

# The Relationships Between Student Ratings and Instructor Behavior: Implications for Improving Teaching

PATRICIA A. CRANTON\*, WILLIAM HILLGARTNER\*\*

---

## ABSTRACT

*Although students' ratings of instruction have been examined in detail by educational researchers, the relationship between ratings and actual classroom behavior has not often been investigated. This study explores the relationship between student ratings and classroom observations. Twenty-eight professors from a wide range of academic disciplines participated in the study. Mean student ratings and frequencies of behavior in several categories were obtained for each professor. It was found that instructor behavior significantly predicted student questionnaire responses in three general areas. (1) When instructors spent time structuring classes and explaining relationships, students gave higher ratings on logical organization items. (2) When professors praised student behavior, asked questions and clarified or elaborated on student responses, ratings on the effectiveness of discussion leading were higher. (3) When instructor time was spent in discussions, praising student behavior, and silence (waiting for answers), students tended to rate the classroom atmosphere as being one which encourages learning.*

---

## RESUMÉ

*Si les évaluations faites par les étudiants sur le mode d'enseignement ont été étudiées en détail par les chercheurs des sciences de l'éducation, le rapport entre ces évaluations et le comportement réel en salle de classe a rarement retenu l'attention des chercheurs. Notre étude porte sur le rapport entre les évaluations faites par les étudiants et les observations faites en classe. Vingt-huit professeurs représentant une vaste gamme de disciplines universitaires ont participé à cette étude. Pour chaque professeur, nous avons établi la moyenne des évaluations faites par les étudiants et la fréquence des divers comportements dans plusieurs catégories. Nous avons trouvé que c'est le comportement des enseignants qui expliquait largement les réponses aux questionnaires remplis par les étudiants dans trois domaines principaux. (1) Lorsque les enseignants passaient leur temps à structurer les classes et à expliquer les rapports, les étudiants accordaient des notes élevées aux items portant sur l'organisation logique. (2) Lorsque les professeurs louangeaient les étudiants sur leur comportement, posaient des questions et élucidaient ou développaient les réponses des étudiants, les notes sur l'efficacité de la conduite des discussions étaient*

\* Centre for Teaching and Learning Services, McGill University

\*\* Instructional Communications Centre, McGill University

*plus élevées. (3) Lorsque l'enseignant passait son temps à discuter avec les étudiants, à faire l'éloge du comportement de ceux-ci et à marquer des silences (pour attendre les réponses), les étudiants tendaient à juger que l'atmosphère qui régnait en classe favorisait l'acquisition des connaissances.*

---

The University or college instructor who decides to improve his teaching almost inevitably uses student questionnaires to uncover the strengths and weaknesses in his classroom performance, then attempts to make changes in the weak or low-rated areas. Although extensive research has been done on the reliability and validity of student ratings (cf. Kulik & Kulik, 1974; Meredith, 1976), few attempts have been made to determine which teacher behaviors actually yield high student ratings. Consequently the instructor who receives low ratings in an area often does not know what changes to make in order to improve those ratings. Instructors who participated in a teaching improvement program will, at best, "feel more positive" about their teaching (Erickson & Erickson, 1979); they do not otherwise utilize student ratings to actually change their behaviors (Pambookian, 1972; Centra, 1973).

This study explores the relationship between student ratings of instruction and observed classroom behaviors for professors who are attempting to improve their teaching by using student evaluations. Once such relationships are established, the next step would be to determine whether changes in these classroom behaviors produce corresponding changes in student ratings and other student outcomes (Cranton, Note 1).

## INSTRUMENTS

The questionnaire used in this study was the Teaching Analysis by Students (TABS) questionnaire, an instrument designed to be an integral part of a teaching improvement process (cf. Anon, 1973). The basic questionnaire consists of 38 items, but additions and deletions may be made by the professor in consultation with the teaching improvement specialist. For a list of the actual items, the reader should consult Bergquist and Phillips (1975, pp. 80-82).

Observation data were collected using a category observation system developed by Shulman (Note 2) and based on Flanders Interaction Analysis (Flanders, 1970). In the system, an instructor's class is video-taped, with a digital clock providing a time reference on the tapes. A trained rater unaware of questionnaire results enters a category number on a score sheet every five seconds yielding a record of all classroom interactions. The fifteen categories of behavior are described in Table 1. Inter-rater reliability coefficients of .86 and .87 have been found; test-retest coefficients are reported to be .90, .88 and .94.

