe their academic performance of these two students is quite inconsistent. In the case of Student A, 60.4% describes the overall performance quite well. But in the case of Student B, 76% captures neither how well the student performed in most of his courses, nor how poorly he performed in one course. In many cases, the use of an average of marks to describe academic performace can be misleading. Krzanowski, Mead and Thorne (1985) made note of this in describing a mixed model analysis of variance used to reduce the dilutionary effect of low marks in examination data.

At the student level, the hypothetical case of Students A and B illustrates the problem upon which this study focuses at the course level. In some instances, as for Introductory Chemistry, we intuitively "know" (we think) that the degree of course difficulty has a measurable effect on overall performance. Showing this to be true is quite another matter. The method developed in this study provides a simple yet powerful statistic which has the capacity to measure the extent to which individual courses exert an influence on academic averages in a population of students. The procedure is capable of producing course profiles which can provide a measure of either negative or positive effects of courses on student averages. Measuring the true impact of first year courses can assist in academic counselling and assessment of department policies with respect to heavily populated first year courses.

Literature Review

Both the focus and approach of the study described in this paper represent significant departures from the existing literature. Where current research deals primarily at the student level, in terms of student perception, student anxiety, and factors affecting performance, the effort in this study is to establish a method of extracting a profile which describes a course, particularly as it relates to overall performance of a population of students.

Course difficulty, especially in mathematics, has been approached from a number of different perspectives in higher education research. Methods of measuring the extent of math anxiety and the relationship between anxiety and performance, have been the focus of considerable attention (Siegel, Galassi and Ware, 1986; Cooper and Robinson, 1986; Adams and Holcomb, 1987).

Most research concludes (not surprisingly) that extent of anxiety about course material is inversely related to success in that course. Other studies have examined such factors as gender (Gadzella and Davenport, 1985), effectiveness of instructor (Goolsby, 1986) and predictive capacity of SAT scores (Boli, Adam and Payne, 1985) as they relate to success in difficult courses.

Without exception, these studies have been undertaken at the student level, often using data collected from psychological survey instruments which assign a score on an attitude or self-assessment scale. Some of the drawbacks to such research include the necessity of conducting surveys, and the difficulties associated with controlling differences in individual respondents' perception and understanding of the issues in the instrument.

Despite substantial interest in "difficult" subjects, studies which focus on the properties of the course itself, are either non-existent or extremely scarce. Course profiles could be instrumental in such processes as that suggested by Hanna and Cashin (1988), who recommended the establishment of a college grading system consistent across all instructors within a given course. Profiling could also be useful in the evaluation process within academic departments, providing an indication of fluctuations in student performance.

Data

The Introductory Chemistry course was selected as a test case because its reputation as a particularly difficult course suggested that effects, if they existed, would readily emerge. Enrolment in the course is very high (around thirteen hundred each semester), and consists of a fairly homogenous mix of students from all degree programs.

Examination of the Grade Summary Reports (Office of the Registrar, 1980 to 1990) suggest that Introductory Chemistry has had a consistent negative effect on the averages of the majority of students who have suffered through it. Over the years, 70% of students have had a decrease in their averages because of it. The failure rate for the course has hovered around the 25% mark over the ten year period, and the average mark in the course over the same period was about 58%. For almost 40% of the students registered in Introductory Chemistry, it was their lowest mark.

The analysis focused on full-time first year students from each fall semester from 1980 to 1990, and was limited to students carrying the usual full-time course load of five courses, to ensure that the relative weight of a single course was consistent across the database. Total size of the study population was 10,184.

The exclusion of part-time students, students carrying more than five courses and students in higher than first years had the effect of improving the appearance of performance in the course - the overall averages and failure rates were better for the study population than for the total population. Some of these differences for a few sample years are illustrated in Table 1.

