1-1-1995

Exploring Student Ratings Through Computer Analysis: A Method to Assist Instructional Development

Robert Lewis

Exploring Student Ratings Through Computer Analysis: A Method to Assist Instructional Development

Robert Lewis

St. Thomas University

This paper demonstrates how computer analysis of student ratings can assist professors in instructional development. Student ratings of three courses taught by the author were placed in computer databases and were then manipulated using BASIC programs. The needs of different client groups within a class were considered. Ratings of students who had different professional goals and differing preferences for learning goal structures were compared. The responses of students to nine different procedures within a class were analyzed to study the interaction between enjoyment, challenge, and learning on the class components. The author argues that computer analysis of the raw data from student course evaluations offers a valuable method for improving teaching practices.

Student ratings of university courses remain an established strategy for faculty evaluation. After reviewing the major studies of faculty evaluation procedures Cashion (1989) concluded, "Many believe — and I share that belief — that student ratings are the only primary data that are systematically gathered at many colleges and universities." (p. 4). The large number of studies exploring biasing factors (Abrami, d'Appollonia, & Cohen 1990; Benton, 1982; Gaski 1987; Marsh 1984) have resulted in general acknowledgement of the validity of

student rating forms (Cohen, 1981; Seldin, 1988). Baird (1987) found "that perceived learning is a significant predictor of both summative course and professor ratings" (p. 91). Aleamoni (1987) has pointed to numerous myths surrounding the use of student ratings, and stressed that "...gathering student ratings can provide the instructor with first hand information on the accomplishment of particular educational goals and on the level of satisfaction with and influence of various course elements" (p. 144). But the validity of course ratings is a question of little importance if the results are not carefully used in attempting to reach one of the three purposes of faculty evaluation defined by Cohen (1980): 1) to aid in administrative decisions, 2) student course selection and 3) instructional improvement. This paper demonstrates the use of student ratings for improvement of teaching, focusing on ways to tease out information that is unavailable from the usual reports consisting of summary or normative data.

There is conflicting evidence about the value of student ratings for aiding faculty development. Wright and O'Neill (1994) surveyed university staff involved in faculty development and found little confidence in the power of summative course evaluations alone to stimulate teaching development. They suggested that rating systems function more as part of an institutional press rather than providing specific information or insights upon which to build better teaching and learning. But institutional press may at times have a negative effect. Davey and Sell (1985) described a tendency among junior faculty at a large research university to "compromise the personal and organizational purpose of ongoing improvement and development for the practical requirements surrounding the promotion and tenure process" (p. 63). They further suggested student ratings were used in preference to evaluation methods with more potential meaning. There is potential value in student ratings, however. Marlin (1987) found that students feel they can make fair and accurate judgments with little bias, but also that they have little confidence that their views are used. In a longitudinal study, Stevens and Aleamoni (1988) found that faculty who used evaluation data made more use of resources and had better student ratings. Centra (1979) concluded that instructional improvement is greatest when a discrepancy exists between the teacher's self-rating and student ratings and when consultation is

available. However he concluded, "Student ratings may lead to some changes when only the teachers see the results, but there are probably many ways to increase their impact" (p. 38).

Computer analysis of their students' course evaluations by individual professors may offer one way to increase impact. Databases can be constructed from ratings in ways that effectively hide the identity of raters while allowing teachers to explore the ratings in more depth. Different groups within a class may have very different responses, and if identified without violating rater anonymity a professor may find helpful clues to better serving all students. Knowledge of how those who rate one aspect of a class low (or high) react to other aspects of a class might offer clues to improved class structures. Computer analysis of raw data from student ratings, if combined with that from other sources, can add to the diagnostic strength of a faculty evaluation system, a factor included by Cashion (1990) in a list of recommendations for programs. The added depth of exploration may even help avoid assigning undue precision to ratings, a problem cautioned against by Centra (1979).

In 1985 I began constructing my own course evaluations, which were then transformed into databases of student responses which could be manipulated by computer. Descriptions of three such procedures follow, each designed to analyze a different course structure. They are offered as examples of how summative course evaluation can be individualized and the raw data manipulated to answer professor generated research questions about teaching and learning. In each case the paper explores only one or two questions, however the exploration involved several others and raised many additional questions.

