DOCUMENT RESUME

ED 073 176 TM 002 426

AUTHOR	Centra, John A.
TITLE	The Utility of Student Ratings for Instructional Improvement: I. The Effectiveness of Student Feedback in Modifying College Instruction; II. Self-Ratings of College Teachers: A Comparison with Student Ratings.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Esso Education Foundation.
REPORT NO	ETS-PR-72-16
PUB DATE	Aug 72
NOTE	77p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Analysis of Variance; College Students; College Teachers; Effective Teaching; Multiple Regression Analysis; *Rating Scales; *Self Evaluation; *Student Opinion; Student Teacher Relationship; Teacher Evaluation; *Teacher Improvement; *Teacher Rating; Teacher Response; Teaching Quality; Technical Reports

ABSTRACT

Two studies are reported. In the first, the effectiveness of student ratings in improving college teaching was investigated. Teachers in five diverse colleges were assigned randomly to a feedback (treatment), no-feedback (control), or posttest group. The feedback and no-feedback groups used a student rating from in one of their classes at midsemester and responded to a self-evaluation form. The feedback group received a summary of student responses. Both groups administered the form again at the end of the semester, as did the posttest group. If student feedback improved instruction, end-of-semester ratings of the feedback group should have been better than the other groups. But multivariate analysis of variance results indicated no significant differences among the groups. Regression analyses indicated that student feedback did effect changes in teachers who had rated themselves more favorably than their students had rated them. In the second study, teacher self-ratings were compared to student ratings to investigate the extent to which instructors learn something from students about their teaching. Teachers generally evaluated their teaching somewhat differently from the way their students evaluated it. Comparisons across items between student and faculty responses indicate that instructors are aware of many of their strengths and weaknesses but see themselves more favorably in absolute terms. (Author/KM)

PR-72-16

THE UTILITY OF STUDENT RATINGS FOR INSTRUCTIONAL IMPROVEMENT

John A. Centra

August 1972

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

THE UTILITY OF STUDENT RATINGS

FOR INSTRUCTIONAL IMPROVEMENT

I. The Effectiveness of Student Feedback in Modifying College Instruction

II. Self-ratings of College Teachers: A Comparison with Student Ratings

John A. Centra

August 1972

This study was supported by a grant from the Esso Education Foundation.

Educational Testing Service
Princeton, New Jersey

Contents

Preface

This report is presented in two parts. In the first part, the effectiveness of student ratings in improving college teaching is investigated. An experimental study was conducted at five diverse colleges to investigate the question: To what extent do college teachers alter their instructional practices after receiving feedback from students? The "treatment" in this study purposely paralleled what is typically done at most colleges that use student ratings--that is, the results were seen only by the individual teacher, and the interpretation of the ratings as well as what to do about them were left up to the teacher.

In the second part of this report, teacher self-ratings are compared to student ratings in order to investigate the extent to which instructors learn something from students about their teaching. Because this second part will eventually be published separately descriptions of the instruments, sample, etc. are repeated.

Thanks are due to the teachers and students who participated, and in particular to the staff member at each college who helped coordinate the study. Also, to Charles Hall and Donald Rubin for advice with statistical analyses, to Gordon Davis and Dorothy Thayer for computer programming, and to Marian Helms for unerring attention to the many deadlines and details.

John A. Centra
Research Psychologist

I. THE EFFECTIVENESS OF STUDENT FEEDBACK IN MODIFYING COLLEGE INSTRUCTION

Formal student evaluations of courses and teaching have been receiving a good deal of attention and use at many colleges and universities. The results of these evaluations most often are seen only by instructors and are intended to help improve their teaching. Underlying this intended use is the assumption that the instructors will use the information to alter and improve their teaching. It is an assumption open to question.

Presumably, instructors value student opinion enough to change their teaching behavior when it is evaluated less favorably than the instructors might desire or expect. The theoretical justification for this belief, as developed by Gage, Runkel, and Chatterjee (1963) and by Daw and Gage (1967), may be found in equilibrium theory. Accordingly, when student feedback creates a condition of imbalance (Heider, 1958), asymmetry (Newcomb, 1959), or dissonance (Festinger, 1957) in an instructor, one might expect the instructor to change in the direction desired by students in order to restore a condition of "equilibrium." Following a suitable lapse of time, such changes should be reflected in a second description of teacher behavior.

There is some evidence that student feedback does indeed have a positive effect on teaching performance, although the evidence is far from conclusive, particularly at the college level. Tuckman and Oliver (1968), using 286 teachers of vocational subjects in high school and technical institutes,

found that instructors who received student feedback showed greater "gains" in student ratings, as measured by changes in students' ratings after a 12-week interval, than instructors who received no feedback (actually all the change scores were negative, with positive changes or "gains" being simply less of a negative score). Changes in ratings of teaching were also reported by Bryan (1963), using teachers of academic subjects at the secondary level, and by Gage, Runkel, and Chatterjee (1963), who experimented with sixth-grade teachers.

The results at the college level, however, have thus far been less positive. Miller (1971) reported that end-of-semester student ratings for teaching assistants who had received midsemester feedback did not differ from end-of-semester ratings for teaching assistants who did not receive the feedback. But because of the small and limited nature of the sample (36 teaching assistants assigned to discussion sections in three courses), the results of the Miller study are very tentative.

The preceding studies neglected to include a number of relevant variables that might be expected to be related to changes in teaching. None of the studies investigated the instructor's awareness of his own teaching practices as indicated by self-ratings. On the basis of equilibrium theory, one could hypothesize that the greater the gap between student ratings and faculty self-ratings, the greater the likelihood that there would be change in instruction, since large differences would create the greatest amount of imbalance or dissonance in instructors. None of the preceding studies, furthermore, looked at possible variations in changes across subject areas nor did they investigate the sex of the instructor as still another variable.

Finally, the number of years of teaching experience is a particularly critical variable which was included in only one of the preceding studies. In that study (Tuckman & Oliver, 1968), the expectation that less experienced teachers would be more likely to change was not supported by the results.

The primary purpose of the present study was to investigate in some depth the effects of student feedback on teaching at the college level. Included as variables in the study were the instructor's sex, teaching experience, and self-ratings, as well as the subject area of the course. These instructor or course characteristics were examined to determine their relationship to instructional changes. The study was carried out at several types of post-secondary institutions in order to investigate also the possibility that changes occur at some colleges but not at others.

A second purpose of the study, presented as a second part of this report, was to learn more about teacher self-ratings and, in particular, their relationship with ratings provided by students. To what extent do instructors learn something from students about their teaching? Comparisons between student and teacher ratings over a wide range of instructional practices might underscore the need for student feedback, as well as highlight specific areas of instruction where feedback is most essential.

Method

The Sample of Institutions

Five colleges which did not have a formal program of student ratings of instruction participated in the study. These were among 50 to 60 colleges contacted in the late spring of 1971 with a brief letter describing the

project (see Appendix A). Each of the institutions was asked to return a preprinted postcard with their replies to three questions:

1. Does your institution have a formal program of student ratings of faculty and/or courses (either student conducted or institution sponsored)?