## PROCEDURE

Twenty-eight professors were involved in the study. All were interested in evaluating their teaching for the purpose of improvement and were working with a teaching improvement specialist to this end. Professors were from a variety of academic disciplines: law, engineering, management studies, library science, physics, biology, education, psychology, arts, anthropology and continuing education. Class sizes ranged from 20 to 100 students,

## 75 The Relationship between Student Ratings and Instructor Behavior:

Table 1  
Descriptions of Categories

1. Data Lecturing: giving facts or opinions about content; expressing one's own ideas; asking rhetorical questions; includes problem solving.
2. Data A.V.: presenting data with the aid of audio-visual materials. Includes using the blackboard.
3. Data Illustration: illustrating data with personal anecdotes, real case presentations and role playing.
4. Data Linking: in presenting data, using the specific skills of generalizing (relating content to other academic disciplines and identifying connections between concepts) or summarizing (reviewing data) or providing connections between student interest and the data.
5. Management: administrative tasks; statements or questions dealing with schedules, deadlines, reading lists, etc. Includes the act of handing out or collecting materials; giving quizzes or written exercises.
6. Structuring: contracting and organizing the class in regard to content and procedure. Includes briefly summarizing past material and activities, setting objectives, and giving commands and directions to be followed.
7. Silence: pauses, short periods of silence. Indicates confusion or laughter when scored simultaneously with another category.
8. Questions: asking a question about content with the intent that someone answer.

Table 1 (continued)

- |     |              |  |
|-----|--------------|--|
| 9.  | Discussion:  | encouraging or facilitating interaction and discussion between students. For example, asking class members to respond to a student's comment.  |
| 10. | Clarifying:  | statements and questions by the instructor designed to encourage a student to elaborate an idea or question initiated by the student. Includes paraphrasing which attempts to clarify another point of view. |
| 11. | Crediting:   | praising ideas, performance or work patterns.  |
| 12. | Criticizing: | direct or indirect criticizing, in a destructive manner, of ideas, performance or work patterns.   |
| 13. | Demand:      | making a demand for work. Includes constructive criticism and insisting on focus.  |
| 14. | Monitoring:  | calling attention to process in order to identify and explore blocks or potential blocks to effective classroom work. Includes periodically checking for attention, comprehension, etc.                      |
| 15. | Affect:      | clarifying the feeling of others in the classroom. Offering one's own feelings. Feelings may be positive or negative. Includes predicting or recalling feelings.   |

with an average size of about 50 students. Courses were both graduate and undergraduate (the majority being the latter), both half year and full year courses, and included a wide variety of teaching styles and methods.

During the fourth or fifth week of the semester, the TABS questionnaire was administered and the class video-taped. One trained rater viewed all tapes, and was unaware of questionnaire results.

The number of occurrences of each behavior category was determined for an individual professor. Data were then recorded in the form of percentages of the total time spent on each category. A professor would have 15 category "scores", each one being a percentage of time spent showing the behavior (lecturing, structuring, etc.).

Questionnaire data for each professor were recorded in terms of averaged student ratings on each of the 38 TABS items.

77 The Relationship between Student Ratings and Instructor Behavior:

Table 2  
 Predicted and Significant\* Relationships  
 Between Behaviors and Ratings

Item #	Item	Category #	Behavior	R <sup>2</sup>
1.	Explanation of objectives	5	Structuring	-
		6	Management	-
4.	Explanation of work expected	5	Structuring*	.21
		6	Management*	.24
5.	Relationship between content and objectives	4	Data Linking*	.14
		6	Structuring	-
6.	Relationships among topics	4	Data Linking*	.21
		6	Structuring	-
7.	Distinction between major and minor topics	4	Data Linking*	.23
		6	Structuring*	.25
8.	Pacing	14	Monitoring	-
9.	Ability to clarify material	2	Data A.V.	-
		3	Data Illustration	-
		4	Data Linking*	.17
		10	Clarifying*	.22
11.	Asking easily understood questions	7	Silence	-
		8	Questions	-
		10	Clarifying	-
12.	Asking thought-provoking questions	7	Silence	-
		8	Questions	-
		9	Clarifying	-
14.	Effectiveness as a discussion leader	8	Questions	-
		9	Discussion*	.20
		10	Clarifying	-
		11	Crediting*	.36
		13	Demand*	.29
		14	Monitoring	-
	15	Affect	-	
15.	Ability to get students to participate	8	Questions	-
		9	Discussion*	.22
		10	Clarifying	-
		11	Crediting*	.31
		13	Demand*	.33
		14	Monitoring	-
	15	Affect	-	
16.	Facilitating discussion among students	8	Questions	-
		9	Discussion*	.24
		10	Clarifying	-
		11	Crediting	-
		13	Demand	-
		14	Monitoring	-
	15	Affect*	.18	

Table 2 (continued)

Item #	Item	Category #	Behavior	R <sup>2</sup>
27.	Management of administrative details	5	Management	-
28.	Flexibility in offering options	10 15	Clarifying Affect*	- .16
29.	Taking action when students are bored	3 4 8 14 15	Data Illustration Data Linking Questions Monitoring Affect*	- - - - .16
32.	Atmosphere to encourage learning	2 4 7 9 11	Data A.V. Data Linking Silence* Discussion* Crediting*	- - .44 .26 .35
33.	Ability to inspire interest	2 4 7 9 11	Data A.V.* Data Linking* Silence Discussion Crediting*	.35 .23 - - .30
36.	Getting students to challenge	9	Discussion	-
37.	Relationship between personal values and course content	9	Discussion	-
38.	Making students aware of value issues	9	Discussion	-

The hypotheses were tested by a series of stepdown multiple regression analyses, with rating items used as the dependent variables and percentages of time spent showing the behaviors as the independent variables.

