DOCUMENT RESUME

ED 047 630 HE 001 998

AUTHOR Purohit, Anal; Magoon, A. Jon
TITLE The Validity of Student-Run Course Evaluations.
INSTITUTION Delaware Univ., Newark.
PUB DATE Jan 71
NOTE 22p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS College Faculty, *College Students, Course Evaluation, *Evaluation, *Higher Education, *Teacher Evaluation, Teaching Quality, *Validity
IDENTIFIERS *Delaware University

ABSTRACT

After a review of the literature of evaluations by students of instructors and courses, this paper discusses 3 different evaluation questionnaires given in successive years (1968 through 1970) at the University of Delaware. Each of these forms represented an attempt to make the ratings less susceptible to the "halo effect," which was defined as the "marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of qualities by this general feeling." The results of these forms were factor analyzed and the findings indicated that only 4 factors were in these course evaluations. The major factor was characterized as "instructor impact" and was interpreted as having a large "halo effect." The other factors were characterized as dimensions of instructional procedures, course work load, and quality of instructional materials. Several suggestions are offered on how to improve the validity of the evaluation instruments. (AF)

THE VALIDITY OF STUDENT-RUN COURSE EVALUATIONS[1]

Anal Purohit and A. J. Magoon

University of Delaware

The student government associations on a large number of college and university campuses currently run course evaluations. The purpose of typical student-constructed course evaluations is to act as a "valuable source of feedback for the faculty" and to "provide students with a guide in selecting courses and instructors which best suit their needs and interests." It has also been claimed that the course evaluation "should be considered as the honest effort of students to provide valid, unbiased information about teaching ability and course structures," and that it is a "stimulus for more encompassing, more penetrating, and more frequent dialogue among all members of the campus community concerning the nature of instruction."[2] Students felt the need for a public analysis of courses and instructors and so have put together short rating forms by which all instructors and courses could be rated, and have provided a summary analysis (usually mean rating information on each item) in published form for community consumption.

The response of various community members to the new student-run course evaluations has been mixed. Students generally are pleased with much of the information so obtained, for the ratings appear to identify courses which are known to be notably good or poor by the usual student standards. The ratings are valid for two purposes: (1) publicizing information among students, information that previously flowed along an inefficient grapevine; and (2) signaling to faculty and administrators the likes and dislikes of students. Student-run course evaluations are the reactions of students to faculty and course structures, and thus validity is axiomatic, as Thorndike and Hagen (1969, p. 433) point out. On the other hand many students, faculty,

---

[1]We wish to express our sincere appreciation to the Student Government Association of the University of Delaware for their interest and cooperation which made this study possible.

[2]Course Evaluation, Fall 1968, University of Delaware, Newark, Delaware p. v.

and administrators feel that the ratings are ambiguous in a number of ways when an attempt is made to infer that good or poor teaching is reflected by the ratings. Many faculty committees and administrators feel the need to use student course evaluation information, but at the same time wish to know what factors must be considered when interpreting such ratings.

One of the most frequently voiced criticisms of instructor and course ratings is that the ratings on a number of supposedly distinct instructor traits merely reflect a "halo effect" of the instructor's personality, or "showmanship" (Slobin and Nichols, 1969). A halo effect is technically defined as the "marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of qualities by this general feeling" (Thorndike, 1920). One result of such halo errors is to force ratings on separate items in the direction of the general impression, which in effect introduces a spurious amount of positive correlation between distinct instructor rating items (Guilford, 1954, p. 279). Large halo effects would obviously result in large components of variance attributable to the general feeling about the instructor when a set of instructor ratings were submitted to a principle components analysis. This study focuses on principle components structures for three sequential course evaluation rating forms. Each successive form represented an attempt to make the ratings less susceptible to halo errors. The results of this effort to construct better halo-free rating forms give us cause to suspect that this cannot be accomplished.