2. In the absence of an institutionalized program, do most of the faculty obtain student ratings of their own?

3. Will your institution be on the semester system next fall?

Only colleges which did not have a formal student rating program, whose faculty generally did not obtain ratings of their own, and which were on a semester system were considered for the study. At these institutions the faculty would not be familiar with the way students viewed their instruction, and the formal feedback might therefore result in changes in teaching practices. Because the initial set of ratings was to be collected at midsemester (and the second set at the end of the semester), it was also important that colleges be on a semester system rather than on the shorter quarter system, thereby enabling a longer period of time for changes to occur. By collecting both sets of ratings within one semester, moreover, the same students and the same courses could be used.

Nine of the colleges that met the above specifications were invited to participate in the study, with the hope that five or six would be able to do so. The five that did accept were located in three states and represented several types of colleges, including two state colleges (one of which had a predominantly black enrollment), a liberal arts college, a multipurpose college,

and an urban community college. This diversity among the particpating colleges provided a wide context of teaching environments.

Procedure

All teaching faculty from four of the five institutions were asked to participate in the first phase of the study. At the fifth institution, which was also the largest, all but 30 members of the faculty were invited to participate in the study. Those 30, chosen at random, were subsequently asked to participate in the study at the end of the second semester.

In only one of the four colleges was the faculty told the full details of the study and, in particular, that student feedback would be purposely withheld from some of them. At the other four colleges the faculty were told that the project was "investigating what students are able to evaluate in the classroom and how useful this information might be to the individual instructor." (See Appendix A for examples of letters; in one instance the writer spoke at a faculty meeting in general terms about the project.)

Faculty members were assured that only they would see their individual rating reports. This assurance undoubtedly contributed to the excellent cooperation from the faculties; in fact, between 70-90% of those at each institution participated in one or more phases of the study.

Teachers within each department of each institution were randomly assigned to one of three groups:

1. the feedback group, which administered a rating form at midsemester and received a summary of results (feedback) within a week; this group also received comparative data based on responses in 75 classes at a sixth institution which had tried out several of the items during the Spring 1971 term (see Appendix B);

2. the no-feedback group, for which student ratings were collected but withheld at midsemester; and

3. the "posttest" group, which used the rating form only at the end of the semester in order to determine whether the midsemester ratings had a sensitizing effect on student raters or teachers.

Each teacher was asked to use the questionnaire in one class of their choice. End-of-semester as well as midsemester ratings were collected for both the feedback and no-feedback groups. Both midsemester and end-of-semester ratings were collected during the fall semester of 1971. A single semester instead of two successive semesters was used initially in order to enable the same students to provide both sets of ratings in the same course. Moreover, a suggestion that has sometimes been made is that instructors should obtain feedback from students at midsemester so that those students who provided the information might benefit.

In Table 1 the number of teachers participating in the study at midsemester and at the end of the semester are listed by college and by group. As might be expected, some of the teachers who used the form at midsemester did not remain in the study for the critical end-of-semester administration. The question then is whether those who dropped out of the study after using the form at midsemester biased the final sample. Were the dropouts, for example, generally the more poorly rated teachers? To examine this question, comparisons were made between three sets of scores: teachers who dropped out vs. those who stayed in for group 1, the same comparisons for group 2, and teachers who stayed in for both group 1 and group 2. Thus in-drop comparisons within the two groups and in-in comparisons across the groups were made for

Table 1

Number of Teachers Participating in the Study at Midsemester and at End of Semester, by College and Group

College	Group 1 Feedback		Group 2 No Feedback		Group 3 Posttest
	Mid-Semester	End of Semester	Mid-Semester	End of Semester	End of Semester Only
1	26	22	25	22	25
2	50	35	55	45	34
3	33	21	32	23	19
4	49	42	52	48	45
5	19	17	24	21	17
Totals	177	137	188	159	140

the 23 items used at midsemester. Out of 69 tests of significance (3 x 23), differences were statistically significant (.05 level) for four of the items. Since differences for that many items could be expected to occur on the basis of chance (at the 5 per cent level of significance), it is safe to conclude that the teachers who dropped out of the study were very much like those who continued (at least in ratings given by students) and, equally important, that the feedback and no-feedback groups were very similar in their student ratings at midsemester.

Instruments

Student ratings or descriptions of instruction were measure at mid-semester by a 23-item "Student Instructional Report" (SIR). Included were items that faculty members in an earlier study had identified as providing information they would like from students.[1] In particular, items that reflected instructional procedures or behavior that teachers presumably could change were used in the study. Among the areas included were course objectives, instructor preparation and organization, student-faculty interaction, student effort, and course difficulty and scope.

The end-of-semester SIR contained the same 23 items; the items were slightly rearranged and presented on a machine-scoreable answer sheet. In addition, it contained several additional items eliciting overall rather than specific ratings; since most of the teachers would be administering the

[1]The questionnaire was developed the previous year with the cooperation of Northwestern University faculty and students. Further information on its development may be found in Centra (1972).

items for the second time in the same course, it was hoped that the additional items would encourage its repeated use. A copy of the midsemester and end-of-semester questionnaires may be found in Appendix C.

Item responses for 19 of the items were on a four-point agree-disagree scale. The remaining four items employed a four- or five-point scale with varying responses. Each instructor received at the appropriate time a summary report that included the mean and standard deviation for each item and the percentage of students in the class that gave each response. As mentioned earlier, the feedback group also received with their midsemester results comparison data from 75 instructors who had administered SIR at the end of the Spring 1971 semester (Appendix B).

Item reliabilities. One way to estimate the reliability of each item is through analysis of variance, a method that is somewhat similar to averaging all possible split-half correlations among student responses. By comparing the variance of responses within classes to the variance between classes, an estimated reliability coefficient may be obtained (see Winer, 1962, pp. 124-132, for additional details). A low reliability estimate, according to this method, would suggest a great deal of variance in responses among students within a typical class, or very little variation in the average responses over a sample of instructors, or both. Ideally it is desirable to have just the opposite--more differentiation between instructors than between the student responses in each class. Through use of the Spearman-Brown formula, estimated reliabilities for classes of various sizes may then be computed.

Analysis of variance reliability estimates for the SIR items were generally above .70 for 20 or more students in a class and slightly less for 15 or more students. These values would seem to be quite acceptable. The data are

presented in Appendix D. The lowest reliabilities may be noted on items dealing with such aspects as the level of difficulty of the course, for which students within a class might more likely be expected to differ.

Test-retest is another frequently employed type of reliability. The correlations between midsemester and end-of-semester responses within the no-feedback and feedback groups separately could be considered as a form of test-retest reliability. Those correlations, as indicated in Appendix E, were moderately high, with the majority of items near or above .70. While a shorter time lag would likely have resulted in higher stability of responses within each group, it is important to keep in mind that good stability with each of the groups does not negate the possibility of feedback effects, i.e., of finding differences between the two groups. Instructors within each group could be ordered or ranked similarly at both midsemester and end-of-semester but the mean difference between the groups could be significant.

Instructor's Form. Instructors in the feedback and no-feedback groups also completed an "Instructor's Form" at midsemester (Appendix F). This form elicited instructor self-ratings on 21 of the 23 SIR items; each of the items was reworded slightly for this purpose.

Instructor's Cover Sheet. Instructors in all three groups also completed a "cover sheet" which contained several questions dealing with the course and the instructor (Appendix G). The number of years of teaching experience and the field in which the subject most appropriately belonged were obtained in this way. The instructor's sex, also a variable of interest, was ascertained from the instructor's name or, when necessary, by inquiring at the college.

