DOCUMENT RESUME

ED 109 235 TM 004 697

AUTHOR Villano, Maurice W.
TITLE The Relationship of Certain Course Characteristics to Student Ratings of Science and Mathematics Teaching at Four-Year and Two-Year Colleges.
PUB DATE Apr 75
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Academic Rank (Professional); Class Size; Colleges; College Students; *Course Evaluation; *Higher Education; Junior Colleges; *Mathematics Instruction; Participant Satisfaction; *Predictor Variables; Rating Scales; *Science Instruction; Statistical Analysis; Student Evaluation; Teacher Rating
IDENTIFIERS Student Appraisal of Teaching

ABSTRACT

The effect of class size, course level, instructor rank, campus, and subject area on 14,366 student ratings of 488 science and mathematics classes at a main university and 18 branch campuses was investigated. Class means were factor-analyzed and four resultant components of teaching performance were submitted to a 2x2x2x2x3 MANOVA. Discriminant functions of significant main effects and interactions did not sustain popular faculty notions ascribing the highest student evaluations to so-called favorable characteristics. Recommendations were made for controlling bias through norming, revision of the rating instrument, and future areas of study. The research instrument used in this investigation was a local adaption of the Student Appraisal of Teaching and is included in the appendix. (Author/RC)

THE RELATIONSHIP OF CERTAIN COURSE CHARACTERISTICS TO STUDENT RATINGS OF SCIENCE AND MATHEMATICS TEACHING AT FOUR-YEAR AND TWO-YEAR COLLEGES

Maurice W. Villano

The Pennsylvania State University

Paper presented at the Annual Meeting
American Educational Research Association
Washington, DC April, 1975

THE RELATIONSHIP OF CERTAIN COURSE CHARACTERISTICS TO STUDENT RATINGS OF SCIENCE AND MATHEMATICS TEACHING AT FOUR-YEAR AND TWO-YEAR COLLEGES

Maurice W. Villano

The Pennsylvania State University

Background

The purpose of this paper was to report the effect of selected course variables on the outcome of student instructional ratings at two 4-year colleges and at seventeen 2-year branch colleges or learning centers of a major state university where the College of Science had systematically evaluated all science and mathematics teaching of approximately 15,000 students in almost 500 classes during a recent academic term. The specific aspects of teaching performance being rated were described in the context of relevant course characteristics such as class enrollment, course level, instructor rank, type or location of campus, and major subject or discipline area. Multivariate statistical analyses were used to explain the complex interrelationships. Results were interpreted with a view toward sustaining or rejecting the sometimes popular notion that ascribes the highest student ratings to classes with so-called "favorable" course characteristics. Implications for the evaluation of science and mathematics teaching were discussed, including the use of controls for "bias" in ratings, revision of the instrument, and possible future areas of study.

Reviews of the literature by Costin, Greenough & Menges (1971), McKeachie (1973) and others have found the extent to which student course ratings are influenced by instructional variables generally to be inconclusive or even contradictory. Research in this area has continued, however, as witnessed by some of the more recent validation studies. Aleamoni & Graham (1974) investigated the effect of class size, course level, and instructor rank on Course Evaluation Questionnaire (CEQ) results at the University of Illinois. Bejar & Doyle (1974) noted that factor structures of student ratings were generalizable over those attributes that separated main campus students from those at academic centers. Miller (1974) discussed the impact of subject matter on research findings. In one of the few studies in science teaching, Cornwall (1974) reported the differential effects of instructional variables on chemistry courses rated at Wisconsin.

Despite the lack of convincing support for their position, many faculty members view the highest course ratings as being assigned to small classes of upper-division courses taught by senior ranks at main university campuses. Persistence of this notion has hindered full acceptance of student evaluation systems by professors who expect rating devices to measure what is being realistically perceived by students within the context of relevant instructional variables. Many institutions control for possible "bias" introduced by these variables by providing separate norming of evaluation results. There is increasing evidence, however, that student ratings vary also as a function of the specific facets of teaching performance being evaluated, and that these complex interrelationships can best be studied using large data samples and a multivariate statistical analysis approach.