Table 1

Selected Summary Statistics for Introductory Chemistry

Total and Study Populations

		Total Population	Study Population
Fall 80	Number	1,312	1,122
	Average	57.3%	63.2%
	% Failed	25%	24%
Fall 83	Number	1,304	1,037
	Average	58%	65.2%
	% Failed	28%	24%
Fall 86	Number	1,139	820
	Average	59.3%	64.9%
	% Failed	24%	18%
Fall 89	Number	1,154	673
	Average	59.4%	68.3%
	% Failed	24%	14%

Method

The initial approach to the problem was inspired by a re-sampling plan devised by Quenouille (1949), designed to decrease bias in estimates of variance components. Termed "jackknifing" by Tukey (1958), this method consists of the repeated structured subsampling of a sample of data. For a sample of size n, n subsamples each of size n-1 are formed by deleting each observation in turn and calculating the statistics on the remaining sample. In a similar manner, parameters describing academic performance in a population of students might be examined by systematically deleting one selected course mark from all marks of each student registered in that course, and re-calculating the parameters of mark distributions. Comparisons of the variances and averages with and without the course in question, might reveal much about the impact of that course on overall performance. As the intended application involved a population without sampling, the method was not so much a use of the jackknife as the development of a technique inspired by the jackknife.

Originally, the distributional properties of student averages and mark variances were of greatest interest. The rationale for the approach taken was this: for each student considered, if Introductory Chemistry was the lowest mark, then deleting that mark would have the effect of raising the average and decreasing the variance (because the range would be smaller). Therefore, if the course had a systematic effect on population performance, one would expect that comparisons of the mean averages and mean variances with and without the course would reveal significant differences.

Mean and variance were calculated twice for each student registered in Introductory Chemistry during an eleven year period: once including and once excluding the Chemistry mark.

The overall frequency distributions for mean and variance including Chemistry were then compared to the distributions excluding Chemistry. Paired t-tests were used to compare the mean average marks by year. This test was based on the following definition of DIFF:

$$DIFF_i = X'_i - X_i, \tag{1}$$

where X'_i = the average of the ith student's 4 courses excluding Chemistry

X_i = the average of the ith student's 5 courses.

The placement of X'_i in the equation ensures that the value of DIFF is generated with a sign which relects the direction of its overall effect.

The paired t-test tests the hypothesis that the value of DIFF is zero; in other words that there is no difference in average marks with and without Chemistry. For reasons explained below, no statistical test was used to compare the variances with and without Chemistry. Results of the analysis are reproduced in Table 2.

Table 2

Results of Analysis of DIFF

for Introductory Chemistry, 1980 to 1990

Year	Value of DIFF	Standard Deviation (averaged over all students)	
		with Chemistry	without Chemistry
1980	-1.4*	8.3	8.3
1981	-0.6*	8.6	8.6
1982	-0.7*	8.8	8.7
1983	-1.4*	8.4	8.0
1984	-1.1*	8.1	7.9
1985	-1.1*	7.7	7.5
1986	-0.8*	7.5	7.5
1987	-1.4*	8.1	7.6
1988	-1.1*	7.5	7.3
1989	-1.2*	7.9	7.8
1990	-0.8*	7.7	7.8

* For the test DIFF = 0, $p < .0001$

The initial results of this procedure were somewhat disappointing. As Table 2 confirms, the negative effect of Chemistry was highly significant in each year, yet these results represent one of those situations in which the statistics clearly indicate something which in practical terms is meaningless. Knowing, for example that having Introductory Chemistry in a timetable will on average decrease a student's average by .6% to 1.4% is not particularly enlightening or useful. For the many courses with less apparent but still significant effects, it was difficult to imagine how the results could be interpreted in any meaningful way. Prior knowledge of the course strongly suggested a substantial negative impact of Chemistry on population performance, and the DIFF procedure provided strong statistical evidence that the impact was real. Yet DIFF was defined in such a way that it could not provide a strong indicator that would clearly illustrate the scope and fluctuations of course impact. A large part of the difficulty lay in the construction of DIFF, particularly X_i, which effectively diluted the Chemistry mark in the averaging process.

In addition, the distributions of standard deviation with and without Chemistry were generated, but the differences in mean standard deviations were so small as to appear insignificant. While similarly small differences in mean averages did test as highly significant, there was a problem in finding or establishing an analogous statistical test for paired variances (or standard deviations), which would enable the appropriate significance level to be calculated.