Literature review

It seems that the course and the instructor evaluations run by students is the "now" thing for the students. Such student evaluations have actually been taking place for decades and various biases have been investigated for nearly as long. If we look at the literature, it is found that even in the 1920's and 30's, Remmers and Guthrie did studies on instructor ratings by high school or college students. As early as 1927, Guthrie discussed whether the college students were competent judges of the quality of teaching in their courses. In that study, high reliability was found indicating that student opinion of teachers is at least consistent at stable when the acquaintance with the instructors was extensive. Guthrie suggested that there is perhaps no method by which the ultimate validity can be determined, unless we assume that a general agreement discovered between student opinion of teaching and various

other criteria, such as faculty opinion examination results, subsequent records of students, indicates an objective validity." (p. 1927) Remmers (1929) did a study using the Purdue Rating Scale for instructors in connection with the departmental differences in the quality of instruction as seen by students. He concluded in his study that departmental as well as individual instructor patterns of teaching personality exists as far as the students view the situation; also that the desirable traits tend to vary together within a given department but the variations from trait to trait with a given department are likely to be significant and that the interdepartmental variations are quite extreme and point to considerable differences in teaching effectiveness. In another research article in the 1930's, Remmers found that reliable judgments of classroom traits of instructors can be obtained from both high school and college pupils and that it was probable that high school pupils will invest the practice teachers with less halo than college students will with their instructors.

From the research of Bendig (1944, 45), Isaacson, et al (1963,64), Coffman (1952), White (1964), Costonas (1962) and Smalzreid (1943), it is found that there is a certain generality to the factors derived from the questionnaires that were used. For example, Bendig found 3 factors for 10 scales of the Purdue University rating scale 1) a general factor, 2) instructional competence, and 3) instructional empathy. He described a general factor as a "halo effect." Isaacson found 6 factors which accounted for 95% of the response variance from 46 items. The factors were 1) a general halo effect, 2) over load factor, 3) structure factor, 4) feedback factor, 5) group interest, and 6) friendly, democratic behavior. Note that both Bendig and Isaacson attribute the first factor to a "halo effect," implying this may be common to many course evaluations.

A more recent study by Deshpande, et al (1970) utilized a rating form where critical incidents provided the focus. The study is fraught with technical limitations resulting from misapplied factoring procedures as well as small sample size. The results of this study, nevertheless reveal little evidence of systematic halo influences.

There have been several suggestions as to how halo influences are to be avoided when constructing rating forms. Symonds (1925) quite early suggested that rating items need to be based on clearly observable behavior that is clearly defined, and that character traits or traits of high moral importance be avoided.

Thorndike and Hagen (1969, pp. 434-436) have listed several examples of how rating variables can be made more explicit, mainly consisting of elaborations on definitions of the traits being rated. The cost of such embellishment and complexity is of course a much lengthier instrument.

The present study focuses on three different questionnaires given in successive years at the University of Delaware. The focus is on inter-item relationships in these typical student questionnaire instruments. It should be noted that the construction of the second and third questionnaires represent an attempt to make such an instrument more specific in its assessment of instructional quality, i.e., the students tried to place a heavier emphasis on evaluation of clearly visible instructor behavior in order to make the instruments more halo-resistant.

Data for the analyses of the first two sets of questionnaire items consisted of mean ratings from randomly-selected classrooms (N_1 = 100, N_2 = 198), and for the third questionnaire randomly-selected individual responses (N_3 = 127). Ratings for each item could be made along five-point scales (e.g., poor to excellent, etc.). The analyses carried out was a principal components analysis (followed by a varimax rotation of the components with eigenvalues greater than 1.0) of the correlation matrix for all the questionnaires.

1968 Questionnaire

The results revealed that variation between classrooms may be described by approximately seven different factors (See Table I). Each factor represents an independent way that classroom ratings differed. These factors have been labeled as to their apparent meaning and are described in order from least significant to most important.

1) Term papers: this factor represents a set of two items dealing with

   a) the number of term papers assigned and
   b) the amount of help the instructor gave outside of class. Instructors who assigned relatively more term papers were generally rated as being more helpful outside the classroom.

2) Instructional procedure: This retains items dealing with

a) whether the course was mostly lecture or mostly discussion.
b) the difficulty of the text, and to a lesser extent,
c) the instructor's interest in teaching and
d) the fairness of examinations.

Courses which are mostly lectures quite often have textbooks rated as relatively more difficult, have instructors rated as slightly more interested in teaching the course, and are perceived as having somewhat less fair examinations.

3) Course difficulty. This is a composite of

a) an item bearing directly on course difficulty, and to a lesser extent, items concerning
b) the importance of attending class,
c) the work load, and
d) the reading load.

Course reported as difficult were also usually rated as having a heavier work and reading load, and more often than not, one had to attend classes in order to do well.

4) Homework requirements: Items having to do with

a) the utility of the text,
b) the value of readings, and
c) the work load substantially define this dimension.

