

# Low-Inference Classroom Teaching Behaviors and Student Ratings of College Teaching Effectiveness

Harry G. Murray

University of Western Ontario, London, Ontario, Canada

Trained observers visited classes taught by university lecturers receiving either low, medium, or high student ratings. The observers estimated the frequency of occurrence of 60 specific, low-inference teaching behaviors. Significant differences among low-, medium-, and high-rated lecturers were found for 26 individual behaviors divided among seven categories of teaching. Group differences were largest for "attention getting" behaviors such as speaking expressively, moving about while lecturing, using humor, and showing enthusiasm for the subject. Factor analysis of individual teaching behaviors yielded nine interpretable factors, of which three (Clarity, Enthusiasm, and Rapport) differed significantly across groups, and all but one showed correlations with various teacher and course characteristics. Results are discussed with reference to the pivotal role of attention-getting behavior in classroom teaching, the validity of student instructional ratings, and the design of teaching improvement programs in higher education.

Research over the last 20 years indicates that student ratings can provide reliable and valid information on the quality of college teaching. Although results are sometimes contradictory, the weight of evidence suggests that student ratings of a given instructor are reasonably stable across courses and time periods; are affected to only a minor extent by extraneous factors such as class size and severity of grading; are consistent with similar ratings made by alumni, colleagues, and trained classroom observers; and most important of all, are significantly correlated with more objective measures of teaching effectiveness such as student exam performance (McKeachie, 1979; Murray, 1980).

Although considerable research has been done on the reliability and validity of student ratings, much less is known as to the specific things that teachers receiving high or low ratings actually *do* in the college

classroom. Factor analysis of student instructional ratings represents one approach to this question. Factor analytic studies (e.g., Marsh, 1982) have identified a number of replicable dimensions underlying student perceptions of effective teaching, including instructor clarity/organization, enthusiasm, breadth of coverage, and feedback/grading. However, the question remains as to whether these global factor-analytic dimensions can be understood in terms of what Rosenshine and Furst (1971) have referred to as "low inference" teaching behaviors, or in other words, specific classroom behaviors of the instructor which can be recorded by direct observation. For example, can student ratings of "organization" be predicted from specific classroom behaviors such as "signals transition to new topic," or "puts outline of lecture on blackboard"? Knowledge of specific teaching behaviors that contribute to positive or negative evaluations from students would be helpful in improving the diagnostic and remedial value of student instructional ratings. As things stand now, an instructor who receives poor ratings on global dimensions such as "clarity" or "overall effectiveness" usually has no idea of the specific behaviors that led to these ratings, or the specific changes that would yield improvement on these dimensions.

The present study compared the fre-

---

This research was supported by grants from Imperial Oil Limited of Canada and the Ontario Universities Program for Instructional Development.

The author gratefully acknowledges the cooperation of teachers and classroom observers who participated in this study.

Requests for reprints should be sent to Harry G. Murray, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

quency of occurrence of 60 low-inference teaching behaviors in groups of university lecturers who had consistently received either low, medium, or high "overall effectiveness" ratings from students in previous courses. Previous studies of the relationship between low-inference teaching behaviors and student instructional ratings have typically used the same group of students both as observers of classroom behavior and as raters of teaching effectiveness, thus leaving open the possibility that correlations obtained were due to judgement biases such as "halo effect" or "implicit personality theory" (e.g., Mintzes, 1979; Solomon, 1966; Tom & Cushman, 1975). Other studies have relied solely on videotaping, a potentially very obtrusive measure, in recording classroom behavior (e.g., Cranton & Hillgartner, 1981; Hiller, Fisher, & Kaess, 1969). In the present study, perceived teaching effectiveness was measured in terms of student ratings from previous courses, whereas low-inference teaching behaviors were independently recorded by trained observers who unobtrusively visited regular classes taught by participating lecturers. It was assumed that this procedure would minimize judgement bias while maintaining conditions of unobtrusive observation.

## Method

### Subjects

The sample of teachers consisted of 54 full-time faculty members in the Faculty of Social Science, University of Western Ontario. The breakdown of the sample in terms of sex and academic rank was as follows: 46 male, 8 female; 25 assistant professors, 20 associate professors, and 9 full professors. Each of the participating instructors had been at the university for at least three years, had taught and received student ratings in at least five previous undergraduate courses, and was scheduled to teach a lecture or lecture/discussion course with an enrollment of at least 30 students during the 1977-78 academic year.

Low-, medium-, and high-rated groups of teachers were selected on the basis of previous ratings on the overall effectiveness item of the Faculty of Social Science questionnaire for student evaluation of teaching. Eighteen of the 54 teachers had consistently received high ratings on this item (4.20 or higher on a 5-point scale;  $M = 4.38$ ,  $SD = .18$ ), whereas 18 had consistently received medium ratings (3.20 to 3.80;  $M = 3.49$ ,  $SD = .12$ ); and 18 had consistently received low ratings on the same item (2.90 or lower;  $M = 2.67$ ,  $SD = .29$ ). For the Faculty of Social Science as a whole, the

correlation between instructor mean ratings in successive years averaged .76 during the 4-year period preceding this study, indicating that assignment of instructors to groups was based on stable, replicable data. Each group of teachers included at least two members from each of the following departments: Economics, Geography, History, Political Science, Psychology, and Sociology. Low-, medium-, and high-rated groups were also approximately matched (i.e., did not differ significantly) in terms of teacher age, sex, and academic rank and in terms of size, level, time of day, and duration of the class in which observations were made.