These results led to some speculation about the approach to the problem. Perhaps the question was being asked in the wrong way. Perhaps the question was not how much does a student's average decrease when Chemistry is part of the timetable, but rather, by how much does the Chemistry mark deviate from the average of the student's other four marks?

This process led to the construction of a new statistic, called D, and defined:

$$D_i = C_i - X'_i, \quad (2)$$

where C_i = the ith student's Chemistry (or Course of interest) mark

X'_i = the average of the ith student's marks excluding C_i.

Mathematically, it can be easily demonstrated that D differs from DIFF only by a constant factor, and that both statistics address the same question in slightly different ways. In terms of relating the impact of a course on the performance of a population, D tends to be an indicator to which the association of course performance with overall performance can be readily made.

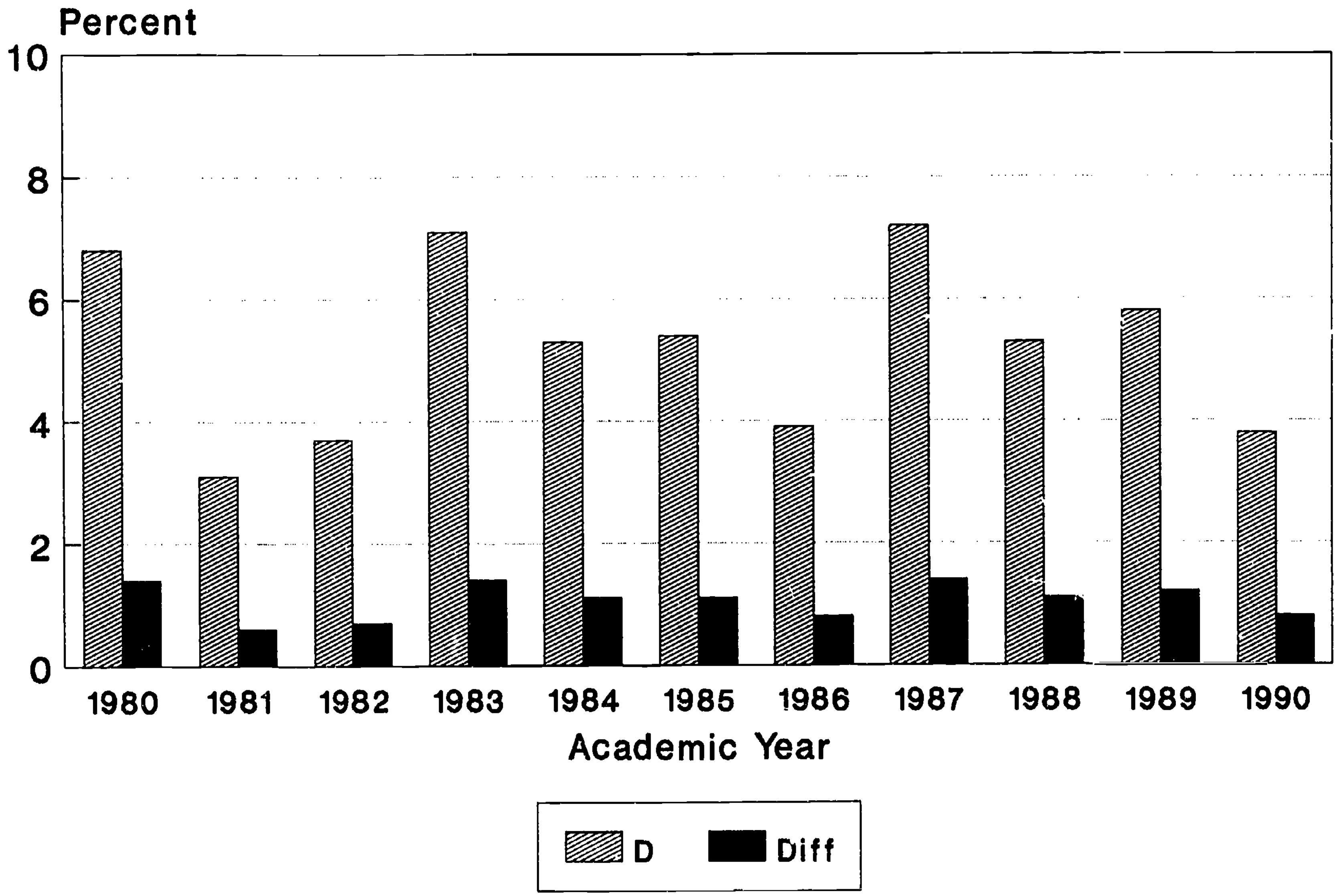
D is normally distributed with a mean which indicates precisely how closely student performance in course C differs from performance in any other four courses. For courses which are not particularly difficult or easy, the expected mean is 0. This can be tested using the standard test of the hypothesis that mean of the variable D is 0. The placement of C_i in the equation ensures that the value of D is generated with a sign which reflects the direction of its overall effect.