5) Examination objectivity. Items which determined

a) whether the examinations were mostly objective or mostly essay,
b) whether the examinations were fair, and
c) reading load were determinants of this factor.

When essay examinations were the rule, these surprisingly were seen as a fairer test of knowledge. Reading load was also heavier where the examinations were given. There is considerable evidence in the psychological testing literature to show that objective tests are indeed fairer to examinees. Thus, it can be argued that

the raters of these questionnaires were either unaware of what an objective test should be or objective test construction on the University of Delaware campus is not what it should be or perhaps both.

6) Examinations grading fairness: Items offering a measure of

   a) the number of hourly examinations given,
   b) test grading fairness,
   c) examination difficulty, and
   d) examination fairness make up a very interesting domain.

When relatively more examinations are given in a course, these are perceived as being graded more fairly, and more difficult than usual, and are rated as fairer tests of the examinees knowledge.

7) Instructor impact: By far the most prominent independent source of variation in ratings is defined by a subset of items concerned with instructor behavior. This set of items (accounting for one quarter of all the variance) is composed of

   a) the instructional effectiveness of the instructor,
   b) instructor knowledge of subject matter,
   c) how well the instructor organized the course,
   d) instructor delivery,
   e) instructor interest in teaching his class, and
   f) the amount of help the instructor gave to students outside of class. Other variables related to this dimension were
   g) satisfaction with the course, and
   h) recommendation of the course. This dimension may be considered due to the "halo effect."

1969 Questionnaire

The items for this questionnaire were revised in an attempt to specify unique types of instructor behavior and were essentially different in phraseology from the 1968 questionnaire. However, the rotated factor pattern was very similar in meaning to that of the first questionnaire (See Table II). In this questionnaire, the analysis resulted in approximately 4 factors. Again, each factor

has been named with respect to what it seems to be measuring. The variables that are given for each factor are given by importance, meaning that the first ranked in variable listed is more important in measuring the factors than the next one.

1) Work load: Items consisting of

   a) the reading difficulty,
   b) the difficulty of the material covered in the class,
   c) the amount of total work load,
   d) the exam difficulty,
   e) the amount of reading load, and to a lesser extent,
   f) the value of assignments are elements of this factor. It seems that if the amount of the total work load is high but correspondingly the value of assignments is also higher.

2) Textbook: The items comprising this dimension are

   a) the rated quality of textbook used,
   b) the value of assignments,
   c) the relevance of the course, and
   d) the difficulty of the readings.

Apparently, the textbook is partially important in determining the relevance of the course and overall evaluation of the course.

3) Classroom dialogue: The items involved in this factor are

   a) the relative amount of conformity,
   b) the emphasis on creativity,
   c) the opportunity to question in the classroom,
   d) the instructor's effectiveness in moderating class discussion,
   e) the value of the class discussion,
   f) the amount of intellectual stimulation,
   g) fairness in grading,
   h) the overall evaluation of the instructor and
   i) the overall evaluation of the course.

It is interesting to note that creativity and conformity reflect more on class discussion and format than on the instructor's presentation - or interest in the course.

4) Instructor impact. This factor is the strongest underlying dimension, and is highly related to the overall evaluation of the instructor. It consists of

   a) the overall evaluation of the instructor,
   b) the instructor's organization of the course,
   c) the instructor's presentations and explanations,
   d) the overall evaluation of the course,
   e) the value of lecture,
   f) the instructor's apparent interest,
   g) the degree of intellectual stimulation,
   h) the instructor's relative effectivenss in moderating discussion,
   i) the instructor's grading fairness,
   j) his respect for the students,
   k) the value of the discussion,
   l) the frequency of opportunity to question in class,
   m) the relevance of the course, and
   n) the availability of the instructor outside the classroom.

All instructor behavior aspects listed above load positively on this factor, indicating that if the instructor is rated highly on one item, he will usually be rated highly on all others in the group. Items which would not ordinarily be thought to be related, e.g., "intellectual stimulation" and "fairness in grading" are strong bedfellows in this instructor impact composite.

1970 Questionnaire

Noting the high degree of overlap in ratings for the revised 1969 rating form, a more severe change in rating items was undertaken in order to "escape halo effects." Utilizing the results of Deshpande et al (1970) in the selection of 17 critical incidents which purportedly tapped 14 separate instructor trait dimensions, a new rating form was constructed. The results of a principal components analysis for .19 items and 127 randomly-selected raters are tabulated in Table III, and reveal only four dimensions with eigenvalues greater than 1.