Analyses

Multivariate analysis of variance (MANOVA) and regression analyses were the major methods of analyses. In analyzing the possible effects of subject area of the course, sex of the instructor, and number of years of teaching experience, the sample was combined across colleges in order to insure adequate cell frequencies. Differences between the feedback, no-feedback, and posttest groups were also analyzed within each college to examine the possibility that teacher changes may vary by college.

Results and Discussion

MANOVA results appear in Tables 2 through 5. In the first three analyses, presented in Tables 2, 3, and 4, end-of-semester item scores were compared among the feedback (treatment) group, the no-feedback (control) group, and the posttest group. The multivariate F ratios presented in each table provide a test of the differences among the three groups. Also presented in each table are the F ratios for differences in ratings due to other factors (e.g., years of teaching, sex, subject area, etc.) and, most importantly, the interaction or each of these factors with group effects.

Because of the large number of dependent variables (items), the MANOVA analyses were done in two stages. First, the 15 items that were thought to have the best chance of reflecting instructional changes were analyzed. Then, using those 15 items as covariates, the remaining items were analyzed. Because the SIR items are not independent, using the 15 items as covariates served to minimize their effect on the succeeding analysis. The second group

of items consisted of the remaining eight repeated items plus, for two of the analyses, two items from the end-of-semester form dealing with the overall effectiveness of the instructor and the overall value of the course to students. The overall ratings, it was reasoned, might reflect improvement in instruction even though the more specific items did not. Specifically, student feedback could influence teachers to change in ways that improve their overall effectiveness but not in the specific ways indicated by the 23 items.

Years of teaching experience and subject area of the course were included in the first MANOVA analysis (Table 2) because they were presumed to be particularly critical variables. That is, less experienced teachers or those in certain subject areas might more likely be expected to be influenced by student feedback. Teachers were separated into the three categories of experience: those in their first or second year of teaching, those who had been teaching three to six years, and those who had taught for seven or more years. The subject area divisions were natural sciences, social sciences, humanities, and education and applied subjects.

As the results in Table 2 indicate, there were no differences among the three groups nor were there differences in any of the interactions of subject area, years of teaching, and groups. There were, however, differences in the end-of-semester ratings given to teachers in various subject areas (F ratios of 2.40 and 2.41, $p < .001$) and, to a lesser extent, for teachers with varying numbers of years in teaching ($F = 1.91$, $p < .01$). These findings will be discussed later in greater detail.

In light of the differences in ratings related to the subject area of the course, the second MANOVA analysis included that factor once again, along with

Table 2

Summary of Results of Multivariate Analysis of Variance (MANOVA)

for Treatment, Number of Years Teaching, and Subject Area

(N = 435 classes)

For 15 items	df hypothesis	df error	F	p<
Years Teaching	30	752	1.24	.174
Subject Area	45	1117.8	2.40	.001
Groups (Treatment)	30	752	.83	.731
Years Teaching x Subject Area	90	2121.1	1.04	.387
Years Teaching x Groups	60	1469.9	.86	.774
Subject Area x Groups	90	2121.1	.96	.602
Years x Subject x Groups	180	3571.4	1.04	.342
For 10 items, with the first 15 as covariates				
Years Teaching	20	730	1.91	.01
Subject Area	30	1072	2.41	.001
Groups (Treatment)	20	730	.79	.731
Years Teaching x Subject Area	60	1917.4	1.10	.288
Years Teaching x Groups	40	1385.9	.82	.774
Subject Area x Groups	60	1917.4	.93	.626
Years x Subject x Groups	120	2847.5	.87	.841

the sex of the instructor. The results are presented in Table 3. Group differences were again insignificant, as were the two-way and three-way interactions of groups, subject area, and sex. As expected, subject area was highly significant ($F = 4.67$ and 2.68, $p < .001$), as was sex of the instructor ($F = 3.15$, $p < .001$; $F = 2.27$, $p < .01$).

The fourth variable investigated for its possible interaction with treatment effects was the college. As the results in Table 4 indicate, feedback did not result in significant instructional changes at any of the five colleges, although the faculty ratings across the colleges did differ significantly ($F = 3.63$ and 3.49, $p < .001$; college differences are discussed later in this section, as well as in the second section of this report).

The final MANOVA analysis of the first-semester data employed the midsemester ratings as covariates. Differences in end-of-semester ratings, therefore, were compared for the two groups (feedback and no-feedback) for which midsemester ratings were available. Colleges and subject areas were also included in this analysis. The results, which appear in Table 5, indicate clearly that treatment and interaction effects were once again not significant. In other words, the feedback group did not differ from the no-feedback group even after controlling for midsemester ratings.

To summarize the findings to this point, end-of-semester ratings of instructors who were given midsemester feedback did not differ, as hypothesized, from either the no-feedback or the posttest groups. Moreover, teacher ratings for the three groups did not differ when subject area, sex of instructor, college, or amount of teaching experience were taken into account. In short, subgroups of teachers identified by the above categories did not appear to

Table 3

Summary of MANOVA Results for Treatment, Subject Area,

and Sex of the Instructor

(N = 435 classes)

For 15 items	df hypothesis	df error	F	p <
Sex	15	388	3.15	.001
Subject Area	45	1153.4	4.67	.001
Groups (Treatment)	30	776	1.07	.368
Sex x Subject Area	45	1153	.66	.958
Sex x Groups	30	776	.58	.97
Subject Area x Groups	90	2188.6	1.01	.45
Sex x Subject x Groups	90	2188.6	1.05	.35
For 10 items, with the first 15 as covariates				
Sex	10	386	2.27	.01
Subject Area	30	1133.7	2.68	.001
Groups (Treatment)	20	772	.77	.757
Sex x Subject Area	30	1133.7	.98	.497
Sex x Groups	20	772	1.24	.205
Subject Area x Groups	60	2027.4	.91	.664
Sex x Subject x Groups	60	2027.4	1.19	.154

Table 4

Summary of MANOVA Results for Colleges and Treatment
(N = 435 classes)

For 15 items	df hypothesis	df	F	p <
Colleges	60	1587.0	3.63	.001
Groups (Treatment)	30	812	1.02	.440
Colleges x Groups	120	2902.8	1.02	.428
For 8 items, with the first 15 as covariates				
Colleges	32	1469.3	3.49	.001
Groups (Treatment)	16	796	1.21	.254
Colleges x Groups	64	2302	.86	.781

Table 5

Summary of MANOVA Results for Treatment, Colleges, and Subject Areas Using Midsemester Scores as Covariates (N = 294 classes)

For 15 items	df hypothesis	df error	F	p<
Groups (Treatment)	15	227	1.07	.390
Groups x Colleges	60	888.3	.98	.518
Groups x Subject Area	45	675.1	1.12	.285

change significantly after receiving feedback. Instructors in their first or second year of teaching were considered most likely to change since they generally had not yet established rigid teaching habits; but student feedback did not result in changes for even this less experienced group.

While so-called treatment effects (i.e., those related to the feedback conditions) were not significant, there were a number of differences that are worthwhile noting. As Tables 2-5 indicate, ratings varied considerably by subject area, sex of instructor, number of years in teaching, and college. The specific ways in which student ratings differed for each of these categories are outlined below.

Subject area. Courses in the natural sciences, relative to those in humanities, social sciences, and education and applied areas, were seen by students as having a faster pace, as being more difficult, and as being less likely to stimulate student interest. In addition, students perceived natural science teachers as less open to other viewpoints.