A course serving three groups

In 1985 I taught a course in Educational Psychology, required of Bachelor of Education students and elective for Bachelor of Arts students. There were three distinct groups, each using the course for a different purpose. Psychology majors entered the course to complete their programs. B.A. students with some psychology background often used the course to explore education as a career. Education

students with a variety of backgrounds in psychology were required to enroll unless exempted because of previous courses. I had serious concerns about whether the course was serving all students well, and whether the selected areas of Educational Psychology were appropriate for all. It was impossible to tell when all students were grouped together in course rating reports.

TABLE 1
Mean Ratings of Four Groups in Educational Psychology, Dec. 6, 1985.

Ranked on a scale of:
   1)Very low 2)Below average 3)Average 4)Above average 5)Very high
Breakdown of ratings for groups below:
(1) Majors in Psychology (2) 3 or more Psych courses
(3) Introductory Psych only (4) No previous Psychology

	(1)	(2)	(3)	(4)
n= 18	18	19	21	9
How much did this course replicate your past learnings in:				
Unit I Cognitive theory and development	3.56	3.05	2.84	1.56
Unit II Educational Planning and Measurement	2.89	2.53	2.58	1.56
Unit III Behaviorism, Social Learn., Motivation	3.72	3.21	3.06	1.44
To what extent did the course develop your:				
Factual knowledge	3.79	3.47	3.73	3.55
Principles and theories	3.53	3.56	3.68	3.64
Understanding of the discipline.				
Rate the amount and kind of work involved in this course:				
Amount of reading required	3.58	3.32	3.41	3.27
Amount of non-reading work	2.84	2.95	2.71	3.27
Difficulty of course	3.11	3.16	3.23	3.09
How effective was the professor in accomplishing:				
Communication of goals and content	3.95	4.37	4.27	4.45
Involving students in the course	4.11	4.5	4.05	4.09
Creating enthusiasm/stimulating effort	3.89	4.21	4.23	4.27
Evaluating learning accurately	3.79	4.05	4.43	4.45
How much did the following procedures aid your learning?				
Lectures and class presentations	3.47	3.47	3.36	4.00
Quality of text and handouts	3.58	3.61	4.05	4.00
Contract system of measurement	3.74	4.05	4.00	4.09

The course evaluation shown in Table 1 was completed by 67 of the 74 students in the class. Student volunteers entered the data into a DOS Text file. I was able to analyze the judgments made by each client group through sorting, while maintaining the covert nature of the ratings. Table 1 is a replication of the computer printout obtained by sorting for psychology background.

The most important outcome of this analysis was the endorsement it offered of the selected content and procedures. Differences between the four groups on familiarity of material were striking, while all other ratings failed to show dramatic differences. There did seem to be less reliance on professor and course structures on the part of psychology majors however the demands of the class were not dissimilar for the four groups. The contract system, given the highest rating of course procedures, may have been a mediating factor.

A similar sort comparing B.Ed. student ratings to those of B.A. students showed few differences when psychology background was controlled.

A course involving three teachers

By 1993 the Educational Psychology course evaluated in 1985 had been redesigned. A module structure was devised, and three professors taught the modules. Each module had its own structure and procedures for evaluation of student learning. The university's mandatory course evaluation form was inappropriate for this structure, however the policy allowed an individually designed evaluation. Four questions were important to answer for each module:

1) How important was this module to your overall program of study?
2) How effective was the "constructive mismatch" - was the material within your abilities yet challenging?
3) How effective was the teaching in reaching the module's objectives?
4) How appropriate and valid was the evaluation in assessing your learning?

Several demographic variables were included on the course evaluation form; the most important being the level of student teaching the student had chosen, and whether the person preferred cooperative, individual, or competitive goal structures in learning. With modules varying dramatically in method and measurement procedures, the overall effect of the course on students was of particular interest and it was reasonable to ask whether demographic factors had any effect on student assessment of the methods or evaluation system. Table 2 shows the mean ratings for each demographic group.

The most interesting outcome of the computer analysis was the differing views on evaluation procedures between those oriented toward the elementary school and those aiming for junior high school or high school teaching. This was especially evident in two modules I taught, which involved contract (Metacognition) and portfolio (Learning Styles) evaluation. A third, (Cooperative Learning), involving elements of a contract but closer to traditional measurement, showed a smaller difference. The basis for this difference still eludes me, however it seems that elementary candidates feel more traditional evaluation methods reflect their learning better. This is an area to be explored, either in subsequent evaluations or by using formative methods. It is of particular interest considering the faculty's emphasis on the use of less traditional methods in evaluating elementary students.