Research on instructor and course ratings in science is especially meager. Further studies are required to improve student evaluation systems with their important implications for education at the university level.

Procedures

The research instrument used in this investigation was a local adaptation of the Student Appraisal of Teaching (SAOT) form by R. I. Miller (1972) (see Appendix). This rating device has 15 Likert-type item-statements to which subjects respond on a four-step scale. The response data were collected on over 18,000 students in undergraduate and graduate science and mathematics classes at two 4-year colleges (106 classes including the main university campus) and 17 two-year colleges or learning centers (382 classes) of a major state educational institution during the Winter 1974 academic term. The sample was screened for missing item responses on individual SAOT forms and reduced to 14,366 Ss representing a total of 488 class-sized groups.

The response data described above were processed and the item means for each class were computed. The generated 15 by 15 intercorrelation (R) matrix representing between-instructor covariation was submitted to a principal-components analysis (Hotelling, 1933) with unities inserted in the diagonal. By slightly over-factoring (using an eigenvalue cut-off of 0.96 as the root criterion), three psychologically-meaningful components were extracted and subjected to an orthogonal (Varimax) rotation (Kaiser, 1958). Component scores were calculated for the three dimensions of the SAOT that were descriptive of teaching performance using the direct measurement method:

$$X = (F'F)^{-1} F' Z$$

where X = factor (component) scores, $(F'F)^{-1}$ = inverse of the original item R matrix, F' = orthogonally-rotated factor loading matrix, and Z = standard scores (Harman, 1967).

These data (component scores) were then treated as dependent variables in a 2 x 2 x 2 x 2 x 3 completely-crossed factorial design with two levels each for fixed factors of Size (class enrollment of 1-20 and 21 & above), Level (introductory and advanced courses), Rank (instructor & assistant professor and associate & full professor), Campus (four-year colleges and two-year learning centers), and three levels of subject area or Discipline (biological, physical, and mathematical sciences). Multivariate analysis of variance (MANOVA) was conducted to examine the main effects and interactions of the five factors. Step-down F tests were employed to indicate the dependent variables that contributed most to the multivariate F ratios. Multiple discriminant function analysis was used to assess the extent and direction of significant effects and interactions. The foregoing procedures were executed using the SUNY (Buffalo) MULTIVARIANCE Program (Finn, 1968).

A possible limitation of this study from the methodological standpoint was the arbitrary collapsing of several levels of the factors in the MANOVA design to only two or three levels. While this was based, in part, on an a priori analysis of the data, some information may have been lost, thus affecting the results. There were some missing cells or cells with only one observation; consequently, the means of the null subclasses were estimated and the residual sum of squares and the within sum of squares were not equal. The three-way and higher-level interactions were not tested. Several inestimable sets of contrasts for the higher-level interactions wer removed from the model and these interaction sums of squares were pooled with the residual to form the error term, with the degrees of freedom adjusted accordingly.

Results and Conclusions

The principal-components analysis yielded three distinct, independent, and robust dimensions underlying the SAOT that accounted for all of the items and over 77-percent of the total instrument variance (see Table 4). These factors, descriptive of the teaching performance being rated, were labelled, with their respective Coefficient Alpha reliabilities, as follows: Component I -- Instructional Qualities (.96); Component II -- Thinking-Participation (.84); and Component III -- Instructor Fairness (.88).

The factor array comprising the first component (see Table 1) consisted of seven items that appeared to relate to general instructional qualities of organization, clarity, motivation, and interest. The final instructor rating and "recommending the course to a friend" also helped define this factor indicating that all seven items in the aggregate were inseparable from the overall evaluation itself.

The second component (Table 2), encompassing three items only, addressed itself to the encouragement of thinking, participation, and discussion in the classroom. The third and last component (Table 3), composed of five items, was indicative of instructor interaction with the students regarding fairness in grading, rendering assistance, and other classroom relationships.