Humanities teachers also had a few shortcomings according to the student ratings: in comparison to teachers in the other three subject fields, they were less likely to inform students of how they were to be evaluated, and there was less agreement between announced objectives and what was actually taught in their courses.

Generally speaking, students rated courses or teachers in the social sciences and those in education and applied areas similarly. (The item mean scores for each of the subject areas are listed in Appendix H.)

Instructor's sex. Students rated women teachers more favorably than men teachers on several of the items, particularly those dealing with what is

frequently referred to as "teacher-student interaction." Women teachers, students indicated, were more likely to know when students didn't know the material, were more concerned with student progress, made more comments on papers or exams, and generally made better use of class time. Courses taught by male teachers, on the other hand, were more stimulating to students, although they were also viewed as more difficult.

Number of years of teaching experience. There was only one way in which teachers with varying numbers of years of teaching experience differed in their ratings: those with more than six years of experience were less likely to inform students of how they would be evaluated in the course. This is somewhat surprising. The general expectation is that teachers improve with experience. However, the more experienced teachers in this study were rated no better than those in their first or second years of teaching; in fact, on this one item they were rated less favorably.

College differences. The major difference in the student ratings among the five colleges was College Five's poor ratings. On a third of the items, students rated instructors at College Five less favorably than at the remaining four institutions. In particular, students at College Five thought that instructors were not concerned enough about students (items 5 and 8) nor did they do enough to challenge students (item 10). The differences will be discussed more fully in the second part of this report, where they will be related to instructor self-ratings. The mean item scores for each of the colleges are listed in Appendix I.

In addition to the differences noted for College Five, there were also some minor variations among the colleges. All of these variations would suggest very different learning (or teaching) environments among the institutions, or quite possibly different levels of expectations among students. Further studies are needed to explain why ratings differ among colleges and to investigate what student ratings might contribute toward a better understanding of teaching climates at particular colleges.

Student Feedback and Self-ratings of Teachers

A major hypothesis of this study was that changes in instruction would be related to instructor self-ratings. Specifically, the expectation was that student feedback would lead to improved instruction for those teachers who had rated themselves much better than their students had rated them. The relationship, moreover, was predicted to be linear: the greater the discrepancy, the greater the likelihood of improvement. To test this hypothesis, the following regression equation was employed with the feedback and no-feedback groups:

$$R_2 = a_1 + b_1 R_1 + c\ (I - R_1) \quad ,$$

where R_2 is the predicted second-semester rating, R_1 is the midsemester rating, I is the teacher self-rating (thus $I - R_1$ is the difference between the instructor self-rating and the midsemester rating), and the a's, b's, and c's are the regression weights. If the hypothesis is supported, there should be a significant difference between the regression weights for $I - R_1$ (i.e., c) for the feedback and no-feedback groups, with c for the feedback group being positive and greater.

For these analyses, instructors in both groups were divided into those who rated themselves more favorably and those who rated themselves less favorably than their students rated them on each item. Of particular interest were teachers who rated themselves more favorably, since the prediction was that student feedback would effect changes only for those teachers. For most of the items, about 60-65 per cent of the sample had rated themselves more favorably; also, for this group the size of the discrepancy between self and student ratings was much greater than for the group that rated themselves less favorably. (A further analysis and discussion of instructor-student discrepancies in ratings is given in the second part of this report.)

Results of the regression analyses for 17 of the 19 agree-disagree items appear in Table 6 (instructors did not respond to items 10 and 20 because they were not appropriate as self-rating items; also, the first four items were not scored appropriately for this analysis). Listed are regression weights c and t test results of the difference in c between the feedback and no-feedback groups. These results are presented for instructors who had rated themselves less favorably (left side of the table) and more favorably (right side) than their students rated them. Results for instructors who rated themselves less favorably indicate fairly random differences in c , and on only one item did the feedback and no-feedback groups differ.

But differences in c for teachers who rated themselves more favorably were significant ($p < .05$) for 5 of the 17 items, as indicated in the last column. Equally important is the fact that for 13 of the 17 items, the direction of the differences also supports the hypothesis. That is, for those items the c 's for the feedback group (group 1 in the table) are higher than those for the no-feedback group (group 2).

Table 6

Summary of Results of Regression Analyses[a]

Item[b]	Group[c]	For instructors who rated themselves less favorably[d] Regression Weight c	t difference[f]	For instructors who rated themselves more favorably[e] Regression Weight c	t difference[f]
5	1	-.077	-.407	.189	1.23
	2	-.009		.045	
6	1	.120	-.0356	.026	-1.24
	2	.129		.167	
7	1	-.234	-1.97*	.310	1.11
	2	.137		.168	
8	1	.079	.248	.094	.738
	2	.049		-.033	
9	1	.007	-.182	.010	-.232
	2	.026		.033	
11	1	.202	.443	.142	2.466*
	2	.083		-.077	
12	1	see footnote b		.167	-.368
	2			.217	
13	1	.097	.451	-.043	.484
	2	-.001		-.075	
14	1	.117	-.411	.015	1.44
	2	.176		-.070	
15	1	.228	-.067	.235	2.196*
	2	.254		-.150	
16	1	.077	-.363	-.120	-.826
	2	.130		.049	
17	1	-.118	-.625	.104	.883
	2	-.002		.003	
18	1	.079	.607	.172	2.062*
	2	.015		-.031	
19	1	-.042	-1.337	.130	3.932*
	2	.132		-.155	
21	1	-.005	-1.174	.111	2.618*
	2	.087		-.093	
22	1	-.035	.831	.124	1.21
	2	-.240		.004	
23	1	-.098	-1.225	.190	1.45
	2	.061		.002	

*Significant at the .05 level.

[a]For the formula: $R_2 = a_1 + b_1 R_1 + c (I - R_1)$

[b]Item numbers refer to the midsemester form. Instructors did not respond to numbers 10 and 20. For item 12, all instructor responses for one of the groups were identical; thus c could not be computed.

[c]1 = Feedback group, 2 = No-feedback group.

[d]Less favorably defined as $I - R_1$ = greater than 0.

[e]More favorably defined as $I - R_1$ = less than 0.

[f]Test of the difference in regression weights c for $(I - R_1)$ in groups 1 and 2.

The five items for which there were significant differences in regression weights c for instructors who rated themselves more favorably were:

(11) The instructor encourages students to think for themselves

(15) In this class I feel free to ask questions or express my opinions

(18) The instructor summarizes or emphasizes major points in lectures or discussions

(19) My interest in the subject area is being stimulated by this course

(21) I have been putting a good deal of effort into this course.

For these items in particular, then, the findings suggest that instructors changed after receiving feedback if their own ratings were especially better than their students' ratings--that is, if instructors were "unrealistic" in how they viewed their teaching. Even with only a half semester left, instructors apparently were able to stimulate student interest, effort, and thinking; to become more open to questions or opinions; and to do a better job of summarizing or emphasizing major points.

Changes In Instruction Following A Longer Time Period

Thus far, the results indicate that changes in instruction at the end of a single semester were rather limited. Only those teachers who rated themselves much better than did their students appear to have changed after receiving mid-semester feedback. It may be, however, that more teachers do in fact change, but not until the following semester. Perhaps teachers need more time to think about and develop new practices, and perhaps they find changes easier to make with the start of a new course.