### Procedure

Each of the 54 teachers was observed in three separate 1-hour class periods by each of six to eight trained classroom observers. Thus each teacher was observed for a total of 18-24 hours over a period of approximately 3 months. The classroom observers were 49 students in an Educational Psychology course who participated in the study to fulfill a practicum requirement. Each observer was assigned a sample of eight teachers, with instructions to unobtrusively visit regular classes taught by these individuals and record low-inference classroom behaviors on a standardized behavioral observation form described below. Prior to visiting classes, observers were given approximately 4 hours of group training in recording classroom behaviors from videotaped lecture segments. The assignment of observers to teachers was random with the following restrictions: (a) each observer was assigned at least one low-, one medium-, and one high-rated teacher; (b) the observer had taken an introductory course in the teacher's general subject area (e.g., economics), but had never taken a course from that specific teacher; and (c) the class to be observed fit the observer's academic timetable. Observers were unaware that the study involved comparison of low-, medium-, and high-rated teachers, and were given no information as to prior student ratings of the teachers they observed. The 54 teachers who participated in the study were similarly unaware of the overall research design, although each had given explicit permission for outside observers to visit classes at unannounced times during the academic term.

### Behavioral Observation Instrument

Observers summarized their three hours of observation of each teacher on a standardized rating form called the Teacher Behaviors Inventory (TBI). The TBI consists of 60 items divided among the following eight categories of classroom behavior: Speech, Nonverbal Behavior, Explanation, Organization, Interest, Task Orientation, Rapport, and Participation. TBI items were derived from previous college-level classroom observation-instruments (e.g., Tom & Cushman, 1975); research on lecturing and explaining (e.g., Gage & Berliner, 1979, Chap. 20); and informal discussions with students and faculty members. As may be seen in Table 1 below, each TBI item refers to a specific, observable behavior, such as "gestures with hands and arms," "writes key terms on blackboard," and "addresses students by name." Observers rated the frequency of occurrence of each of these behaviors on a

Table 1

*Frequency of Occurrence of Specific Classroom Behaviors in Low-, Medium, and High-Rated Groups of Teachers*

Teaching behavior	Mean frequency rating			<i>MS<sub>e</sub></i>	Univariate <i>F</i>
	Low group	Medium group	High group		
Speech					
Speaks slowly	1.94	1.38	1.24	.43	5.47*
Speaks in monotone	2.09	1.50	1.49	.55	3.36*
Stutters, mumbles, or slurs words	1.64	1.51	1.49	.68	0.15
Speaks expressively or emphatically	1.98	2.79	3.81	1.21	11.18*
Speaks softly	1.66	1.35	1.34	.65	.84
Speaks loudly	1.70	1.33	1.85	.55	2.82
Speaks clearly	3.94	3.63	4.14	.37	2.82
Nonverbal behavior					
Moves about while lecturing	2.68	3.46	4.15	2.09	4.03*
Gestures with hands and arms	3.03	3.35	3.95	1.01	3.31*
Avoids eye contact with students	2.24	1.95	1.56	.36	5.18*
Shows distracting mannerisms	1.65	1.23	1.23	.31	3.11
Smiles or laughs	2.20	2.34	2.54	1.39	.33
Shows energy and excitement	2.42	2.46	3.26	.81	4.40*
Relaxed and confident	4.00	4.54	4.65	.56	3.46*
Shows facial expressions	1.86	2.08	3.21	1.12	7.50*
Explanation					
Uses concrete examples	4.28	4.58	4.73	.61	1.39
Repeats difficult ideas	3.46	4.26	4.10	.99	2.89
Uses graphs and diagrams	3.19	3.73	4.26	1.35	3.82*
Stresses important points	3.89	4.71	4.85	.38	11.43*
Suggests mnemonic aids	1.53	1.51	2.14	.59	3.43*
Writes key terms on board	2.56	2.41	3.18	1.89	1.38
Uses audiovisual aids	2.26	2.28	2.16	2.32	.03
Gives multiple examples	3.73	4.23	4.59	.86	3.47*
Organization					
Uses headings and subheadings	3.43	3.50	3.43	1.46	.02
Explains how each topic fits in	2.73	3.01	3.23	1.26	.80
Puts outline of lecture on board	1.60	1.88	1.56	.79	.59
Signals transition to new topic	2.80	2.89	3.34	.21	4.01*
Summarizes periodically	3.89	4.24	4.21	.67	.99
Gives preliminary overview of lecture	3.25	3.66	3.84	.30	5.50*
Interest					
States own viewpoint on issues	3.27	3.45	3.74	1.39	.63
Uses humor	2.90	3.50	4.36	1.60	5.40*
Suggests practical applications	3.62	4.16	4.45	.75	3.79*
Shows strong interest in subject	3.04	3.99	4.74	.74	15.69*
Uses variety of media	1.49	1.76	1.85	.93	.62
Reads lecture verbatim from notes	2.49	2.36	1.41	1.11	4.98*
Presents thought-provoking ideas	2.30	3.30	3.59	1.38	5.29*
Relates subject to current events	3.71	3.74	4.20	1.01	1.20
Task orientation					
States teaching objectives	2.46	2.70	2.68	1.12	.24
Advises students re tests	2.43	2.46	2.55	.91	.07
Provides sample exam questions	1.21	1.50	1.76	.80	1.51
Digresses from topic of lectures	1.55	1.38	1.68	1.11	.42
Dwells on obvious points	1.77	1.38	1.25	.62	1.96
Proceeds at rapid pace	2.10	2.73	2.75	1.01	2.14
Sticks to point in answering	4.58	4.70	4.60	.29	.21