The MANOVA disclosed statistically significant main effects for all five factors of Size, Level, Rank, Campus, and Discipline and for three 2-way interactions: Size by Level, Size by Rank, and Level by Rank. The discriminant functions separated cells with high and low mean scores on the dependent variables into identifiable groups which took on special meaning in this study. An extended discussion of the interpretations follows.

For the Size main effect, the Thinking/Participation variable was solely responsible for the highly significant (.0001) multivariate F ratio (Table 6). The discriminant function for this effect (Table 8) indicated that ratings high on Thinking/Participation separated the large-sized classes, highest on this function, from the small-sized classes.

For the Level main effect, the Thinking/Participation variable again accounted singularly for the multivariate F (.02). The discriminant function for this effect signified that evaluations high on Thinking/Participation discriminated between the lower-level courses, highest on this function, from the upper-level courses.

For the Rank main effect, both the Instructional Qualities and the Instructor Fairness variables contributed significantly to the multivariate F ratio (.01). The discriminant function for this effect suggested that low ratings on Instructional Qualities and Instructor Fairness differentiated between junior-ranked faculty, highest on this function, and senior-ranked faculty.

For the Campus main effect, the Thinking/Participation variable, once more, most individually influenced the multivariate F (.03). The function for this effect revealed that high ratings on Thinking/Participation tended to distinguish the four-year colleges, highest on this function, from the two-year colleges or learning centers.

For the Discipline or subject area main effect, both the Thinking/Participation and Instructional Qualities variables explained tha highly significant F (.0001). The function for this effect noted that high ratings on these variables somewhat disassociate the biological sciences, highest on this function, from the physical and mathematical sciences.

For the Size by Level interaction, the Instructor Fairness variable was mainly responsible for the significant (.01) multivariate F ratio (Table 7). The discriminant function for this interaction indicated that low ratings on Instructor Fairness separated the small-sized upper-level courses, highest on this function, from the large-sized lower-level courses.

For the Size by Rank interaction, the Thinking/Participation variable primarily accounted for the multivariate F ratio (.04). The discriminant function for this interaction showed that high ratings on Thinking/Participation and to a lesser extent Instructor Fairness tended to differentiate between large-sized classes taught by senior faculty, highest on this function, and small-sized classes taught by junior faculty.

For the Level by Rank interaction, the Instructor Fairness variable contributed most to the multivariate F ratio (.03). The discriminant function for this interaction was based on a low rating on Instructor Fairness and to a lesser extent Thinking/Participation and appeared to distinguish between upper-level courses taught by junior faculty, highest on this function, from lower-level courses taught by senior faculty members.

Based on the above findings, it was concluded that the hypothesis that assigned the highest student evaluations to courses with "ideal" characteristics was not entirely supported. The results were somewhat equivocal in that the means for small-sized classes were less than for large-sized classes, and were less for upper-level than lower-level courses; however, senior faculty members fared better than junior ones, and ratings at the four-year institutions exceeded those at the two-year colleges and learning centers. The two-way interactions and discriminant functions underscored these relationships in a multivariate setting for most instances.

From an inspection of the cell mean scores following the multiple discriminant function analysis (Table 5), a vague but discernable pattern emerged that tended to separate the class-sized groups into two gross evaluation categories: 1) high ratings--large-sized, lower-level biological science courses taught by senior faculty at the four-year colleges; and 2) low ratings--small-sized, upper-level physical and mathematical science courses taught by junior faculty members at the two-year colleges and learning centers. While this dichotomy is not always clearcut--there are several anomolies--it points to a need for the separate norming of evaluation results if they are to be meaningful.

The inordinate influence of the Thinking/Participation variable on three main effects and one interaction was puzzling. This component, consisting of only three items and accounting for 22-percent of the variance for the SAOT, comprises a minor part of many evaluation forms. The role of critical thinking in science evaluation may be underestimated and should bear further investigation in future studies.

It was recommended that the Student Appraisal of Teaching (SAOT) instrument be retained in its present form pending further study. Such investigation should include but not be limited to an exploration of the factorial invariance or stability of the teaching performance aspects underlying the SAOT and an examination of changes in instructor scores for succeeding terms along the same dimensions brought out in the present study.