To investigate this possibility, additional data were collected at the end of the Spring semester at one of the five colleges. The particular college

was one at which 30 teachers had been randomly selected to use the SIR rating form at the later time rather than during the Fall semester. In short, they were to be the Spring control group. In addition to this group, teachers in the Fall feedback and posttest groups at this same college were asked: (1) if, during the Spring semester, they were teaching the same course in which they used the SIR form in during the Fall, and (2) if they would be willing to administer the form in that course at the end of the Spring semester (they were surveyed about one month prior to the end of the semester). Eight teachers from the feedback group and 13 from the posttest group responded affirmatively to both questions and did administer the form once again at the end of the Spring semester. Although a larger sample would have been desirable, the teacher ratings were mean scores and thus more reliable than individual scores. In sum, the eight teachers in the feedback group were using the form for the third time, while the posttest group, which had administered the form only at the end of the Fall semester, were using it for the second time. Of the 30 instructors asked to use the SIR form for the first time, 21 were able to do so. A summary of the Spring groups, including the number in each group and the specific times when SIR had been used by the groups, is given in Figure 1.

	Time of Administrations		
Group	Midsemester, Fall	End of Semester, Fall	End of Semester, Spring
Fall Feedback Group, N = 8	X	X	X
Fall Posttest Group, N = 13		X	X
Spring Group, N = 21			X

Figure 1

Time of Administrations of SIR for each of the Three Spring Groups

While the multivariate analysis of variance of the Fall data did not reveal significant differences between the feedback and comparison groups, there were eight items for which the univariate F values for one or more of the analyses had approached significance (i.e., p between .05 and .20). On the basis of this prior finding, and because they would appear to be most sensitive to change, these eight items rather than the entire set of 23 were selected for further analysis with the Spring groups. To test the differences among the three groups on these eight items at the end of the Spring semester, multivariate analysis of variance was used once again.

The MANOVA results clearly indicated that the groups differed ($F = 2.18$; $df = 16,264$; $p < .015$). An inspection of the univariate F values and the means for each group, presented in Table 7, further indicated that for most of the items, the feedback groups received more favorable scores than either of the other two groups (with the exception of item 16, lower scores are more favorable).

In particular the feedback group differed from the posttest and Spring-only groups in that they more likely summarized or emphasized major points in lectures or discussions (item 14), were better prepared for class (12), were more likely to accomplish course objectives (20), made better use of class time (3), and finally, according to their students, more likely made helpful comments on papers or exams (9). On the only other item on which the groups differed, the Spring-only group received the most favorable score; that was on item 19, "the instructor was open to other viewpoints."

These results suggest then that student feedback did effect some changes in instruction over time, in that teachers who had received feedback twice during the previous semester did receive better ratings than instructors who had received feedback once or not at all. Before jumping to that conclusion,

Table 7

Summary of Univariate F Test Results and

Means for the Spring Data, Eight Items

Item	$p <$	Means for[a] Fall Feedback Group	Fall Posttest Group	Spring Administration Group Only
		N=8	N=13	N=21
3. Instructor used class time well	.07	1.45	1.80	1.85
5. Inst. knew when students didn't understand material	.91	2.04	2.05	2.00
9. Inst. made helpful comments on papers or exams	.07	1.79	2.27	2.06
12. Inst. was well-prepared for each class	.03	1.21	1.53	1.56
14. Inst. summarized or emphasized major points in lectures or discussions	.01	1.40	1.75	1.71
16. Scope of the course is too limited	.61	3.31	3.25	3.22
19. Inst. was open to other viewpoints	.05	1.81	1.98	1.71
20. Inst. has accomplished his objectives for the course	.05	1.51	1.82	1.74

[a]Lower score indicates agreement, except for item 16.

however, some alternative explanations need to be considered. Perhaps teachers in the feedback group who chose tc readminister the items again in the Spring were better to begin with. Indeed, they may have chosen to readminister the rating items because they had already received favorable ratings and sought further reinforcement for their efforts. This potential bias, to some extent, was minimized because only courses that were taught in both the Fall and Spring semesters were included in the Spring feedback and posttest groups. Yet the possibility of prior differences existed, and to investigate that possibility, differences in scores on each of the eight selected items at the end of the Fall were tested for three pairs of groups:

(1) the 8 teachers from the feedback group vs. 35 from the same group who did not participate in the Spring;

(2) the 13 teachers from the posttest group vs. 32 from the same group who did not participate in the Spring; and

(3) the 8 teachers in the feedback group vs. the 13 in the posttest group.

If none of these differed significantly in the Fall, then a strong case could be made for the Spring differences being due to student feedback. MANOVA tests for the three pairs of comparisons did not in fact yield significant differences. End-of-fall ratings for the eight teachers from the feedback group, therefore, were not unlike their colleagues' ratings in that group, nor were they significantly different from the 13 in the posttest group; likewise, end-of-fall ratings for the 13 in the posttest group were similar to the other faculty ratings from that sample. Differences at the end of the Spring term, consequently, were less likely to be due to prior differences or self-selection. It would seem safe to conclude that student feedback did

effect changes in the feedback group, and that these changes were reflected in certain Spring ratings.

But what about the Fall posttest group? Because their Spring ratings were very similar to the group that used the form for the first time, it would appear that the single feedback had little effect in changing instruction. There are several possible explanations for this lack of change. The posttest group, unlike the feedback group, had not received any comparative or "normative" information to help them interpret their scores. It may be, therefore, that the lack of interpretive information did not enable those instructors to understand fully their ratings (particularly since student ratings are typically skewed in a positive direction); consequently, they may not have thought they needed to change. Or it may be that again not enough time had lapsed for changes to be made (one semester vs. a semester and a half for the feedback group). Or finally, perhaps at least two sets of student ratings are needed before teachers see a pattern of weaknesses that they might improve.

Summary and Conclusions

The purpose of this study was to investigate the effects of feedback from student ratings on changing instructional practices at the college level. Teachers within each of five diverse colleges were assigned randomly to a feedback (treatment), no-feedback (control), or posttest group. The feedback and no-feedback groups used a 23-item student rating form in one of their classes at midsemester during the Fall 1971, and also responded to an "Instructor's Form" with self-ratings of instruction. A summary of their students'

responses along with some comparative information to aid in interpretation were sent back to each instructor in the feedback group within a week, while results were withheld from the no-feedback group. Both groups used the rating form (the Student Instructional Report) in the same class again at the end of the semester; the posttest group, which was using it for the first time, also administered the form in one class at the end of the Fall semester.

If student feedback improved instruction, end-of-semester ratings of the feedback group should have been better than either the no-feedback or the posttest groups. Multivariate analysis of variance results for the end-of-semester ratings, however, indicated no significant differences among the three groups. Furthermore, no differences were noted when various interactions were investigated. Specifically, regardless of the particular college, subject area of the course, sex of the instructor, or the number of years the instructor had taught, the feedback did not appear to produce a difference. (There were, however, some noteworthy differences in student ratings of instructors by college, subject area, sex, and years of experience.)

But a major hypothesis of this study was that student feedback would effect changes in teachers who had rated themselves more favorably than their students had rated them. Results of the regression analyses indicated this to be the case; the findings suggested that the greater the discrepancy--where the discrepancy reflected the extent to which students rated teachers less favorably than the teachers apparently expected--the greater the likelihood of change. On the basis of equilibrium theory, teachers were predicted to change because they valued student opinion and because students had evaluated them less favorably than they had expected. In order to restore a condition

of "equilibrium," teachers who received such feedback changed in the direction suggested by the student ratings.