All results were generated using SAS (1985).

Results

Evaluation of the D statistic as an indicator of individual course effect was used to extract a profile of the trend over an eleven year period in Introductory Chemistry. Figure 1 illustrates the strength of D compared to DIFF for the eleven year period. As was the case for DIFF, in each year D is significantly less than 0 ($p<.0001$). In 1980, for example, students registered in Introductory Chemistry had, on average, a Chemistry mark that was almost 7% lower than the average of all their other marks.

Figure 1. Absolute Values of D and DIFF
Introductory Chemistry



All D and DIFF significantly
less than 0 (p<.0001)

Because D is a function of student performance in other courses, varying student quality is accounted for in the statistic. Just how this works at the student level can be determined by examining the hypothetical students A and B. Where Student A's DIFF value is -.6%, the D value is -3%, which illustrates that although the Chemistry mark is low, it is generally in keeping with the other marks that student achieved. For Student B, DIFF is -4.5%, but D is a considerable -22%, showing that this otherwise good student had substantial problems with this one course. If a student is generally good or generally bad, the value of D will be about the same. What D enables us to measure is precisely how much specific course performance deviates from general performance.

The general patterns exhibited by DIFF and D in Figure 1 are identical, as expected; however, the strength of D is in its ability to demonstrate fluctuations over time using a scale that is both meaningful and sensitive to change.