Table 1 (continued)

Teaching behavior	Mean frequency rating			<i>MS<sub>e</sub></i>	Univariate <i>F</i>
	Low group	Medium group	High group		
Rapport					
Shows concern for students	3.65	4.48	4.51	.79	4.80*
Tolerant of other viewpoints	4.66	4.84	4.75	.22	.57
Addresses students by name	1.57	1.80	2.41	.80	4.25*
Offers help with problems	2.68	3.41	3.91	1.51	4.02*
Talks with students after class	1.69	1.66	2.04	1.15	.61
Sensitive to student needs	4.52	4.75	4.83	.23	1.78
Friendly, easy to talk to	4.10	4.70	4.66	.56	3.23*
Participation					
Encourages questions and comments	2.96	3.78	3.78	.90	3.96*
Asks questions of individual students	1.56	1.43	1.80	.80	.70
Asks questions of class as whole	3.06	3.70	4.19	1.26	4.02*
Praises students for good ideas	2.83	3.28	3.49	1.29	1.42
Fails to take initiative in class	2.74	2.13	1.99	1.28	1.98
Provides opportunity for participation	2.26	2.10	2.54	2.59	.30

\*  $p < .05$ .

5-point scale labeled as follows: 1 = almost never, 2 = rarely, 3 = sometimes, 4 = often, 5 = almost always.

## Results

### Reliability of Behavioral Observations

Interrater reliability coefficients were computed for each of the 60 TBI items by sorting observers assigned to each teacher into even-numbered and odd-numbered subgroups, correlating mean ratings for these two groups across teachers and applying the Spearman-Brown formula to estimate the reliability of the mean rating for all observers combined. Reliability coefficients computed in this way ranged from .24 to .97 for individual TBI items, with a median value of .76. The estimated median reliability for a single classroom observer was .32, whereas the projected median reliability for an increased complement of 15 observers per teacher was .88. These figures compare favorably to Centra's (1975) corresponding estimates of .27 and .85 for student ratings, and .08 and .57 for colleague ratings of overall teaching effectiveness based on classroom observation.

Three of the 60 TBI items were judged to have reliability coefficients too low to justify their inclusion in subsequent statistical analyses: "criticizes students for errors"

(.24); "speaks rapidly" (.31); and "uses big words" (.40). The remaining 57 items had reliabilities ranging from .51 to .97, with a median reliability of .77. Ratings of these items were averaged across observers for each teacher to obtain mean ratings of 57 behaviors for each of 54 teachers.

### Analysis of Individual Teaching Behaviors

Table 1 shows the mean estimated frequency of occurrence of each of the 57 reliably judged teaching behaviors for low-, medium-, and high-rated groups of teachers. A preliminary multiple discriminant analysis indicated that low-, medium-, and high-rated groups differed significantly with respect to mean scores on all 57 teaching behaviors taken together, Wilks's lambda = .0000287,  $p < .01$ . Follow-up univariate analyses of variance showed that 26 of the 57 teaching behaviors differed significantly among low-, medium-, and high-rated groups. Comparison of the mean interrater reliability of behaviors that did ( $M = .76$ ) and did not ( $M = .71$ ) differentiate among groups yielded a nonsignificant difference,  $t(55) = 1.16$ ,  $p > .05$ , indicating that reliability of measurement was not of critical importance in determining which behaviors differentiated significantly among groups.

It may be noted in Table 1 that the teaching behaviors that showed significant univariate  $F$ s ranged widely in content, with all but one of the eight TBI categories including at least one significant behavior. However, group differences tended to be larger and more frequent for Nonverbal Behavior, Interest, Explanation, and Rapport categories than for remaining categories. The five teaching behaviors showing the largest raw-score differences among low, medium, and high groups were (a) speaks expressively or emphatically, (b) shows strong interest in subject, (c) moves about while lecturing, (d) uses humor, and (e) shows facial expressions. These behaviors seem to share elements of stimulus variation, expressiveness, and spontaneity. Thus, they are perhaps best interpreted as ways of communicating enthusiasm for the subject matter and thereby eliciting and maintaining student attention to lecture material. The fact that these behaviors showed the largest group differences suggests that the use of effective methods of engaging and holding student attention is a very important factor in differentiating between successful and less successful college lecturers. On the other hand, attention-getting behavior was by no means the only factor differentiating among low-, medium-, and high-rated teachers. Significant group differences were also found for more traditional "cognitive" or information-giving behaviors such as using graphs and diagrams, stressing important points, giving multiple examples, and signaling the transition to a new topic and for behaviors aimed at establishing rapport or encouraging student participation, such as asking questions of students, addressing students by name, and offering help with problems. Perhaps the reason that group differences tended to be larger for attention-getting behaviors than for information-giving or rapport-establishing behaviors is that the effectiveness of the latter behaviors depends on whether or not students are already paying attention to the lecturer's presentation.