It would appear, then, that student ratings can effect changes in teaching practices even after only a half-semester. Those changes, however, were evidenced only for those teachers who had unrealistically high (compared to their students' views) opinions of their instructional practices. Others may have rated themselves average or poorly, just as their own students rated them, and did not change even though there was room for improvement.

But would more teachers change if they had more time--that is, if more than a half-semester had lapsed? It appeared that they would, especially if they are provided with information to help them interpret their own results. A segment of the teachers in the Fall sample used the rating form again in the Spring. The feedback group, which had at that point administered the form for the third time, scored better on several items than a random group of instructors who were using it for the first time, or the so-called Fall post-test group, which had used it once before but did not receive any comparative information to aid in interpretation. The feedback group, therefore, not only had the benefit of more time to change (a semester and a half) but they also were the only group that had been given interpretative information. This information consisted of simply the mean scores based on some 80 teachers who had used the form during the previous year.

The need for comparative or "normative" data to help the teacher understand better his or her students' rating is, therefore, underscored by this study. Comparisons with other teachers at their institution, with instructors from similar colleges, or with a national sample of teachers in the same

subject area would undoubtedly help teachers recognize strengths and weaknesses otherwise overlooked. Other interpretative mechanisms, such as providing each teacher with a narrative summary of their student ratings along with specific suggestions for change, or providing individual or group counseling to teachers with poorer ratings, could also add to the understanding and use of student rating results.

This study was designed to show what student feedback coupled with some rather minimal comparative information could do toward changing instruction. The "treatment" in this study purposely paralleled what is typically done with student ratings at most colleges: the results were seen only by the instructor and, while some comparative information was provided, interpretation of the results as well as what to do about them was left up to the instructor. Some might view the changes that occurred in this study as modest. But in view of the ease with which student ratings can be employed for instructor self-improvement--indeed the extent to which they are already in use among colleges and universities-- they appear to have had a sufficient impact to warrant continued use as one method of improving college teaching.

References

Bryan, R. C. Reactions to teachers by students, parents and administrators. (United States Office of Education, Cooperative Research Project No. 668.) Kalamazoo: Western Michigan University, 1963.

Centra, J. A. The Student Instructional Report: Its development and uses. SIR Report Number 1. Princeton, N. J.: Educational Testing Service, 1972.

Daw, R. W., & Gage, N. L. Effect of feedback from teachers to principals. Journal of Educational Psychology, 1967, 58, 181-188.

Festinger, L. A theory of cognitive dissonance. Evanston, Ill.: Row, Peterson, 1957.

Gage, N. L., Runkel, P. J., & Chatterjee, B. B. Changing teacher behavior through feedback from pupils: An application of equilibrium theory. In W. W. Charters & N. L. Gage (Eds.), Readings in the social psychology of education. Boston: Allyn & Bacon, 1963. Pp. 173-181.

Heider, F. The psychology of interpersonal relationships. New York: Wiley, 1958.

Miller, M. T. Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 1971, 62, 235-239.

Newcomb, T. M. Individual systems of orientation. In S. Koch (Ed.), Psychology: A study of a science. Vol. 3. New York: McGraw-Hill, 1959. Pp. 384-422.

Tuckman, B. W., & Oliver, W. F. Effectiveness of feedback to teachers as a function of source Journal of Educational Psychology, 1968, 59 (4), 297-301.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

II. SELF-RATINGS OF COLLEGE TEACHERS: A COMPARISON WITH STUDENT RATINGS

Teacher self-ratings have been proposed as a possible source of information for performance improvement and, to a lesser extent, as an input into performance evaluation. As a basis for decisions on promotion or salary, self-evaluations are not likely to have much validity. But it is possible that some form of systematic self-evaluation could be helpful to the teacher trying to improve instruction, particularly if combined with external evaluations provided by students or colleagues.

There has been little research on teacher self-ratings. In particular, the relationship between self-ratings and those provided by students or colleagues is not yet fully known. With 51 instructors in a military setting, Webb and Nolan (1955) reported a correlation of .62 between instructor self-ratings and student ratings. Clark and Blackburn (1971), however, reported a correlation of .19 between student ratings and faculty self-ratings at a small college, and a similarly moderate correlation (.28) between self-ratings and colleague ratings. In both of these studies, overall teaching was rated rather than specific instructional practices.

The purpose of this study was to further investigate college teachers' self-ratings and ratings given by students by comparing these two sets of ratings over a wide range of specific, student-oriented instructional practices. Discrepancies between self-ratings (or self-descriptions) and those provided by students would underscore the need for student feedback to the instructor as well as highlight specific areas of instruction where feedback is most essential. Differences in ratings will also be studied to investigate their relationships to selected teacher and course characteristics.

Procedure

The sample for the study consisted of 343 teaching faculty at five institutions of higher education. Between 75 to 90 per cent of the teachers invited from each college participated in the study. The five institutions included two state colleges (one of which had a predominantly black enrollment), a selective liberal arts college, a multipurpose college, and an urban community college. None of these institutions had, at the time of the study, a systematic program to collect student ratings, nor did a significant portion of their faculty collect student ratings on their own. The majority of teachers in this study, therefore, were not familiar with how students might rate their instruction.

Students and teachers responded to 21 items dealing with instructional practices. The student questionnaire was titled the "Midsemester Student Instructional Report" and actually contained 23 items, 21 of which were judged appropriate for instructor self-ratings. Included were items that faculty members in an earlier study had identified as providing information they would like to receive from students (Centra, 1972). Among the dimensions of instruction included were the organization of the course, student-teacher interaction, instructor communication, student effort, and stimulation of students. Previous factor analytic studies had identified several of these as dimensions that effectively differentiated among instructors (Coffman, 1954; Gibb, 1955; Hodgson, 1958; Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwalt, & Zinn, 1964).

Responses to 17 of the items were on a four-point agree-disagree scale, with a "not applicable" option also provided. The four remaining items used a four- or five-point scale with different response options for each

item. The wording for each of the statements in the questionnaire differed slightly for students and instructors. For example, an item on course objectives was worded as follows for each group:

For students: The instructor's objectives have been made clear

For teachers: I feel my objectives for the course have been made clear to students

Teachers were asked to "describe this course, your teaching, or the students enrolled." They were told that the reason for obtaining this self-report was to see which items were tapping information already known to most instructors.

The data were collected at midsemester of the Fall 1971 term. Instructors administered the rating form in one class of their own choosing, with the understanding that only they would receive a summary of their students' responses.

Analyses

Faculty-student comparisons were made in a number of ways. First, the relationship between the two sets of ratings was studied by correlating instructor responses to each of the 21 items with the mean responses of students in their class (N = 343 classes). Secondly, differences between the way faculty as a group and students as a group rated or described instruction were investigated by a comparison of means; i.e., the mean score for all teachers on each item was compared to the average of the student class means.

Finally, the discrepancy between each instructor's response and the mean response of his class was of particular interest. The extent of that discrepancy and its relationship with specific teacher or course variables

(i.e., sex, years of teaching experience, subject area of the course) were analyzed through multivariate analysis of variance.