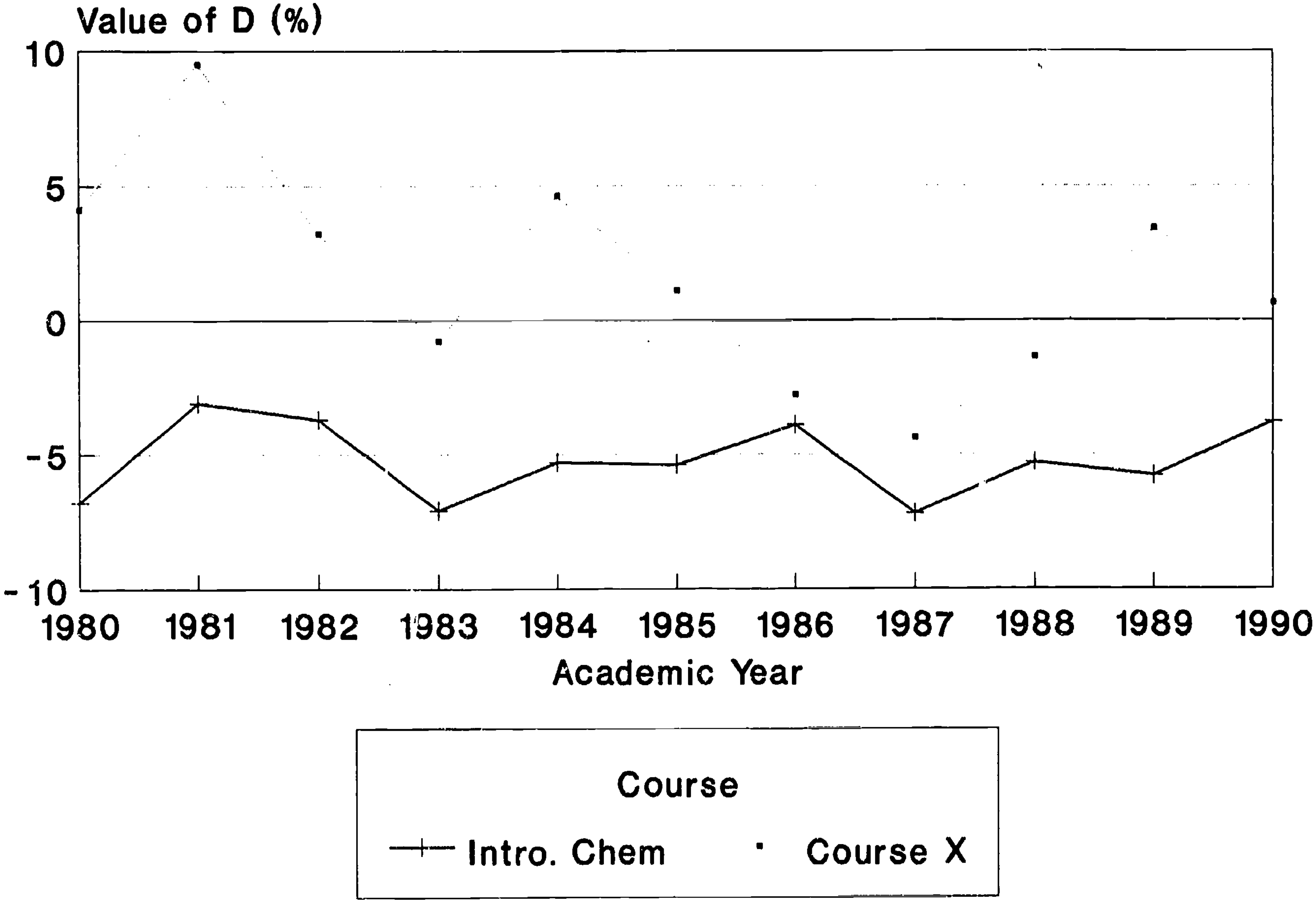
Extensions to other courses and subgroups

The procedure was used to generate a course profile for another first year course (coded Course X), which has had high enrolment over the past eleven years. Results are reproduced in Figure 2, along with the Introductory Chemistry results for comparisons.

Compared to the pattern displayed by Chemistry, the profile for the course of interest is extremely erratic, starting with a very high D in 1980. In later years, 1983, 1985 and 1988, D for this course was not significantly different from zero, and in some years, 1986 and 1987, it was significantly less.

These results suggest that it might be worthwhile for the originating department to examine teaching or testing procedures for the course, in order to establish some consistent processes which might stabilize this pattern.

Figure 2. Introductory Chemistry and Course X, D Profile 1980-1990



To examine high schools, a school board was selected from which large numbers of first year students had enrolled at the university. The study population was divided into two groups, those from the school board, and those from elsewhere. The Introductory Chemistry course was used as the course of interest. Figure 3 summarizes these results.

Consistently lower Ds for the sample board indicate that the impact of Chemistry on their overall performance was less pronounced than for other students. The danger here is in assuming that their performance in Chemistry was therefore better than other students'. D only tells us that the Chemistry marks for students from this board are closer to their other marks. One possible explanation is that they performed better in all courses, including Chemistry; another is that they performed worse in all courses, including Chemistry; still another is that their overall performance was comparable to other students', but they were better in Chemistry. The use of D alone is insufficient to accurately evaluate the impact of individual courses on population subsets.

To complete the picture, additional analyses were undertaken examining the differences in average Ds and average student averages, between students from the board and all other students. Results are summarized in Table 3.

General indications from Table 3 are that students from the sample board are indeed stronger in Chemistry than other students, as well as being generally better students. Their Ds are consistently lower, and their averages are consistently higher. Significant differences in both D and averages occur in some years, without appearing to conform to any pattern. Explanations for these occurrences are likely beyond the scope of this study.