1) Instructor impact: This factor again is the strongest underlying dimension here accounting for more than one fourth of the total set variance. Again such rating

items as explanation of course policies, the logic of course planning, teaching effectiveness, accuracy of the instructor's method of evaluation, the clarity of presentation, the instructor's advice as to how to improve coursework, and overall instructor evaluation, and to a lesser extent, course evaluation are items which define this factor in a major way. To an impressive extent it appears that this factor matches in definition the instructor impact dimensions found in the previous evaluation forms.

2) Instructor Rapport: Items focusing on encouragement to ask questions, courteousness of the instructor, encouragement of creativity, informedness of the instructor, and emphasis on seeing beyond the course limits combined to form the second largest component. This appears to function much like the "Classroom Dialogue/Instructional Procedure" dimension discussed in the first two evaluation forms.

3) Textbook Quality: A third dimension of interest centers on the clarity and relevance of the textbook, as well as a smaller relation to overall course evaluation, instructor supplementation of the text and how far the instructor looked beyond the limits of the course. Items are very similar to the textbook dimension in earlier rating forms.

4) Course Difficulty: The difficulty of examinations and the difficulty of the total workload formed a virtually independent dimension again, quite familiar as the difficulty domain of earlier studies.

Table IV presents an item-for-item match on the four similar factors for all three rating instruments. Items for which no matches could be found are also tabled. It should be noted that items were supposedly improved so that halo effects would be less apparent in each successive form, i.e., where very global traits or general aspects of the course were rated in 1968, specific critical incidents were used in 1970. It was expected that as behavior or course aspects to be rated were made more specific, a greater number of factors would emerge from the rating form, indicative of an inherent complexity of classroom structure. The results are at variance with this supposition. Four main dimensions were always observed, except for the 1968 ratings where three separated dimension of differences [(1) frequency of exams,

(2) frequency of term papers, (3) type of examinations were mistakenly (by students) included as rating items. It appears that students for all intents and purposes naturally evaluate courses and instructors along a maximum of four dimensions.

Instructor impact is a strong dimension in each situation, and a minimum of eight items can be matched in terms of content across all three instruments. Apparently, this impact dimension is a fairly reliable phenomena when many items focus on instructor behavior. It is also indicative of an overall "halo effect" due either to a general ambiguity as to meaning of many items, or the raters' inability to rate distinct aspects of instructor behavior and hence only represent a broad, general evaluation of the instructor rather than a precise evaluation of particulars.

The question of validity of these evaluations because of the "halo effect" of course remains. Let us again go back to the two main purposes of typical student-constructed course evaluations: (1) "a valuable source of feedback for the faculty"; and (2) "to provide students with a guide in selecting courses and instructors which best suit their needs and interests." Because of the "halo effect," rating results may not be valid for specific variabales. But in general, if the faculty member wants to have feedback on whether there is overall student satisfaction or not, these evaluations are valid. It has been argued "since ratings on specific traits correlate closely with final estimates of personal fitness . . . an overall judgment is more likely to be correct if made after the rater's attention has been focused successively on several of the candidate's specific traits" (Bingham, 1939, p. 226). For the second purpose of the questionnaire, it should be more valid, for a majority of the students will find the same types of instructor characteristics as the students who rated the instructor.

For other usages beyond these it is difficult to say that such ratings are valid. Take for example the faculty committee which wishes to use course evaluation information to make promotional decisions regarding faculty. The crucial information contained in the instructor impact information apparently reflects an overall general impression as to the quality of the instructor. It is quite reminiscent of a classic study by Ewart et al (1941) of ratings of worker competency; all characteristics, however logically independent, were moderately correlated and thus indicated that a single evaluative rule colored all separate ratings. In another study, it was found that rated qualities like "productivity" correlate only slightly with actual productivity, (Stockford and Bissel, 1949). Thus the faculty committee

or administrator must realize that the instructor traits rated by students are not necessarily the causative agents of the students' high or low valuation of the instructor, but that just the reverse could be true: the student valuation of the instructor, which could well be based on criteria that are not understood but which in turn vary from student to student, causes ratings on various evaluative items to vary concomitantly.