## RESULTS

In Table 2, the behaviors marked by an asterisk were those which accounted for a significant proportion of the variance in their prediction of the student questionnaire item.

Categories 5 and 6 (Structuring and Management) accounted for a significant amount of variance ( $R^2 = .24$ ) in their prediction of Questionnaire Item 4, (explanation of work expected). The professor who spent a higher proportion of time in structuring and management behaviors was rated higher by students on ability to explain work expected. However, this relationship did not appear for Item 1 (explanation of course objectives). It is likely that behavior related to this item took place earlier in the semester and was not observed in our videotapes.

Questionnaire items concerned with the clarification of relationships among topics (# 5, 6 and 7) were predicted by the Data Linking category ( $R^2 = .14, .21, .23$ ). The instructor who spends time generalizing, summarizing and providing connections, will tend to be rated higher on ability to clarify relationships.

No relationship was found between adjusting the rate of presentation (Item 8) and monitoring behavior (Category 14). Instructors who checked for attention, comprehension, etc. did not necessarily adjust their pacing, or students did not necessarily perceive them as adjusting their pacing.

Student ratings of ability to clarify material which needed elaboration (Item 9) were predicted by the Data Linking and Clarifying categories of behavior ( $R^2 = .22$ ). That is, when a higher percentage of time was spent generalizing, summarizing or providing connections, students perceived clarification skills more positively.

Questionnaire Items 11 and 12 (asking questions) were not related to the Questioning, Silence or Clarifying behavior categories. A professor spending a higher proportion of time in these activities did not necessarily receive a higher rating on his or her questioning skill.

Questionnaire items concerned with the instructor's effectiveness as a discussion leader were predicted by a number of behavior categories. Discussion, Crediting and Demand, came first into the regression equations for Items 14 and 15, accounting for 36% and 33% of the variance, respectively. Other significant predictors were Categories 8 and 10 (Questions and Clarifying) for Questionnaire Item 14 and Category 14 (Monitoring) for Item 15. Item 16, on the other hand, was concerned with facilitating discussion *among* students, and was best predicted by the Affect category ( $R^2 = .16$ ) followed by the Discussion category ( $R^2 = .24$ ).

Contrary to expectations, Item 27 on management of administrative detail was not related to time spent in Management activities. It is possible that students perceive higher percentages of time used for administrative tasks as an indication that the instructor is poorly organized and is wasting class time.

A professor who was perceived as being flexible in offering options for individual students (Item 28) was also one who was spending a higher proportion of time exhibiting behavior in the Affect category ( $R^2 = .16$ ).

Item 29, taking action when students are bored, was also predicted by the Affect category ( $R^2 = .16$ ) but not by the other expected categories (Data Illustration, Questions, Monitoring, etc.).

The Questionnaire items concerned with creating a learning atmosphere were related to several categories. For Item 32, the Discussion, Crediting and Silence categories were the first to enter the equation ( $R^2 = .44$ ); for Item 33, the Data Linking, Crediting and Data A.V. categories were the best predictors (35% of the variance was accounted for). In other words, students saw time spent on discussion as contributing to an atmosphere which encouraged learning. However, when rating the instructor's ability to inspire interest in the content of the course, more time spent on generalizing, summarizing, providing connections, praising students, and using audiovisual aids was related to higher questionnaire ratings.

Finally, items concerned with value issues, and getting students to challenge points of view were not predicted by the Discussion category: professors spending more time in discussion activities did not necessarily raise these issues.

## DISCUSSION

The instructor who is embarking on a program to improve his teaching often has very little information as to what changes in classroom behavior should be made in response to poorly rated questionnaire items. This research has attempted to establish some relation-

ships between behavior and ratings, to begin to answer the question, "what does the highly rated instructor actually do?"

For those teaching skills related to "structuring" (clarifying expectations and clarifying relationships) it was found that the proportion of time spent on the activities is relevant to the student ratings of ability in those areas. The relationship is straightforward: devoting time to the area is probably a good first step in improving. This also appears to be true for clarification of material: generalizing, summarizing, providing connections and paraphrasing are relevant to student ratings of the instructor's ability to clarify.