The utility of course profiling on population subsets is somewhat limited by the need for additional information in the interpretation of the results. Taken solely as a measure of the

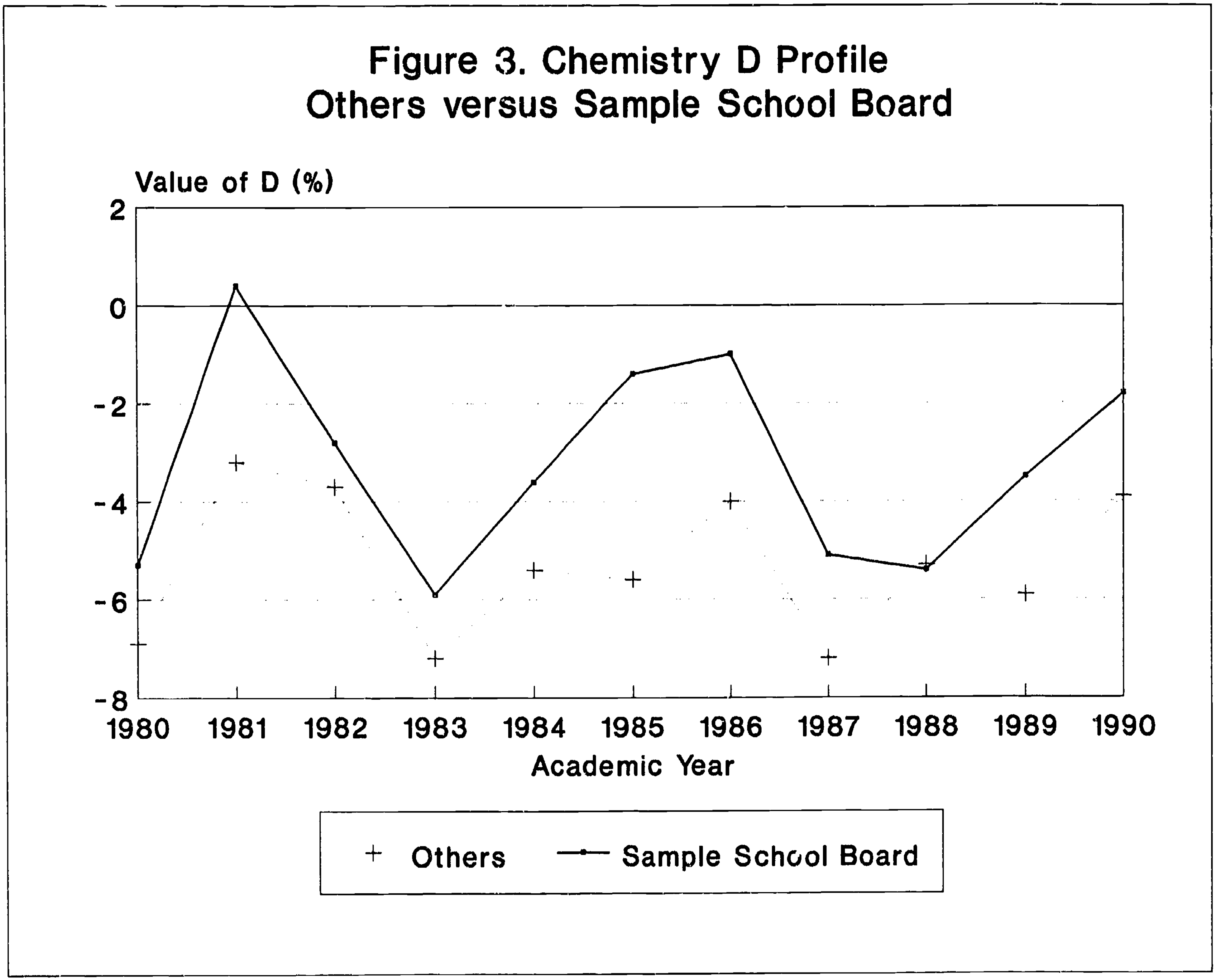


Figure 3. Chemistry D Profile
Others versus Sample School Board
Value of D (%)
2
0
-2
-4
-6
-8
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
Academic Year
Others
Sample School Board