Cronbach (1970, pp. 574-576) has noted recently that many studies support the contention that rating information of humans by humans is essentially three dimensional in nature. By far the largest dimension is an "evaluative" one, followed by "potency" and "activity." Semantic differential procedures have long been based on this principle, but the connection to ratings has not been clear. Recently, some good studies of rating dimensionality have been conducted. (Norman & Goldberg, 1966; d'Andrade, 1965). It could easily be argued that evaluative trait descriptions such as those in course evaluations yield the halo effect merely as an artifact of linguistic structure: adjectives or trait descriptions which show evidence of an evaluative cast tend to function concomitantly in the language. In a sense it might be said that the instructor is placed along a one-dimensional bad-to-good continuum, for reasons that may differ for each student rater, and ratings on evaluative descriptions of the instructor reflect quite generally that very simple relative position.

In summary, it appears that typical course and instructor rating forms are subject to very concrete interpretational difficulties when a standard low-to-high rating scale and fairly short descriptive phrases describing instructor behavior are used. Even when the descriptive phrases describe critical incidents that would not logically be correlated with other behavior, correlations appear in the ratings. This latter phenomena quite possibly has nothing to do with the specific composite of good (or bad) trait qualities that students invest an instructor with, but is rather the result of the instructors "psychological positioning" along a one-dimensional evaluative dimension. While the judgments are reliable ($r = .94$ for average ratings from classes of median size, $N = 28$) and axiomatically valid for student purposes, the reasons for the instructors relative valuation cannot be determined by utilizing the rating items themselves for this purpose.

The discussion of these course evaluation instruments might conclude with a number of suggestions on how to improve the validity of these instruments. As noted earlier, the validity depends upon the purpose for which scores are to be used. From the position of students who wish to know what other students thought of an instructor, the

average ratings may be taken at their face value as valid. Students need not be interested in the precise reasons for each instructor's valuation, much as pollsters need not attempt to describe why opinion on an issue splits along certain lines.

Were the evaluation procedures to become even more specific in a revised instrument there could be a tendency toward phenomena so specific as to be paradoxically irrelevant to student evaluation, i.e., the function of "evaluation" is not to be objective but subjective. Perhaps the most useful suggestion would be to have very few "instructor evaluation" items, since the information in all such items is quite redundant. Other items should deal with other independent aspects of the course which students might be interested in reporting on.

A number of other suggestions for improving the validity of instructor and course evaluations have been made, but each would probably be inadequate in dealing with a halo effect. As has been seen above, a first suggestion of more specificity has paradoxical disadvantages. A second suggestion of having students rate different traits on different occassions would probably show no real differences from traditional instruments unless instructor valuations were time dependent. The prospects for finding such a time dependency do not appear especially bright. A third suggestion follows along the same lines, advising the assignment of subparts of the rating form to random subsets of raters. If subsequent rating items are correlated because of their physical contiguity, then this procedure might prove fruitful. The concept of a single evaluative semantic space dimension mitigates against this possibility, however. A fourth suggestion proposes that student-constructed essay evaluations be content-analyzed via computer, and the frequency of evaluative adjectives and phrases be tabulated. It would appear that semantic spaces quite similar to those found in traditional semantic differential investigations would again be reconstructed.

Although student evaluation of instructor quality is quite unidimensional and <u>in situ</u> apparently resistant to inquiries as to why ratings take on the values they do, several other fruitful avenues of investigation are open. Often rating information is not given alone. At the University of Delaware, students have shown an interest in the problem and have supplied to the investigators a great deal of peripheral information about student raters. Several interesting relationships between instructor and course ratings and student characteristics are presently under investigation.

BIBLIOGRAPHY

Bendig, A. W. "A Factor Analysis of Student Ratings of Psychology Instructors on the Purdue Scale" Journal of Educational Psychology, 1944, 45, 385-93.

Bendig, A. W. "Ability and Personality Characteristics of Introductory Psychology Instructors Rated Competent and Empathetic by their Students" Journal of Educational Research, 1955, 48, 705-709.

Bingham, W. V. "Halo, Invalid and Valid" Journal of Applied Psychology, 1939, 23, 221-28.

Cronbach, L. J. The Essentials of Psychological Testing, Harper & Row, 1970.

d'Andrade, R. G. "Trait Psychology and Componential Analysis" American Anthropologist, 1965, 67, 215-228.

Deshpande, A. S., Webb, S. C., Marks "Student Perceptions of Engineering Instructor Behaviors and their Relationships to the Evaluation of Instructors and Courses" American Educational Research Journal 1970, 7, 289-305.

Ewart, Edwin, Seashore, S. E., and Tiffin, Joseph "A Factor Analysis of an Industrial Merit Rating Scale" Journal of Applied Psychology, 1941, 25, 481-486.