REFERENCES

Aleamoni, L. & Graham, M. The relationship between CEQ ratings and instructor's rank, class size, and course level. Journal of Educational Measurement, 1974, 11(2), 189-202.

Bejar, I. & Doyle, K. Generalizability of factor structures underlying student ratings of instruction. Paper presented at the Annual Convention of the American Educational Research Association, Chicago, April 15-19, 1974.

Cornwall, C. Statistical treatment of data from student teaching evaluation questionnaires. Journal of Chemical Education, 1974, 51(3), 155-160.

Costin, F., Greenough, W., & Menges, R. Student ratings of college teaching: reliability, validity, and usefulness. Review of Educational Research, 1971, 41(5), 511-535.

Finn, J. Multivariance: univariate and multivariate analysis of variance, covariance, and regression. A Fortran program, version 4, June 1968, Department of Educational Psychology, State University of New York at Buffalo.

Harman, H. Modern Factor Analysis. Chicago: University of Chicago Press, 1967.

Hotelling, H. Analysis of a complex set of statistical variables into principal components. Journal of Educational Psychology, 1933, 24, 417-441, 498-520.

Kaiser, H. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23(3), 187-200.

McKeachie, W. Correlates of student ratings. In A. Sockloff (Ed.) Proceedings: First Invitational Conference on Faculty Effectiveness as Evaluated by Students. Philadelphia: Temple University Measurement and Research Center, 1973, 213-218.

Miller, R. Evaluating Faculty Performance. San Francisco: Jossey-Bass, 1972.

Miller, R. Developing Programs for Faculty Evaluation. San Francisco: Jossey-Bass, 1974.

STUDENT APPRAISAL OF TEACHING

Thoughtful student reaction can help improve teaching effectiveness. This questionnaire is designed for that purpose. Your assistance is appreciated. Please do not sign your name.

Rate your professor on each item by indicating the extent of your agreement with the statement.

If you conclude that you cannot give an informed answer, please indicate "Cannot answer" (E). All students should answer item 15.

A	B	C	D	E
Strongly agree	Agree	Disagree	Strongly disagree	Cannot answer

1.	Class presentations are well organized.	A	B	C	D	E
2.	Mastery of the subject matter is evident.	A	B	C	D	E
3.	Critical thinking and analysis are encouraged.	A	B	C	D	E
4.	Examinations that require creative, original thinking are given.	A	B	C	D	E
5.	The presentation motivates your best effort.	A	B	C	D	E
6.	Ideas presented are clearly explained.	A	B	C	D	E
7.	The professor accepts and is willing to discuss student viewpoints different from his own.	A	B	C	D	E
8.	You are encouraged to seek the professor's help when needed.	A	B	C	D	E
9.	Grading policies are adequately explained and followed.	A	B	C	D	E
10.	The professor exhibits a genuine interest and enthusiasm in teaching.	A	B	C	D	E
11.	Questions on examinations are fair.	A	B	C	D	E
12.	The professor encourages student participation and welcomes questions and discussion.	A	B	C	D	E
13.	Learning objectives for the course and/or examinations are clearly stated.	A	B	C	D	E
14.	I would recommend this instructor to a friend.	A	B	C	D	E
15.	Compared to other professors you have had at this university, how effective is the professor in this course (considering the previous points)?	A	B	C	D	E

A	B	C	D
One of the most effective	More effective than most	Not as effective as most	One of the least effective

TABLE 1

COMPONENT I: INSTRUCTIONAL QUALITIES

Item No.	Loading	SAOT Statement
1	.86	Class presentations are well-organized.
6	.76	Ideas presented are c [illegible] plained.
15	.75	Compared to other professors you have had at this university, how effective is the professor in this course (considering the previous points)?
2	.71	Mastery of the subject matter is evident.
14	.67	I would recommend this instructor to a friend.
5	.67	The presentation motivates your best effort.
10	.55	The professor exhibits a genuine interest and enthusiasm in teaching.