Other teaching skills are not so directly related to time spent on the apparently relevant behaviors. Monitoring student behavior, for example, does not necessarily result in the ability to adjust the rate of presentation. Spending time questioning students may not lead to students' satisfaction with questioning ability. Dedicating class time to administrative details does not seem to be related to high ratings of the instructor's management skills. In each of these areas, students probably use other indications of effectiveness. For example, perception of pacing ability may be based on the student's own understanding of material, and therefore would be influenced by student characteristics. It may be possible that the class means are not the appropriate unit of analysis for items where individual student differences are so relevant to the rating (Cranton, Note 3). The same issue arises for questioning ability: student judgements will be made individually as to whether they understand the questions, or find them thought-provoking, and the amount of time spent asking questions does not seem relevant. Further research is required before questionnaire items in these areas can become useful in improving instruction.

In yet another group of teaching skills, the rating-behavior relationships provide some interesting and useful information.

The effective discussion leader appears to be the professor who exhibits more crediting behavior (praising ideas, performance and work patterns), asks questions, and clarifies (in terms of encouraging a student to elaborate an idea or question). The professor who is rated highly on ability to get students to participate in discussion tends to show behavior in the Demand category (including constructive criticism and focus) and the Monitoring category (exploring blocks in classroom work, checking for attention, comprehension). Being an effective discussion leader tends to be related to not only the presence of discussion in the classroom, but also to providing feedback, learning what the students know, encouraging, focusing, and criticizing (constructively). The professor who facilitates discussion among students, as opposed to between the students and himself, has a higher proportion of time spent in the Affect category; clarifying feelings, offering feelings, predicting or recalling feelings. Also the professor who is perceived as flexible in offering options to individual students, and the professor who is seen to take appropriate action when students are bored, both spend a higher proportion of time expressing feelings.

The ability to encourage learning and to create interest in the course content is a priority of most instructors; students must be motivated before other goals can be achieved. A wide variety of behaviors predict these ratings, as would be expected. In creating an atmosphere that encourages learning, discussing, crediting, praising student ideas, and waiting for responses are relevant. Three out of the four Data categories are also related to ratings in this area. Although significant contributors, they only account for an additional 7% of the variance. In other words, when students feel that the classroom activities are conducive to learning, the emphasis is on interpersonal interactions. There is discussion; the instructor reinforces student participation, and waits for students to respond.

Inspiring interest or excitement is related to somewhat different types of behavior – discussion is not as relevant; data linking and crediting are the first contributors. In order to interest students, as most instructional design texts suggest, it is important to relate course concepts to other disciplines and to students' experiences. Crediting is also clearly relevant: students are encouraged to present their own ideas, and these ideas can then be related to the topic being discussed. Silence is again significant, adding 8% to the variance. The importance of silence is probably the importance of giving students time to respond.

## SUMMARY

The study was able to isolate some specific classroom behaviors that are related to student ratings of instruction. These relationships can provide guidance for the instructor who is using student ratings to improve instruction. Research should continue this type of investigation: a larger sample size would permit a more thorough analysis of relationships, follow-up studies (i.e., post-improvement) are needed, and other student outcomes (e.g., student behavior, student ratings of their own learning) should be included.

## REFERENCES

- Anon, Teaching Analysis by Students (TABS). Amherst, Mass.: Clinic to Improve University Teaching, University of Massachusetts, 1973.
- Bergquist, W., & Phillips, S. *A handbook for faculty development*. Washington: Council for the Advancement of Small Colleges, 1975.
- Centra, J.A. Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology*, 1973, 65, 395-401.
- Erickson, G.L., & Erickson, B.L. Improving college teaching: An evaluation of a teaching consultation procedure. *The Journal of Higher Education*, 5 (5), 670-683.
- Flanders, N.A. *Analyzing teaching behaviour*. Reading, Mass: Addison Wesley, 1970.
- Kulik, J.A., & Kulik, C.C. Student ratings of instruction. *Teaching of Psychology*, 1974, 1, 51-57.
- Meredith, I.M. Structure of the teaching analysis by students. *The Journal of Psychology*, 1976, 92, 235-242.
- Pambookian, H.A. Initial level of student evaluation of instruction as a source of influence on instructor change after feedback, *Journal of Educational Psychology*, 1974, 66, 52-56.

## Reference Notes

1. Cranton, P.A. *The effect of evaluation for improvement on teaching performance and student outcomes*. Montreal: McGill University, Centre for Teaching and Learning Services, 1980.
2. Shulman, L. *Development of a category observation system for the analysis of video-taped class sessions*. Montreal: McGill University, Centre for Learning and Development, 1975.
3. Cranton, P.A. *The evaluation and interpretation of student ratings of instruction*. Montreal: McGill University, Centre for Teaching and Learning Services, 1979.