Behaviors whose failure to show significant group differences in this study might be seen as surprising in view of the emphasis they have received in the literature include

stating teaching objectives and using audiovisual aids. It should be emphasized, however, that behaviors that showed small or nonsignificant group differences should not necessarily be interpreted as "unimportant" behaviors but only as behaviors that do not contribute significantly to differentiating between highly rated and less highly rated teachers. In other words, it may be that behaviors such as stating objectives and using audiovisual aids are necessary but not sufficient conditions for perceived excellence in teaching.

Further analyses were undertaken to determine the extent to which ratings of overall teaching effectiveness could be predicted from a limited subset of individual teaching behaviors. To minimize the possibility of spurious results due to post hoc selection of predictor variables (Cohen & Cohen, 1975), a new random sample of 10 low-inference teaching behaviors was selected for use in each of three separate multiple regression analyses, with simultaneous inclusion of predictors in each case. The criterion variable was mean instructor rating in all previous courses. All three analyses yielded statistically significant multiple regression coefficients,  $R(10, 43) = .814, .686, \text{ and } .733$ , respectively,  $p < .01$ , which with correction for shrinkage correspond to variance accounted for estimates of .594, .363, and .443, respectively. These results indicate that student instructional ratings can be predicted with considerable accuracy from frequency estimates of as few as 10 randomly selected low-inference classroom behaviors. In other words, it would appear that student ratings are "valid" in the sense that they are determined primarily by actual classroom behaviors of the instructor rather than by extraneous factors such as popularity or leniency.

### *Factor Analysis*

For purpose of (a) investigating the factor structure of the TBI and (b) deriving a smaller set of uncorrelated teacher behavior variables for use in further analyses, teachers' scores on the 57 reliably judged TBI items were subjected to a principal-axis, varimax-rotation factor analysis. To

achieve a larger number of "subjects" than "variables" (and thus assure a nonsingular covariance matrix), the factor analysis was performed on raw ratings of individual classroom observers ( $N = 392$ ) rather than teacher mean ratings ( $N = 54$ ). Factor loadings from this analysis were then used in the computation of factor scores for the 54 teachers. In defense of this procedure, recent evidence indicates that teacher behavior ratings show essentially the same factor structure regardless of whether observer ratings or teacher mean ratings serve as the unit of analysis (e.g., Linn, Centra, & Tucker, 1975).

The factor analysis yielded nine readily interpretable factors with eigenvalues of 2.00 or higher, which accounted for 69.1% of the total variance in TBI ratings, plus four smaller factors with uninterpretable or single-item loading patterns. Table 2 lists teaching behaviors loading .50 or higher on each of the nine major factors, plus variance accounted for and alpha reliability estimates for each factor. Of the 57 behaviors, 47 loaded on one factor only, whereas 5 loaded on two different factors, and 5 loaded on none of the major factors. It may be noted that the obtained factor structure did not correspond in any obvious way to the logically defined categories of the TBI. For example, Factor 1, which seems to reflect a "clarity of explanation" or "expositional skill" dimension, includes items from four separate categories of the TBI. Factor 2 consists of behaviors relating to expressiveness, stimulus variation, humor, spontaneity, and eye contact, which collectively seem to represent an "enthusiasm" or "attention getting" dimension. Factor 3 consists of behaviors which tend to facilitate student participation and student-teacher interaction. Factor 4 includes items from the Task Orientation and Organization categories of the TBI, and seems to reflect a businesslike, traditionalist, or "directive" approach to teaching. Factor 5, on the other hand, corresponds fairly closely to the Rapport category of the TBI, and reflects a friendly, sympathetic, concerned attitude toward students. Factors 6 through 9, accounting for a total of 16.1% of TBI variance, consist of behaviors relating to organization, use of

media, pacing of lecture presentations, and clarity of speech respectively.

### *Analysis of Factor Scores*

Teachers were assigned factor scores on each of the nine major factors by averaging mean frequency ratings on all TBI items loading .50 or higher on a given factor. Negatively loading items were coded in reverse direction and items loading on more than one factor were used only in relation to the highest-loading factor. Unstandardized factor scores computed in this way showed intercorrelations ranging from  $-.28$  to  $.39$ , despite the orthogonality of underlying factors. However, results reported below were essentially unchanged when factor scores were computed more conventionally, with differential weighting of items according to loading and standardization of composite scores.