TABLE 2

COMPONENT II: THINKING-PARTICIPATION

Item No.	Loading	SAOT Statement
3	.85	Critical thinking and analysis are encouraged.
4	.82	Examinations that require creative, original thinking are given.
12	.64	The professor encourages student participation and welcomes questions and answers.

TABLE 3

COMPONENT III: INSTRUCTOR FAIRNESS

Item No.	Loading	SAOT Statement
9	.79	Grading policies are adequately explained and followed.
11	.72	Questions on examinations are fair.
8	.62	You are encouraged to seek the professor's help when needed.
7	.62	The professor accepts and is willing to discuss student viewpoints different from his own.
13	.62	Learning objectives for the course and/or examinations are clearly stated.

TABLE 4

COMPONENT LOADING MATRIX USING ORTHOGONAL TRANSFORMATION[a]

Item No.	Instructional Qualities (I)	Thinking-Participation (II)	Instructor Fairness (III)
1	86		
6	76		47
15	75		47
2	71	40	
14	67		57
5	67	41	46
10	55	47	44
3		85	
4		82	
12		64	54
9			79
11			72
8		53	62
7		53	62
13	57		62
Total variance accounted for:	29.3%	22%	25.9%

[a]Loadings multiplied by 100 and rescaled to eliminate negative signs. Coefficients less than .40 deleted.

TABLE 5

CELL MEANS (Z SCORES) FOR THREE DEPENDENT VARIABLES

[a]N	[b]Size	Levl	Rank	Camp	Disc	Inst Qual	Instr Fair	Think/Part
1	sml	lo	jr	4	phys	0.590	0.081	0.224
1	sml	lo	jr	4	math	0.581	-0.875	0.224
15	sml	lo	jr	2	biol	0.369	0.604	0.499
31	sml	lo	jr	2	phys	-0.128	0.386	-0.267
86	sml	lo	jr	2	math	-0.166	-0.035	-0.004
2	sml	lo	sr	4	math	-0.316	0.507	-0.701
4	sml	lo	sr	2	biol	0.535	-0.297	-0.779
15	sml	lo	sr	2	phys	0.153	0.162	-0.802
2	sml	lo	sr	2	math	0.119	-1.109	0.320
4	sml	up	jr	4	biol	-0.505	-0.403	-0.718
2	sml	up	jr	4	phys	-2.242	-0.557	1.562
9	sml	up	jr	4	math	0.740	-0.411	-0.548
3	sml	up	jr	2	biol	0.266	-0.369	-1.535
25	sml	up	jr	2	phys	-0.300	0.065	-0.740
10	sml	up	jr	2	math	-0.163	-0.785	-0.207
6	sml	up	sr	4	biol	0.519	-0.064	0.445
5	sml	up	sr	4	phys	0.569	0.210	-0.656
3	sml	up	sr	4	math	-0.579	0.219	-1.095
2	sml	up	sr	2	phys	-1.649	0.305	-0.977
4	lrg	lo	jr	4	biol	0.245	-0.089	1.050
10	lrg	lo	jr	4	phys	0.110	0.260	0.698
12	lrg	lo	jr	4	math	-0.010	-0.386	-0.147
37	lrg	lo	jr	2	biol	0.461	-0.149	0.628
47	lrg	lo	jr	2	phys	-0.238	-0.220	-0.023
68	lrg	lo	jr	2	math	-0.133	-0.138	0.090
8	lrg	lo	sr	4	biol	0.286	0.139	0.986
7	lrg	lo	sr	4	phys	-0.211	0.530	0.267
4	lrg	lo	sr	4	math	0.113	0.228	0.654
6	lrg	lo	sr	2	biol	-0.013	0.554	0.508
13	lrg	lo	sr	2	phys	0.181	-0.057	0.069
10	lrg	lo	sr	2	math	0.540	-0.349	-0.178
3	lrg	up	jr	4	biol	0.319	-0.397	-0.063
4	lrg	up	jr	4	phys	-0.475	0.207	-0.308
6	lrg	up	jr	4	math	-0.465	0.315	-0.305
5	lrg	up	jr	2	phys	0.215	0.242	-0.865
3	lrg	up	jr	2	math	0.441	0.067	1.096
4	lrg	up	sr	4	biol	1.004	-0.002	0.903
10	lrg	up	sr	4	phys	0.403	1.083	0.247
1	lrg	up	sr	4	math	0.314	3.598	-1.762

[a]N=number of sections per cell. [b]See p.4 for definition of factors & levels.