A course with nine components

It is especially difficult to sort out the effects various teaching procedures have on the learning of students. In 1991-92 I taught a course in two successive semesters for which I used a course structure and textbook approach that was completely new to me. As a visiting professor I could not alter text (readings only) or course entry requirements, even though I was initially uncomfortable with them. I designed a course which required weekly critiques of the articles, small group processing, some amount of experiential learning and large group discussion. A research paper, midterm with rewrite, and a final exhibition filled out the required elements of the course. I was eager to obtain student responses, for I was breaking new ground for myself.

What were the interactive effects of each of the elements, and what part did the varying demands and motivational properties of the

TABLE 2
Means of Educational Psychology Course Evaluations
Sorted by Preferred Teaching Level and Goal Structure

Scale - 1 (lowest) to 7 (highest)

	Chosen Level of Teaching			Preferred Goal Structure		
	Elem.	JHS	SHS	Co-op.	Compet.	Indiv.
Nature-Nurture						
Importance	6.13	5.69	5.76	5.91	5.33	5.81
Mismatch	6	5	5.13	5.53	4.67	5.81
Teaching	6.44	6.67	6.69	6.72	6.33	6.56
Evaluation	6.38	6.33	6.38	6.5	6	6.38
Metacognition						
Importance	4.94	5	5.25	5.14	5.38	5.35
Mismatch	4.82	5.1	5.3	5.31	4.88	5
Teaching	5	5.45	5.6	5.6	5.43	5.4
Evaluation	4.39	6	5.95	5.39	5.86	5.3
Special Education						
Importance	6.36	6.13	6.13	6.56	5.8	5.46
Mismatch	6.18	5.63	5.73	6.2	5.2	5.38
Teaching	6.55	6.44	6.53	6.72	5.4	6.23
Evaluation	6.09	6	5.93	6.24	5.4	5.77
Learning Styles						
Importance	6	5.7	5.85	6.05	5.2	5.81
Mismatch	6.5	5.65	5.8	5.95	5.4	6
Teaching	6	5.75	5.85	5.86	5.6	6
Evaluation	4	5.32	5.58	5.5	5.5	5.56
Co-operative Learning						
Importance	5.84	4.93	5.27	5.71	4.71	5.71
Mismatch	5.32	5	5.07	5.47	4.86	5.24
Teaching	5.26	5.53	5.8	5.65	5	5.41
Evaluation	4.94	5.4	5.53	5.34	5	5.29
Classroom Management/Motivation						
Importance	6.2	6.38	6.44	6.36	6.33	6
Mismatch	5	5	5.22	5.21	5.33	3.78
Teaching	5.6	5.63	5.56	5.29	6	4.56
Evaluation	5.2	6.38	6.33	5.71	5.67	5.11

Note: Based on 55 course evaluations from 60 students
The number of students in each demograhic group varies for each module

components play in student learning? Table 3 shows the accumulated ratings from all twenty member of the class.

TABLE 3
Course evaluation of Advanced
Educational Psychology, EDI 361
Summary data for 20 students of 20 enrolled

Evaluation of Course Components:
Lowest rating = 1 Highest rating = 5

	Challenge					Enjoyment					Learning				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
The readings themselves:		2	7	6	5			1	9	10		1		14	5
Writing the critiques:			4	6	8	1	3	6	7	3		1	2	11	5
Prof's memos and marks:	1		2	5	11			3	5	11	1		2	3	14
Small group discussions:		1	2	9	7		1	3	7	9		1	2	7	9
Large group discussion:		1	5	6	6	1	1	3	7	8			3	7	9
Class activities (other than group discussions)			3	9	7			3	6	10			4	3	12
Writing research paper			4	2	14	2	1	7	4	5			2	5	12
Mid term exam/rewrite			2	3	14	4	4	5	4	2	2	1	3	3	10
Final exhibition			2	4	14	1		1	4	14			1	6	13

The results shown in Table 3 tended to confirm my own observations, as such results often do. Despite the overall positive view offered by this survey, some puzzling questions remained. What effect does a low rating on challenge or enjoyment have on a student's learning in the class? Is there a "silent minority" within the class whose needs are not well served by the structure? Is challenge more important for learning in some elements and enjoyment more important elsewhere? To explore these questions, I wrote a program which could divide the group on any factor, displaying comparison ratings on all other elements. Since the writing of critiques was a critical requirement, upon which many others depended, I focused on it first. Table 4 shows the computer display of the mean rankings on all categories for the ten students who rated enjoyment of the critiques as 1, 2, or 3 (FOCUS) compared to the ten students who rated writing critiques as 4 or 5 on

enjoyment (OTHER). The difference between the two groups is also shown (DIFF).