Table 3

Comparisons of D and Overall Averages

Sample School Board and Other Students

Year	Board D	Others D	Board Average	Others Average
1980	-5.3	-6.9	62.4	63.2
1981	+0.4	-3.2**	64.7	64.2
1982	-2.8	-3.7	65.7	64.6
1983	-5.9	-7.2	67.8	64.8**
1984	-3.6	-5.4	65.3	64.5
1985	-1.4	-5.6***	66.5	63.6
1986	-1.0	-4.1*	68.6	64.8
1987	-5.1	-7.2	70.0	65.7*
1988	-5.4	-5.3	68.2	66.0
1989	-3.5	-5.9	73.2	68.2*
1990	-1.8	-3.9	67.8	68.4

For comparisons between either D or Overall Averages:

* $p < .05$; ** $p < .005$; *** $p < .0005$

impact of individual courses on the overall performance of population subsets, it may be useful in providing indicators of the relative strengths of certain groups. Used to describe population performance in individual courses, as it was originally designed to do, course profiling is at its simplest and most powerful.

Conditions on the use of course profiling

The course profiling method has been applied to traditional (with respect to course load and academic level) students in heavily populated first year courses. The results have shown that the procedure has enormous potential in describing population performance over time. There are some precautions which may have to be exercised in extending this method to the rest of the curriculum, to the entire population, or to subsets of a population. Some of these are offered below.

1) In addition to the possible atypical characteristics of students carrying either more or less than 5 courses, the derivation of D becomes a more complex and less reliable when a population contains students with varying numbers of courses. The impact of including all students has not yet been explored in detail, and is an area which future research will explore in detail.

2) The effect of population size has not been examined. For large populations, D appears to work well. The method may not be appropriate for courses which have small enrolments, such as lightly populated high level, professional or specialized courses.

3) Difficulty in interpreting results on population subsets may present a problem. Use of course profiling in population subsets may have to be accompanied by additional analyses to round out the results.

Conclusion

Course profiling is an effective means of describing effects of individual course marks on overall student performance. The statistic D is intuitively sensible and easy to use. Its conceptual simplicity does not compromise its credibility; in the words of Bradley Efron: "Good simple ideas...are our most precious intellectual commodity, so there is no need to apologize for the easy mathematical level." (Efron, 1982).

As for the professor who initiated this research with his innocent request for data, he is still waiting.

References

Adams, N.A. and Holcomb, W.R. (1986). Analysis of the relationship between anxiety about mathematics and performance. Psychological Reports 59 (October, Part 2), pp. 943-948.

Boli, J., Allen, M.L. and Payne, A. (1983). Women and men in introductory undergraduate mathematics and chemistry courses. Paper read at American Educational Research Association, Montreal, April 1983.

Cooper, S.E. and Robinson, D.A.G. (1986). The relationship of gender and anxiety on math performance among math-oriented college students. Paper read at Midwestern Psychological Association, Chicago, May 1986.

Efron, Bradley (1982) The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, J.W. Arrowsmith Ltd. Winterstoke Road, Bristol BS3 2NT, England.

Gadzella, B.M. and Davenport, J. (1985) Achievement and attitudes in mathematics. College Student Journal 19, pp. 398-403.

Goolsby, Charlee B. (1986). Affective variables and success in mathematics. Paper read at Conference for Exemplary Programs in Development Education, Atlanta.

Hanna, G.S. and Cashin,W.E. (1988). Improving college grading. Idea Paper No.19, Centre for Faculty Evaluation and Development, Kansas State University.

Krzanowski, W.J., Mead, R. and Thorne, S. (1985) On the use of average marks in degree examination data. Applied Statistics 34, No.3, pp. 219-226.

Quenouille, M. (1949). Approximate tests of correlation in time series. Journal of the Royal Statistical Society - Series B, 11, pp. 18-84.

Siegel, R.G., Galassi, J.P. and Ware, W.B. (1985). A comparison of two models for predicting mathematics performance: social learning versus math aptitude-anxiety. Journal of Counseling Psychology 32, pp. 531-538.

SAS Institute Inc. (1985). SAS User's Guide: Statistics, Version 5 Edition. Cary, North Carolina.

Tukey, J. (1958). Bias and confidence in not quite large samples, abstract. Annals of Mathematical Statistics, 29, pp. 132-148.