TABLE 6

TESTS OF SIGNIFICANCE: MAIN EFFECTS

Dependent Variable	Univariate F	Step-down F	Multivariate F
SIZE	$(df_h=1,$	$df_e=467)$	$(df_{num}=3, df_{den}=465)$
Instructional Qualities	1.82	1.82	F = 9.54
Instructor Fairness	0.26	0.23	p<.0001
Thinking/Participation	25.91****	26.47****	
LEVEL	$(df_h=1,$	$df_e=467)$	$(df_{num}=3, df_{den}=465)$
Instructional Qualities	0.07	0.07	F = 3.58
Instructor Fairness	0.29	0.29	p<.02
Thinking/Participation	10.52**	10.36**	
RANK	$(df_h=1,$	$df_e=467)$	$(df_{num}=3, df_{den}=465)$
Instructional Qualities	5.68*	5.68*	F = 3.94
Instructor Fairness	5.89*	6.06*	p<.01
Thinking/Participation	0.12	0.03	
CAMPUS	$(df_h=1,$	$df_e=467)$	$(df_{num}=3, df_{den}=465)$
Instructional Qualities	0.27	0.27	F = 3.09
Instructor Fairness	0.94	0.96	p<.03
Thinking/Participation	7.80**	8.01**	
DISCIPLINE	$(df_h=2,$	$df_e=467)$	$(df_{num}=6, df_{den}=930)$
Instructional Qualities	7.03***	7.03***	F = 7.00
Instructor Fairness	2.80	2.76	p<.0001
Thinking/Participation	10.55****	11.15****	

*p<.05, **p<.01, ***p<.001, ****p<.0001

TABLE 7

TESTS OF SIGNIFICANCE: INTERACTIONS

Dependent Variable	Univariate F	Step-down F	Multivariate F
SIZE by LEVEL	(df_h=1, df_e=467)		(df_{num}=3, df_{den}=465)
Instructional Qualities	1.02	1.02	F = 5.37
Instructor Fairness	13.95***	14.05***	$p<.01$
Thinking/Participation	0.83	1.01	
SIZE by RANK	(df_h=1, df_e=467)		(df_{num}=3, df_{den}=465)
Instructional Qualities	0.04	0.04	F = 2.82
Instructor Fairness	2.17	2.18	$p<.04$
Thinking/Participation	6.08*	6.20*	
LEVEL by RANK	(df_h=1, df_e=467)		(df_{num}=3, df_{den}=465)
Instructional Qualities	0.07	0.07	F = 3.10
Instructor Fairness	5.35*	5.37*	$p<.03$
Thinking/Participation	3.69	3.83	

*$p<.05$, **$p<.01$, ***$p<.001$

TABLE 8

STANDARDIZED DISCRIMINANT COEFFICIENTS

Main Effect/ Interaction	Bartlett Chi-Sq	Dependent Variable Instructional Qualities	Instructor Fairness	Thinking/ Participation
<u>SIZE</u>	27.81 df=3, p<.0001	0.30	-0.08	0.97
<u>LEVEL</u>	10.62 df=3, p<.02	-0.03	-0.15	0.99
<u>RANK</u>	11.67 df=3, p<.01	-0.70	-0.72	0.05
<u>CAMPUS</u>	9.18 df=3, p<.03	0.23	0.33	0.93
<u>DISCIPLINE</u>	41.14 df=6, p<.0001	0.66	-0.08	0.79
<u>SIZE by LEVEL</u>	15.85 df=3, p<.01	-0.29	-0.94	-0.25
<u>SIZE by RANK</u>	8.38 df=3, p<.04	0.13	0.52	0.86
<u>LEVEL by RANK</u>	9.22 df=3, p<.03	-0.14	-0.77	-0.65