Element	Challenge			Enjoyment			Learning		
	Focus	Other	Diff	Focus	Other	Diff	Focus	Other	Diff
Readings	3.5	3.9	-.4	3.4	3.5	-.1	4.2	4.1	.1
Critiques	4.11	4.33	-.22	2.5	4.3	-1.8	4.11	4	.11
Memos-marks	4.44	4.2	.24	4.44	4.4	.04	4.2	4.7	-.5
Small gp.	4.22	4.1	.12	4.4	4	.4	4.5	4.1	.4
Large gp.	3.78	4.11	-.33	4	4	0	4.1	4.6	-.5
Activities	4.3	4.11	.19	4.3	4.44	-.14	4.6	4.22	.38
Res. Paper	4.5	4.5	0	2.8	4.22	-1.42	4.3	4.78	-.48
Mid-term	4.89	4.4	.49	2.78	2.8	-.02	4.22	3.7	.52
Final Exhib	4.6	4.6	0	4.8	4.2	.6	4.8	4.4	.4

TABLE 4
Means of Raters Who Judged Enjoyment of Writing Critiques 1-3 Compared to Means of all Other Raters

It is clear that those who found little enjoyment in writing critiques were not an alienated minority. In nearly all other ways they were similar to those who liked writing the critiques. They reported effective learning from all class components, including writing critiques. It may be that these ten students simply dislike writing or prefer to respond rather than to initiate, since the rating on enjoyment of the research paper also shows a large difference. The same kind of analysis was possible using any of the rated categories, and permitted me to explore other hypotheses derived from study of the data in Table 3. I found no evidence of an alienated minority.

Interestingly, those who rated enjoyment of writing the critiques low rated their learning from the midterm higher than their colleagues who enjoyed writing critiques. This, combined with the similar "enjoyment" differences of these two groups on writing the research paper led to considering the correlations between enjoyment and learning and between challenge and learning. The correlations for all course elements are shown in Tables 5 and 6.

TABLE 5
Correlations between Challenge and Learning of Class
Components in EDI 361

		read	crit	memo	sm gp	lg gp	activ.	paper	mid-t	exhib.
					Attributed learning					
C	reading	.31	.40	.65	.17	.51	.27	.41	.43	.14
H	critiques	.34	.43	.05	.25	.34	.34	.66	.35	-.27
A	memos/marks	.08	.24	.70	.26	.28	.24	.17	.30	.56
L	small groups	.05	.10	.01	.55	.44	.51	.13	.41	.02
L	large group	-.08	.16	.41	.23	.59	.19	.16	-.01	.05
E	activities	.04	.39	.06	.51	.46	.77	-.12	.62	.36
N	term paper	.52	.46	-.10	-.08	.32	.17	.54	.25	-.22
G	mid term	.33	.35	.29	.38	.34	.58	.04	.61	.60
E	final exhib	.02	.14	-.03	.66	.60	.51	.13	.44	.10

TABLE 6
Correlations between Enjoyment and Learning of Class
Components in EDI 361

		read	crit	memo	sm gp	lg gp	activ.	paper	mid-t	+ exhib
					Attributed learning					
E	readings	-.05	.42	.16	.43	.68	.67	.03	.68	.37
N	critiques	-.09	.10	.47	-.19	.46	-.06	.45	-.11	-.23
J	memos/marks	.03	.42	.83	.23	.62	.63	.38	.52	.52
O	small groups	.29	.22	.23	.80	.60	.66	-.10	.56	.55
Y	large groups	-.07	.38	.26	.43	.69	.50	.10	.49	.46
M	activities	-.12	.25	.16	.12	.61	.52	.16	.23	.18
E	term paper	.16	.28	.58	-.28	.32	.27	.71	.18	-.17
N	mid term	.15	.35	.41	.20	.40	.60	.20	.75	.17
T	final exhib	-.12	.24	-.03	.12	.17	.44	.33	.43	.52

It seems that enjoyment has a higher relationship with learning than challenge. The particularly higher relationship between these two factors with respect to memos and marks offers some food for thought. Challenge is particularly important to the class activities. Both of these findings were contrary to my intuition, and both have influenced me

to rethink the manner in which I approach assigned work and class activities.