Mean factor scores for low-, medium-, and high-rated groups of teachers were evaluated by one-way analyses of variance and follow-up Tukey HSD tests separately comparing low versus medium and medium versus high groups. Three of the nine factors, namely Clarity, Enthusiasm, and Rapport, varied significantly across low, medium, and high groups,  $F(2, 51) = 5.61, 16.09, \text{ and } 4.53$ , respectively,  $p < .01$  in each case, with Enthusiasm showing by far the largest group differences. Figure 1 shows group mean scores for the three significant factors. It may be noted that all three factors increased from low to medium to high groups. Of greater interest, Tukey tests showed that the difference between low and medium groups was the only significant pairwise contrast for the Clarity and Rapport factors, whereas the difference between medium and high groups was the only significant source of variance for the Enthusiasm factor. This pattern of results suggests that whereas Clarity and Rapport are important factors in differentiating a medium-rated or "average" lecturer from a low-rated or "poor" lecturer, the critical factor in differentiating an "outstanding" lecturer from an "average" lecturer is Enthusiasm, which has been interpreted here as the use of stimulus variation and expres-

Table 2  
Results of Factor Analysis

Factor	Item loadings	% variance	Alpha reliability	Factor	Item loadings	% variance	Alpha reliability
<i>Factor 1: Clarity</i>				<i>Factor 4: Task Orientation</i>			
Uses concrete examples	.84	20.6	.96	Digresses from topic of lecture	-.87	5.9	.90
Gives multiple examples	.80			Proceeds at rapid pace	.73		
Repeats difficult ideas	.75			Dwells on obvious points	-.67		
Asks questions of class as whole	.68			Signals transition to new topic	.65		
Suggests practical applications	.66			Advises students re tests	.59		
Shows concern for students	.65			Uses headings and subheadings	.51		
Uses graphs and diagrams	.60			<i>Factor 5: Rapport</i>			
Stresses important points	.55			Shows distracting mannerisms	-.74	5.0	.81
Suggests mnemonic aids	.54			Sensitive to student needs	.67		
Shows strong interest in subject	.52			Tolerant of other viewpoints	.62		
Fails to take initiative in class	-.50			Friendly, easy to talk to	.61		
<i>Factor 2: Enthusiasm</i>				Sticks to point in answering	.61		
Uses humor	.78	12.3	.93	Shows concern for students	.56		
Speaks expressively or emphatically	.72			<i>Factor 6: Organization</i>			
Shows facial expressions	.72			Gives preliminary overview of lecture	.84	4.6	.86
Moves about while lecturing	.67			Explains how each topic fits in	.71		
Reads lecture verbatim from notes	-.66			States teaching objectives	.61		
Shows energy and excitement	.64			Puts outline of lecture on board	.50		
Smiles or laughs	.64			<i>Factor 7: Use of Media</i>			
Gestures with hands and arms	.62			Uses audiovisual aids	.82	4.3	.77
Shows strong interest in subject	-.58			Suggests practical applications	.59		
Avoids eye contact with students	-.55			Uses variety of media	.57		
Speaks softly	-.51			Relates subject to current events	.56		
<i>Factor 3: Interaction</i>				<i>Factor 8: Pacing</i>			
Asks questions of individual students	.83	9.2	.95	Proceeds at rapid pace	-.78	3.8	.79
Addresses students by name	.77			Writes key terms on board	.60		
Provides opportunity for participation	.76			Summarizes periodically	.57		
Encourages questions and comments	.70			<i>Factor 9: Speech</i>			
Praises students for good ideas	.64			Stutters, mumbles, slurs words	-.80	3.4	.81
Presents thought-provoking ideas	.61			Speaks clearly	.57		
Talks with students after class	.55			Speaks loudly	.54		
Asks questions of class as whole	.55						
Speaks in monotone	-.52						

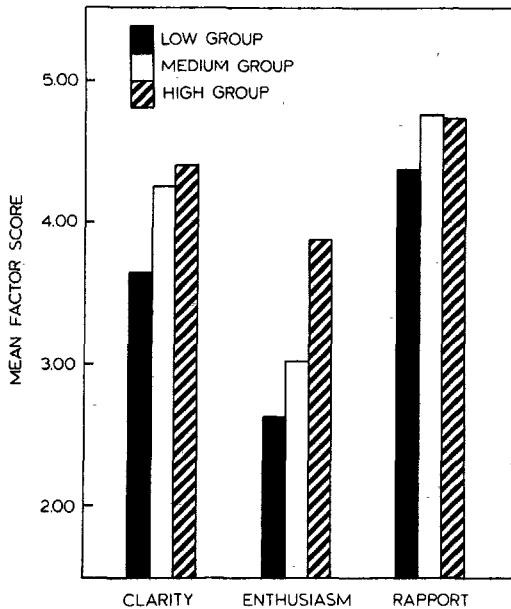


Figure 1. Mean scores of low-, medium-, and high-rated groups on Clarity, Enthusiasm, and Rapport factors.

sive behavior to elicit student attention and interest.

Figure 2 presents a scatterplot of factor scores of individual teachers on the two factors showing largest between-group variation, namely Clarity and Enthusiasm. The overlapping elliptically shaped perimeters indicate areas containing at least two-thirds of the members of low-, medium-, and high-rated groups respectively. Group centroids are indicated by asterisks. It may be noted that low-rated teachers ranged widely in Clarity scores, but were generally low on the Enthusiasm dimension. Consistent with Tukey tests reported above, medium-rated teachers were equal to high-rated teachers in Clarity, but not concomitantly high in Enthusiasm. High-rated teachers, on the other hand, tended to score high on both Clarity and Enthusiasm dimensions, with 72% of this group falling in the upper-right quadrant of the scatterplot. Although most teachers conformed to the above generalizations, there were some noteworthy exceptions or "deviant cases" in which individual teachers resembled members of other groups more than members of their own group. For example, one teacher

in the low-rated group had Clarity and Enthusiasm scores that fell close to the centroid of the high-rated group; and conversely, one member of the high-rated group scored near the centroid of the low-rated group. Presumably factors other than Clarity and Enthusiasm were responsible for these anomalies.