Results and Discussion

The results of the comparison of means and the correlational analysis for items 5-21 are presented in Table 1. The correlation between the two sets of descriptions or ratings was not particularly high, indicating only modest agreement in the way faculty and students perceived instruction. While the correlation between faculty and student responses was significantly different from zero for most of the items due to the large N (343), the median correlation was only .21.

Also listed in Table 1 are the mean faculty responses for each item and a ranking of the items, the mean of the classroom (student) means and a ranking of those scores, the results of the t-tests, and the number of colleges where the difference between the means was significant. A graphical presentation of the data is presented in Figure 1. Responses for items 5-21 could range from one for "strongly agree" to four for "strongly disagree"; thus, lower values represent greater agreement with each statement. The comparisons of the mean values indicate that instructors as a group generally rated or described their teaching more favorably than did their students.(Students' ratings were also skewed toward the more favorable end of the scale, which is usually the case with this type of instrument.) In particular, instructors and students did not agree on the following items: the extent to which students are free to ask questions or give opinions in class (item 14), the extent to which instructors are concerned with student learning (11), the amount of agreement between

Table 1

Faculty-Student Comparisons to Instructional Report Questionnaire,

343 Classes at Five Colleges[a]

		Mean Faculty Response[b]	Mean of Student Means[b]	T Test of Means	Number of Colleges Item Was Significant	Correlations[c]
5	Course objectives made clear	1.55 (9)	1.81 (10)	7.52*	4	.25
6	Agreement between objective and teaching	1.51 (7)	1.82 (11)	9.34*	5	.19
7	Instr. using class time well	1.56 (10.5)	1.72 (7.5)	4.68*	1	.11
8	Instr. availability for students	1.47 (5.5)	1.62 (2)	4.32*	2	.28
9	Instr. knows when students don't understand	1.77 (15)	1.98 (15)	5.43*	3	.21
10	Instr. encourages students to think	1.42 (3.5)	1.71 (6)	8.87*	4	.23
11	Instr. concern with student learning	1.26 (2)	1.68 (4)	13.54*	5	.17
12	Instr. comments helpfully on papers or exams	1.75 (14)	2.03 (17)	5.74*	2	.33
13	Instr. raises challenging questions	1.68 (12)	1.90 (13)	5.34*	3	.22
14	Students are free to question or give opinions	1.15 (1)	1.67 (3)	18.93*	5	.16
15	Instr. preparation for each class	1.47 (5.5)	1.52 (1)	1.59	0	.13
16	Instr. informs students of how evaluated	1.52 (8)	1.84 (12)	8.03*	5	.42
17	Instr. summarizes or emphasizes major points	1.56 (10.5)	1.73 (9)	4.34*	2	.13
18	Student interest stimulated by course	1.85 (16)	2.01 (16)	3.90*	1	.32
19	Students putting effort into course	2.09 (17)	1.97 (14)	-3.10*	0	.33
20	Instr. openness to other viewpoints	1.42 (3.5)	1.72 (7.5)	8.74*	5	.16
21	Instr. accomplishing objectives for the course	1.70 (13)	1.69 (5)	- .19	0	.15

*Significant at .01 level.

[a]The N for each item was often less than 343 due to "Not Applicable" instructor responses; i.e., they did not think the item applied to their course. Lower responses indicate greater agreement or more favorable responses.

[b]Rank of each item mean is in parentheses. The rank correlation equals .77.

[c]Correlations between faculty member responses to each item and the mean of student responses in his class. For an N of 343, r of .14 is significant at the .01 level.

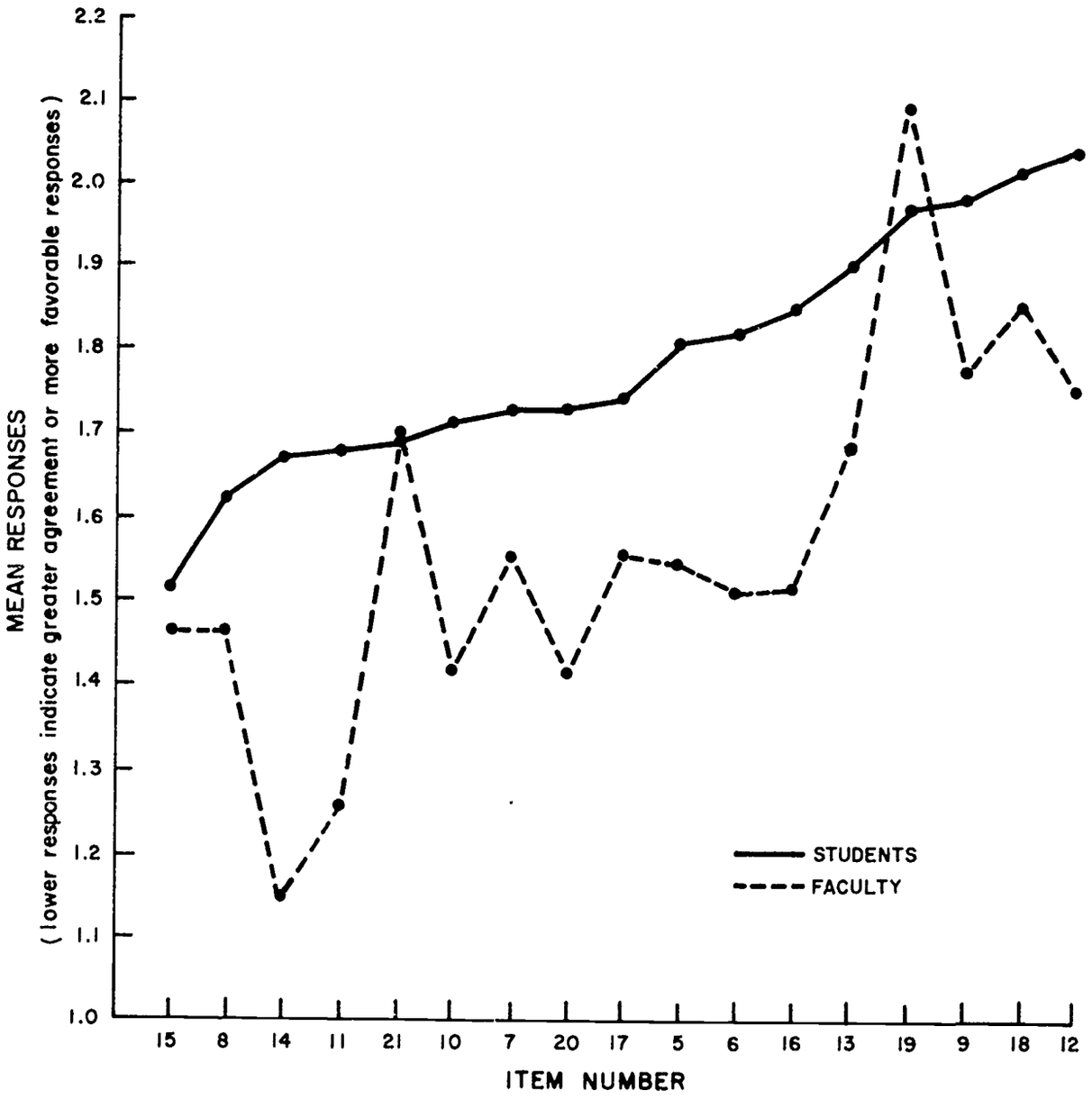

Fig. 1. Faculty and student mean responses to items in instructional report.

objectives and what is being taught (6), instructor openness to other viewpoints (20), the extent to which instructors inform students of how they would be evaluated (16), whether the instructor encourages students to think for themselves (10), and the clarity of course objectives (5). For each of these seven items, instructor-student differences were notable at either four or all five of the colleges.