The survey offered a wealth of information about methods I might use in other classes. The ratings on challenge, enjoyment, and learning offered a gauge of motivation potential for the various elements, and the correlations suggested numerous hypotheses. The computer program made it possible to "play" with the results to follow any hunches that were suggested by the results.

Summary

This paper described the use of computer manipulation of student course evaluations to explore reactions of various demographic groups to course elements or to analyze the interrelations between components of a class. Summative course ratings can be given more meaning if subgroups can be compared and the interaction of student ratings of course elements can be analyzed. Although there is some danger in attributing more precision to ratings than they deserve, computer analysis allows student views to be analyzed in more depth. The very process of analysis tends to reduce overgeneralization. Rather than looking for confirmation of quality teaching or dreading negative ratings, one is drawn to anomaly and irregularity. The ratings shown here were generally positive, yet they offered many hypotheses which led to further study and course changes. If changes actually are derived from student ratings, and that fact is shared with students, it may further their willingness to offer constructive views.

Computer databases, if made available to professors, can provide anonymity for raters while allowing professors the chance to analyze student views in detail. The ratings used in this study were done outside (and in some cases replaced) the official course evaluations, but university wide ratings forms could easily be made available with access limited to each professor's classes. Database or spreadsheet software could easily be adapted for analysis. Allowing faculty access to databases of their student evaluations should provide a powerful tool for teaching analysis and may serve to improve teaching and the faculty evaluation system.

References

Abrami, P. C., d'Appollonia, S., and Cohen, P.A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*, 219-231

Aleamoni, L. M. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education 1*, 111-119

Baird, J.S., Jr. (1987). Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology, 79*, 90-91.

Benton, S.E. (1982). *Rating college teaching: Criterion validity studies of student evaluation of instruction instruments.* (AAHE-ERIC/Higher Education Rep. No. 1). Washington, DC: American Association for Higher Education

Cashion, W.E. (1989). *Defining and Evaluating College Teaching.* (IDEA Paper No. 21) Manhattan, KS: Center for Faculty Evaluation and Development.

Cashion, W.E. (1990). *Student ratings of teaching: Recommendations for use.* IDEA Paper No. 22) Manhattan, KS: Center for Faculty Evaluation and Development.

Centra, J.A. (1979). *Determining Faculty Effectiveness.* San Francisco, CA: Jossey-Bass.

Cohen, P.A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education 13*, 321-341

Cohen, P.A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*, 281-309

Davey, K. B. & Sell, G. R. (1985). *Instructional evaluation for development/improvement: Fact or fiction based on a case study of faculty practices.* (Paper presented at the annual meeting of the American Educational Research Association, Division J, 69th Chicago IL., Mar 31-April 4 1985) ERIC Document Reproduction Service No. ED 260 603

Gaski, J. F. (1987). On "construct validity of measures of college teaching effectiveness". *Journal of Educational Psychology, 79*, 326-330

Marsh, H.W. (1984). Validity of student's' evaluation of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology, 74*, 264-279

Marlin, J.W. (1987). Student perception of end-of-course evaluations. *Journal of Higher Education, 58*, 704-716

Seldin, P. (1988). Evaluating college teaching. In Young, R.E. and Eble, K.E. (Eds.), *College teaching and learning: Preparing for new commitments.* San Franscisco, CA; Jossey Bass.

Stevens J.J & Aleamoni, L.M. (1988). The use of evaluative Feedback for Instructional Improvement. In Arreola (Ed.), *Handbook for development of a comprehensive faculty evaluation system.* Center for Educational Development and Assessment, Memphis TN. (Reprinted from Instructional Science 13 (1985) pp. 285-305)

Wright, W. A. & O'Neil, M. C. (1994). Teaching improvement practice: New Perspectives. *To Improve the Academy 13*, 5-37