A final point worth noting in relation to Figure 2 is the apparent curvilinear regression of Enthusiasm scores on Clarity scores. An orthogonal polynomials analysis of these data, with Category scores grouped into quartile levels, yielded a statistically significant quadratic component,  $F(1, 50) = 4.48$ ,  $p < .05$ . Further evidence of nonlinear regression is seen in the fact that virtually none of the teachers depicted in Figure 2 showed the combination of low Clarity with high Enthusiasm (upper left quadrant), whereas the reverse combination of high Clarity and low Enthusiasm (lower right quadrant) occurred disproportionately often. This result suggests (but of course does not prove) that Clarity is a causal prerequisite of Enthusiasm, or more accurately, that a teacher must operate at a certain level of perceived clarity in order to be perceived as enthusiastic. Further evidence that the causal ordering of these two variables is Clarity  $\rightarrow$  Enthusiasm rather than Enthusiasm  $\rightarrow$  Clarity comes from the fact that the correlation between Clarity scores and overall effectiveness rat-

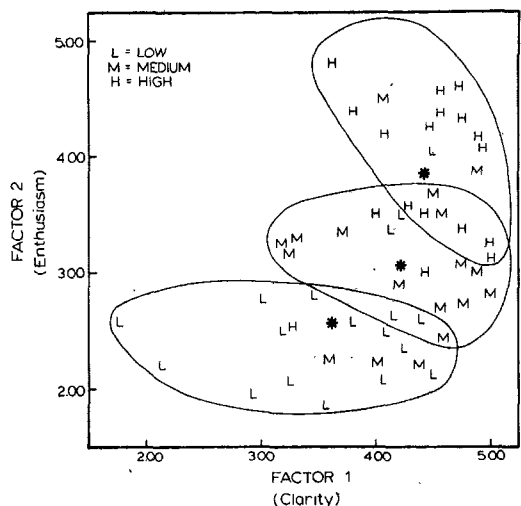


Figure 2. Scatterplot of individual teacher scores on Clarity and Enthusiasm factors.



ings dropped from .37 to .16 when Enthusiasm was statistically controlled (through semipartial correlation), whereas the correlation between Enthusiasm scores and overall effectiveness ratings was essentially unchanged (.64 vs. .62) when Clarity was statistically controlled.

Further analyses were undertaken to determine whether the nine teacher behavior factors were systematically related to any of the following teacher or course variables: teacher sex, age, academic rank, and department; and class size, level, time of day, and duration. Strong correlations between these two sets of variables were not expected, because low-, medium-, and high-rated groups had been roughly matched in terms of teacher and course variables, and group membership was predictable from factor scores. However, moderate levels of association were considered possible in view of the fact that matching of groups was approximate rather than exact, and factor scores were imperfectly related to group membership. Consistent with this expectation, a preliminary canonical correlation analysis (with dummy coding of categorical variables) yielded two significant canonical functions linking factor scores as a set to teacher and course characteristics as a set,  $R = .924$  and  $.864$ , respectively. The corresponding canonical redundancy index, which is less prone to inflation with modest degrees of freedom, was .248, indicating that approximately 25% of factor score variance was attributable, via the two canonical functions, to teacher and course characteristics (Cohen & Cohen, 1975). In that three of the eight teacher and course variables were non-quantitative, and furthermore, regression plots were visibly nonlinear in some cases, it was decided to test individual relationships between teacher/course characteristics and factor scores by one-way analyses of variance, in which quantitative teacher and course variables were blocked into quartile levels. Fifteen of 72 individual relationships between factor scores and teacher/course characteristics were found to be statistically significant, with all factors except Pacing showing at least one significant correlation. For example, teacher behaviors aimed at enhancing student-teacher interaction (Factor 3) were found to be more common in

smaller classes than in larger classes, more common in senior classes than in freshman classes, and more common in younger teachers than in older teachers. Use of media (Factor 7) was more frequent in older than in younger teachers, and more frequent in Geography than in any other department. Teacher clarity (Factor 1) was higher for teachers aged 32 to 40 than for younger or older teachers, and higher in first year than in second or third year classes. The existence of these and other relationships between factor scores and teacher/course characteristics further supports the validity of the TBI as a measure of classroom teaching and points to possible directions for future research on teacher behavior.

### Discussion

In summary, the major result of this study was that university lecturers receiving low, medium, and high ratings from students differed significantly in the frequency with which they exhibited various specific, low-inference teaching behaviors. Significant differences among low-, medium-, and high-rated groups of lecturers were found for 26 individual behaviors divided among seven different categories of teaching. Factor analysis of individual teaching behaviors yielded nine interpretable factors, of which three (Clarity, Enthusiasm, and Rapport) differed significantly across groups, and all but one showed significant correlations with various teacher and course characteristics.