Gage, N. L. Handbook of Research on Teaching, Rand McNally, Chicago, 1965.

Guilford, J. P. Psychometric Methods, McGraw-Hill, 1954.

Guthrie, E. R. "Educational Research and Statistics--Measuring Student Opinion of Teachers" School and Society, 1927, 25, 632, 175-76.

Isaacson, R. L., McKeachie, W. J., and Milholland, J. E., "Correlation of Teacher Personality Variables and Student Ratings" Journal of Educational Psychology, 1963, 54, 110-117.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hoffelar, J., Baerwaldt, J. W., and Zinn, K. L., "Dimensions of Student Evaluations of Teaching" Journal of Educational Psychology, 1964, 55, 344-55.

Norman, W. T., & Goldberg, L. R. "Raters, Ratees, and Randomness in Personality Structure" Journal of Personality and Social Psychology, 1966, 4, 681-691.

Remmers, H. H. "Educational Research and Statistics--Departmental Differences in the Quality of Instruction as seen by Students" School and Society, 1929, 30, No. 767, 332-334.

Remmers, H. H. "Reliability and Halo Effect of High School and College Students' Judgments of their Teachers" Journal of Applied Psychology, 1937, 619-630.

Slobin, D. Y. & Nichols, D. G., "Student Ratings of Teaching" Improving College and University Teaching, 1969, 244-248.

Stockford, L, & Bissel, H. W. "Factors Involved in Establishing a Merit Rating Scale, Personnel, 1949, 26, 94-116.

Symonds, P. M. Diagnosing Personality and Conduct, 1931.

Thorndike, E. L., "A Constant Error in Psychological Ratings" Journal Applied Psychology, 1920, 4, 25-29.

Thorndike, R. and Hagen, E., Measurement Evaluation in Psychology and Education, John Wiley and Sons, 1969.

TABLE I

Rotated Factor Pattern

(N = 150 Class Means)

1968 Student Course Evaluation - University of Delaware

	Rating Items	I	II	III	IV	V	VI	VII
1.	Satisfaction	.868						
2.	Recommendation	.816						
3.	Lecture - Discussion						.800	
4.	Inst. Effective - Ineffec.	-.734						
5.	Easy - Difficult					.842		
6.	Instructor knowledge	.741						
7.	Instructor organization	.837						
8.	Instructor delivery	.868						
9.	No attendance (Fail-Pass)	-.507				-.507		
10.	Instructor help	.530						-.532
11.	Instructor interest	.780					-.348	
12.	Work load				.473	.553		
13.	Value of readings				.758			
14.	Text difficulty						-.732	
15.	Not read text (Fail-Pass)			-.306	-.820			
16.	Reading load			-.515		.500		
17.	No. hour exams		-.830					
18.	No. term papers							-.850
19.	Exam difficulty		-.625					
20.	Objective - essay exams			-.842				
21.	Test fairness		-.461	-.675			.324	
22.	Test grading fairness		-.693					
	Variance Accounted for:	.24	.09	.09	.09	.09	.08	.07

TABLE II

Rotated Factor Pattern

(N = 198 Class Means)

1969 Student Course Evaluation - University of Delaware

	Rating Items	I	II	III	IV
1.	Inst. apparent interest				-.785
2.	Availability of Inst. outside				-.429
3.	Opport. to question in class	.694			-.469
4.	Inst. effect. in moderating	.547			-.735
5.	Inst. organization of course				-.863
6.	Inst. present. and explan.				-.857
7.	Intellectual stimulation	.381			-.773
8.	Inst. respect of student	.567			-.646
9.	Fairness in grading	.351			-.685
10.	Overall eval. of course	.308		.379	-.800
11.	Overall eval. of instructor	.340			-.878
12.	Textbook used			.832	
13.	Value of lecture				-.799
14.	Value of discussion	.523			-.499
15.	Value of assignments		-.360	.610	
16.	Relevance of course			.593	-.449
17.	Material covered		-.771		
18.	Reading difficulty		-.800	-.322	
19.	Exam difficulty		-.713		
20.	Amount of reading load		-.713		
21.	Amount of total work load		-.770		
22.	Amount of conformity	-.802			
23.	Amount of creativity	.760		.216	
	Variance accounted for:	.145	.135	.089	.314

TABLE III

Rotated Factor Pattern*

(N = 127 Randomly Selected Responses)