The fact that student ratings of teaching could be predicted (or more precisely, postdicted) with considerable accuracy from neutral observers' reports of low-inference classroom teaching behaviors provides indirect support for the validity of student ratings. It appears that instructors who receive high ratings from students do in fact teach differently than instructors receiving average or poor ratings. In other words, student ratings appear to be determined more by actual classroom behaviors of the instructor than by extraneous factors such as "personality" or "popularity."

Although the classroom behaviors identified as statistically significant in this study varied widely in content, differences among low-, medium-, and high-rated groups were

largest for "expressive" behaviors such as vocal variation, movement and gesture, use of humor, and eye contact. It is interesting to note that studies of the so-called "Dr. Fox effect" have reported a similar prepotent effect of lecturer expressiveness on student ratings of videotaped lectures (Ware & Williams, 1980). The position taken here is that expressive teaching behaviors serve to communicate the lecturer's enthusiasm for the subject matter, and thereby elicit and maintain student attention to lecture material. In addition to having important direct effects on student attention and interest, expressive behaviors are assumed to play a crucial role in determining the impact of other categories of teacher behavior. For example, behaviors aimed at explaining concepts, enhancing recall, or stimulating independent thinking can be expected to be successful only to the extent that student attention has already been effectively engaged. Consistent with this interpretation, the Dr. Fox studies have usually reported that lecturer expressiveness enhances student retention of lecture material as well as student ratings (Ware & Williams, 1980).

Despite procedural differences, there are clearcut consistencies between the present findings and those of previous observational studies of low-inference college teaching behavior. Mintzes (1979) found significant correlations between student ratings of psychology instructors and student reports of several of the same behaviors identified in the present study—for example, speaking expressively, giving a preliminary overview, using multiple examples, and addressing students by name. Similarly, Tom and Cushman (1975) reported significant correlations between student self-ratings of amount learned in agriculture courses and student reports of 28 different low-inference teaching behaviors, including vocal expressiveness, signaling the transition to a new topic, and use of real-life examples. Tom and Cushman found nonsignificant correlations for movement and gesture and giving a preliminary overview, both of which were significant in the present study. Finally, Cranton and Hillgartner (1981) reported a variety of relationships between student instructional ratings and behavioral frequency counts derived by videotape

analysis, including a tendency for teachers who asked questions, praised students, and elaborated student responses to be highly rated as discussion leaders. Further research is needed to determine the extent to which low-inference teaching behaviors are differentially effective for different subject areas, teaching methods, and types of students. It may be that there are no more than a few categories of teaching behavior (enthusiasm being one possibility) which are effective for all subject areas, all teaching methods, and all types of students.

A vast literature exists on the relationship between classroom teaching behavior and student achievement in elementary and secondary schools. Early research in this area, reviewed by Rosenshine and Furst (1971), yielded rather inconsistent and confusing results, although the two dimensions of teacher behavior identified by Rosenshine and Furst as correlating highest with student achievement, namely "clarity" and "enthusiasm," are similar to the two factors that correlated highest with student ratings in the present study. More recent research, reviewed by Rosenshine (1979), is consistent in showing a positive association between businesslike, task-oriented behaviors such as stating objectives, brisk pacing, and use of factual questions, and student achievement in basic skill subjects such as reading and arithmetic. There appears to be a parallel of sorts between this "direct instruction" pattern of teaching behavior and the Task Orientation and Organization factors identified in the present study. Unfortunately, neither of the latter factors correlated significantly with student ratings of teaching.

Although most of the behaviors that differentiated among low-, medium-, and high-rated teachers were ones that would logically be expected to facilitate student learning or cognitive growth, the possibility exists that these behaviors are related only to student ratings and not to student learning. This question can be resolved only through further research. It should be noted, however, that some of the behaviors in question have already been shown, under experimental or quasi-experimental conditions, to be causally related to student learning. For example, as stated previously,

research on the Dr. Fox effect suggests that lecture material presented in a "dynamic" fashion, involving humor, expressive speech, and movement and gesture, is recalled significantly better than the same material presented in a "static" fashion (Ware & Williams, 1980). In a similar vein, Bertou, Clasen, and Lambert (1972) found that asking content-related questions during a videotaped lecture improved student retention of lecture material; and Smith (1977) showed that teachers who asked questions and encouraged student participation were more successful in fostering students' critical thinking skills than teachers who did not exhibit these behaviors. In view of these results, it seems likely that there is at least some degree of overlap between teaching behaviors that affect student ratings and teaching behaviors that affect student achievement.