1970 Student Course Evaluation - University of Delaware

	Rating Items	I	II	III	IV
1.	Explanation of course policies	.826			
2.	Logic of course planning	.856			
3.	Instructor's improvement advice	.719			
4.	Clarity of the textbook		.886		
5.	Relevance of textbook (personally)		.836		
6.	Difficulty of examinations			-.817	
7.	Accuracy of evaluation	.489			
8.	Difficulty of the work load			-.840	
9.	Relaxed atmosphere	.319			-.397
10.	Teaching effectiveness	.774			-.308
11.	Clarity of presentation	.806			
12.	Supplementation of text	.395	.370		-.354
13.	How informed was the instructor	.496			-.539
14.	Encouragement to ask questions	.327			-.729
15.	Courteousness of the instructor				-.798
16.	Encouragement of creativity				-.790
17.	Seeing beyond course limits		.427		-.572
18.	Overall instructor evaluation	.767			-.443
19.	Overall course evaluation	.550	.630		
	Variance accounted for:	27%	13%	9%	17%

*Only the loadings greater than $\pm$ .30 are given

TABLE IV

Course Evaluation "Instructor Impact" Contrasting Items

in 1968, 69 and 70

	1968	Varimax Loading		1969	Varimax Loading		1970	Varimax Loading
				Related Items				
1.	Satisfaction	.868	1.	Eval. of Instr.	.878	1.	Eval. of Instr.	.767
2.	Recommendation	.816	2.	Eval. of course	.800	2.	Eval. of course	.550
3.	Instr. Interest	.780	3.	Instr. Interest	.785	3.	----	----
4.	Instructor Organization	.837	4.	Organization of course	.863	4.	Logic of course planning	.857
5.	Instructor delivery	.868	5.	Presentation of explanations	.857	5.	a) clarity of presentation	.806
							b) effectiveness of teaching methods	.774
6.	Instr. help outside class	.530	6.	Availability outside class	.429	6.	----	----
7.	Instr. knowledge	.741	7.	Intellectual stimulation	.773	7.	How informed was instructor	.496
8.	Value of attendance	.507	8.	Value of lecture	.779	8.	----	----
9.	Instructor effectiveness	.734	9.	Effectiveness of moderating course	.735	9.	----	----
10.	----	----	10.	Fairness in grading	.685	10.	Accuracy of the method of evaluations	.489
11.	----	----	11.	Opportunity to question	.469	11.	Encouragement to ask questions	.327
				Uncorrelated Items				
			1.	Respect for students	.646	1.	Explanation of course policies	.826
			2.	Value of discussion	.499	2.	Instructor's advice to improve	.719
			3.	Relevance of course	.449	3.	General atmosphere in class	.319
						4.	Supplementation of text	.395

TABLE IV - Cont'd

Course Evaluation "Classroom Dialogue/Instructional Procedure"

Contrasting 1968, 69 and 70 Items

1968	Varimax Loading	1969	Varimax Loading	1970	Varimax Loading
		Related Items			
1. Lecture vs. Discussion	.800	1. Opportunity to question	.694	1. Encouragement to ask questions	.729
2. Exam grading fairness	.324	2. Fairness in grading	.351	2. ----	----
3. ----	----	3. Intellect. Stimulation	.381	3. How informed was the instructor	.539
4. ----	----	4. Eval. of Inst.	.340	4. Inst. Eval.	.443
5. ----	----	5. Amt. of Creativity	.760	5. Encouragement for Creativity	.790
		Unrelated Items			
1. Inst. Interest	.348	1. Eff. in moderating	.547	1. General atmosphere	.397
2. Text difficulty	.732	2. Respect for student	.567	2. Effectiveness of teaching	.308
		3. Eval. of course	.308	3. Supplementation of text	.354
		4. Value of discussion	.523	4. Courteousness of inst.	.798
		5. Amt. of conformity	.802	5. Emphasis on seeing beyond the limits of the course	.572

TABLE IV - Cont'd

Course Evaluation "Course Difficulty" Contrasting

1968, 69 and 70

1968	Varimax Loading	1969	Varimax Loading	1970	Varimax Loading
		Related Items			
1. Easy vs. difficult course	.842	1. a) Difficulty of material covered	.771	1. a) ----	----
		b) Difficulty of reading	.800	b) ----	----
		c) Difficulty of exams	.713	c) Difficulty of exams	.817
2. Reading load	.500	2. Reading load	.713	2. ----	----
3. Work load	.553	3. Total work	.773	3. Difficulty of total work load	.840
		Unrelated Items			
1. Necessity of attendance	.503	1. Value of assignments	.360		