It might be useful at this point to review some of the limitations of the present study. Two such limitations have already been noted: (a) the possibility that the present findings may not generalize to different subject areas, teaching methods, or types of students; and (b) the possibility that teaching behaviors identified in this study affect student ratings but not student learning. A third limitation is that the present research pertains more to the "delivery" aspect of instruction than to the "content" aspect. In other words, a teacher might exhibit all of the statistically significant behaviors identified in this study, and yet, because of inadequacies in subject knowledge, content coverage, or academic standards, be judged as an unsatisfactory or incompetent teacher. A fourth limitation of the present research is that the differences found among low-, medium-, and high-rated teachers are probabilistic rather than categorical in nature. Thus it is possible, although unlikely, that a teacher who shows few or none of the behaviors associated with the high-rated group might nonetheless receive high ratings from students. The "deviant cases" seen in Figure 2 remind us that there are limits on the predictability of overall teaching effectiveness from specific behavioral characteristics. A fifth and final limitation of the present research is that its retrospective/correlational design leaves open the possi-

bility that the direction of causality is not from classroom teaching behaviors to student instructional ratings, as has been implicitly assumed throughout this article, but rather from prior student ratings to present teaching behaviors (i.e., receiving good or poor ratings causes teachers to behave in characteristic ways). Yet another possibility is that some third variable affects teaching behaviors and student ratings concurrently. Although these alternative causal models can not be ruled out with the data at hand, their plausibility is reduced both by the normal temporal precedence of teaching behavior to teaching evaluation and by the fact that several of the behaviors identified in this study have already been experimentally demonstrated to be causal antecedents of student learning.

The above caveats notwithstanding, the results of this study have important implications for the improvement of college teaching. For one thing, knowledge of specific teaching behaviors that contribute to positive or negative evaluations from students provides a basis for interpreting student evaluations in a more meaningful, diagnostic way. In the absence of such knowledge, an instructor receiving poor evaluations on global dimensions such as "clarity" or "rapport" has no way of identifying the specific behaviors that led to these evaluations, or the specific changes that should be made to bring about improvement. Another possible application is in the development of student evaluation instruments which focus on specific, low-inference teaching behaviors, and thus are more useful for diagnostic or formative purposes than the typical global evaluation forms in current use. Finally, the present results are applicable to consultation or training programs for the improvement of university teaching, including teacher self-training. Instructors who have received poor ratings from students and seek to improve their standing might begin by experimenting with or trying to emulate some of the behaviors shown in this research to be characteristic of highly-related teachers. Obviously this should be done in a sensible, noncontrived way, such that the behaviors selected are compatible with the instructor's basic personality and teaching style. It is clear that some of the

behaviors identified in this study (e.g., giving a preliminary overview) can be acquired relatively easily, whereas others (e.g., speaking expressively) may require extensive practice or training. As one example of a college-level training program focusing on specific classroom behaviors known to contribute to effective teaching, Murray and Lawrence (1980) found that university lecturers trained by a professional actress in the use of vocal variation, pausing, movement and gesture, and related behaviors, showed significant improvement in rated effectiveness relative to untrained lecturers.

### References

- Bertou, P. D., Clasen, R. E., & Lambert, P. An analysis of the relative efficacy of advanced organizers, post organizers, interspersed questions, and combinations thereof in facilitating learning and retention from a televised lecture. *Journal of Educational Research*, 1972, 65, 329-333.
- Centra, J. A. Colleagues as raters of classroom instruction. *Journal of Higher Education*, 1975, 46, 327-337.
- Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum, 1975.
- Cranton, P. A., & Hillgartner, W. The relationship between student ratings and instructor behavior: Implications for improving teaching. *Canadian Journal of Higher Education*, 1981, 11, 73-81.
- Gage, N. L., & Berliner, D. C. *Educational psychology* (2nd ed.). Chicago: Rand-McNally, 1979.
- Hiller, J. H., Fisher, G. A., & Kaess, W. A computer investigation of verbal characteristics of effective classroom lecturing. *American Educational Research Journal*, 1969, 6, 661-675.
- Linn, R. L., Centra, J. A., & Tucker, L. R. Between, within, and total group factor analyses of student ratings of instruction. *Multivariate Behavioral Research*, 1975, 10, 277-288.
- McKeachie, W. J. Student ratings of faculty: A reprise. *Academe*, 1979, 62, 384-397.
- Marsh, H. W. SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 1982, 52, 77-95.
- Mintzes, J. J. Overt teaching behaviors and student ratings of instructors. *Journal of Experimental Education*, 1979, 48, 145-153.
- Murray, H. G. *Evaluating university teaching: A review of research*. Toronto, Canada: Ontario Confederation of University Faculty Associations, 1980.
- Murray, H. G., & Lawrence, C. Speech and drama training for lecturers as a means of improving university teaching. *Research in Higher Education*, 1980, 13, 73-90.
- Rosenshine, B. Content, time, and direct instruction. In P. Peterson & H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications*. Berkeley, Calif.: McCutchan, 1979.
- Rosenshine, B., & Furst, N. F. Research on teacher performance criteria. In B. O. Smith (Ed.), *Research in teacher education: A symposium*. Englewood Cliffs, N.J.: Prentice Hall, 1971.
- Smith, D. G. College classroom interactions and critical thinking. *Journal of Educational Psychology*, 1977, 69, 180-190.
- Solomon, D. Teacher behavior dimensions, course characteristics, and student evaluations of teachers. *American Educational Research Journal*, 1966, 3, 35-47.
- Tom, F. K. T., & Cushman, H. R. The Cornell Diagnostic Observation and Reporting System for Student Description of College Teaching. *Search*, 1975, 5(8), 1-27.
- Ware, J. E., & Williams, R. G. A reanalysis of the Doctor Fox experiments. *Instructional Evaluation*, 1980, 4, 15-18.

Received December 3, 1981 ■