TABLE IV - Cont'd

Course Evaluation "Homework/Textbook" Contrasting Items in

1968, 69 and 70

1968	Varimax Loading	1969	Varimax Loading	1970	Varimax Loading
		Related Items			
1. Importance of text	.820	1. Value of text	.832	1. Relevancy of text	.836
2. Value of readings	.758	2. Value of assignments	.610	2. ----	----
3. ----	----	3. Overall Eval. of course	.379	3. Overall Eval. of course	.630
		Unrelated Items			
1. Work load	.473	1. Relevance of course	.593	1. Clarity of text	.886
		2. Reading difficulty	.322	2. Supplementation of text	.370
				3. Emphasis on seeing beyond the limits of course	.427

1970 - 1971

SGA COURSE EVALUATION

STUDENT

SCHEDULE SEQUENCE NUMBER

FACULTY NUMBER

EXPECTED GRADE IN COURSE: A, B, C, D, F, I, L, P/F

CUMULATIVE INDEX: 0 (FRESH TRANS), 0 - .9, 1.0-1.4, 1.5-1.9, 2.0-2.4, 2.5-2.9, 3.0-3.4, 3.5-4.0

SEX: MALE, FEMALE

CLASS STANDING: FRESH, SOPH, JR, SR, EX, OTHER

Why did you select this course?

MAJOR REQUIREMENT | MINOR REQUIREMENT OR RELATED FIELD | REQUIRED ELECTIVE | FREE ELECTIVE | COLLEGE REQUIREMENT | OTHER

If you selected a free elective, college requirement, or other, why?

THOUGHT I COULD MAKE A GOOD GRADE | THOUGHT SUBJECT WOULD BE INTERESTING AND USEFUL | REPUTATION OF INSTRUCTOR | COULD USE P/F OPTION | PREVIOUS S C A EVALUATION | RECOMMENDED BY OTHER STUDENTS

DIRECTIONS

1. Use # 2 PENCIL ONLY.
2. Please fill out the information requested above.
3. Respond frankly to each question.
4. If an item does not apply, leave it blank.
5. If you change your mind, erase your previous answer.

EVALUATE THE FOLLOWING QUESTIONS ON A SCALE OF FROM 1 TO 5, WITH 1 AS THE LOWEST AND 5 AS THE HIGHEST.

1. How clearly did the instructor explain his course policies? 1 2 3 4 5
2. How logically was the course planned and carried out? 1 2 3 4 5
3. How extensive was the instructor's advice on how to study for the course or to improve your work? 1 2 3 4 5
4. How clearly written was the textbook? 1 2 3 4 5
5. How relevant was the textbook to you personally? 1 2 3 4 5
6. How do you rate the difficulty of the examinations? 1 2 3 4 5
7. How accurate a measure of your knowledge was the instructor's method of evaluation (tests, quizzes, papers, etc.)? 1 2 3 4 5
8. How do you rate the difficulty of the work load? 1 2 3 4 5
9. How relaxed was the general atmosphere in the classroom? 1 2 3 4 5
10. How valuable was the discussion section (if applicable)? 1 2 3 4 5
11. How valuable was the lab section (if applicable)? 1 2 3 4 5
12. How do you rate the effectiveness of the teaching method used in this course? 1 2 3 4 5
13. How clear was the instructor's presentation? 1 2 3 4 5
14. How well did the instructor supplement the text from other sources, (including other texts, classroom demonstrations, etc.)? 1 2 3 4 5
15. How well informed was he on materials presented and questions raised? 1 2 3 4 5
16. To what degree did the instructor encourage students to ask questions? 1 2 3 4 5
17. How courteous was the instructor toward different points of view? 1 2 3 4 5
18. To what degree did the instructor foster creativity by encouraging the students to think for themselves? 1 2 3 4 5
19. How much did the instructor emphasize seeing beyond the limits of the course? 1 2 3 4 5
20. To what degree was the instructor available to give individual assistance? (Answer only if you have sought such help.) 1 2 3 4 5
21. Overall how highly would you evaluate this instructor? 1 2 3 4 5
22. Overall how highly would you evaluate this course? 1 2 3 4 5

22

ON THE BACK OF THIS SHEET:
1. List any specific likes and dislikes of this course and/or instructor.
2. List any suggestions you may have for improving this questionnaire.