On the other hand, there was little difference between the faculty and student groups in their ratings of the instructor preparation for class (15) and on the extent to which course objectives were being accomplished (21). For the remaining eight items, the differences were modest and in many instances not significant within a college.

Another way to look at the data is to compare items with each other. The question then becomes: To what extent do the groups of teachers and students order the items similarly? A ranking of item means for each of the two groups indicates fairly high similarity; in fact, a rank correlation of .77. This would suggest that, while teachers and students are generally using different points on the scale in responding to the items (as indicated by the comparison of means), both groups tend to see the same relative strengths and weaknesses among the teachers in this study. For example, while there is a large mean difference between the groups on instructor concern with student lea ng (item 11), both groups rated instructors favorably on this item in comparison to other aspects of teaching. Keeping in mind that higher scores represent unfavorable (disagree) responses, both groups also rated the instructors in this study poorly on stimulating student interest in the course (18).

Generally speaking, combining the ranks of both teachers and students indicates that not stimulating student interest enough (16), the lack of helpful comments on papers or exams (12), and not knowing when students understand the material tended to be rated as the most frequent criticisms of instruction for the teachers in this study. On the other hand, their strengths were in allowing students to feel free to ask questions or give opinions (14) and in their concern with student learning (11).

Individual Teacher-Class Differences

Probably more important than a comparison of the way an average instructor and an average class rated instruction is some knowledge of how many instructors perceived themselves far differently than their students did. A distribution of the differences between each instructor's responses and those of his class (i.e., the class means) provides that information. Presented in Table 2 is a summary of the results of such a distribution. For each item, the percentage of instructors who gave themselves "considerably poorer" or "considerably better" ratings is indicated within each college and for the total sample. A difference of .63 or greater was used to define "considerably poorer or better" because a difference of at least that great would appear to be large enough to have some practical significance; it is also the approximate standard deviation for most of the student item responses.

For most of the items, between a fourth and a third of the instructors described or rated themselves considerably better than their students did. The median, in fact, was just under 30 per cent for all 343 instructors and their classes. Forty-one per cent of the instructors gave themselves

Table 2

Results of the Distribution of Differences between Faculty-Student Responses

to the Instructional Report Questionnaire

	Percentage of instructors who gave themselves:											
	Considerably poorer ratings than the mean of students in their class[a]						Considerably better ratings than the mean of students in their class[a]					
	College						College					
	1 N=50	2 N=99	3 N=66	4 N=95	5 N=37	Total N=343	1 N=50	2 N=99	3 N=66	4 N=95	5 N=37	Total N=343
5 Course objectives made clear	6	4	5	2	1	4	28	25	23	31	46	30
6 Agreement between objectives and teaching	2	9	0	2	0	4	44	31	38	36	32	37
7 Instr. using class time well	4	12	5	7	11	8	16	18	15	26	30	21
8 Instr. availability for students	8	9	8	4	30	10	32	24	15	21	22	23
9 Instr. knows when students don't understand	4	8	6	5	14	7	34	21	37	22	28	28
10 Instr. encourages students to think	2	5	5	11	14	4	34	23	14	35	43	29
11 Instr. concern with student learning	2	9	6	1	6	5	36	24	33	45	33	36
12 Instr. comments helpfully on papers or exams	4	8	10	5	7	7	30	28	18	31	41	31
13 Instr. raises challenging questions	4	7	5	10	7	7	28	26	14	23	33	24
14 Students are free to question or give opinions	0	3	0	2	0	2	38	36	42	41	47	41
15 Instr. preparation for each class	6	14	14	19	16	15	20	14	21	20	16	16
16 Instr. informs students of how evaluated	2	10	5	5	0	6	28	28	32	30	42	32
17 Instr. summarizes or emphasizes major points	2	14	7	7	11	9	18	18	33	28	41	28
18 Student interest stimulated by course	8	6	8	7	9	8	20	18	23	16	38	21
19 Students putting effort into course	24	12	19	19	14	18	12	9	8	8	14	10
20 Instr. openness to other viewpoints	4	6	5	3	5	5	30	25	37	33	38	32
21 Instr. accomplishing objectives for the course	4	14	16	9	10	12	10	14	16	16	34	17

[a]A difference of .63 or greater (on a four-point scale) was defined as a "considerably poorer" rating (i.e., tend to disagree with item) or a "considerably better" (i.e., tend to agree with item) rating.

better ratings on item 14: students are free to ask questions or give opinions in class; and 36 per cent on item 11: the instructor is concerned about whether students learn and tries to be actively helpful. Both items deal with faculty-student interaction as do items 8, 9, 10, and 16 for which fairly high percentages of instructors also gave themselves better ratings. The faculty-student interaction dimension, then, appears to be one on which a sizable number of instructors and their students do not agree and on which student reactions would appear to be especially crucial. Other similar areas would be the instructor's openness to other viewpoints (item 20) and the agreement between announced objectives for the course and what was being taught (6).

A surprisingly large percentage of instructors rated themselves poorer than students did in a few areas. Fifteen per cent rated themselves more poorly on class preparation and 12 per cent were less satisfied that they were accomplishing course objectives. In general, however, only between 4 to 8 per cent of the teachers gave themselves considerably poorer ratings.

One of the items in the form was unique in that it elicited opinions on student effort in the course (19). For students, the exact wording was: "I have been putting a good deal of effort into this course"; for instructors it was worded: "Students seem to be putting a good deal of effort into this course." The results for this item, as one might expect, were much different than those for other items. Compared to students' responses, 18 per cent of the faculty thought students generally were putting considerably less effort into the course, while 10 per cent gave students better ratings on effort than students gave themselves. In other words, in this instance

students have tended to give themselves better ratings just as instructors did on so many of the previous items.

An inspection of the differences within each college indicates fairly similar results with the exception of college five. In comparison to the other four colleges, higher percentages of the instructors at college five rated themselves considerably better than did their students on a majority of the items. While it is not possible to conclude much on the basis of one college, it is interesting to note that college five was the smallest and most selective of the colleges in the study. Moreover, instructors at college five were given the poorest student ratings among the five colleges, whereas their self-ratings were not much different or poorer than those of instructors elsewhere. Thus, the gap between instructor-student ratings at college five was due largely to the poorer ratings by students, perhaps because of higher expectations on their part, rather than on better ratings by instructors.

Presented in Table 3 is a summary of responses to the first four items, which used varied responses rather than agree-disagree options. The items deal with the pace, the level of difficulty, and the work load of the course, as well as the extent to which the instructor used examples and illustrations. Once again there were student-instructor differences although they were not particularly large. Instructors tended to think they more often used examples and illustrations, and at three of the colleges instructors more likely considered the pace at which material was covered to be slow. College five, the selective liberal arts college, was once again noteworthy in that its faculty and to some extent the students reported less frequent use of examples or illustrations in courses.

Table 3

Faculty-Student Comparisons at Five Colleges and Total (N = 343), for Four Items in Instructional Report Questionnaire

	Percentage Responding[a]											
	Students						Faculty					
	College						College					
	1	2	3	4	5	Total	1	2	3	4	5	Total
1 Pace at which material is covered:												
Very or somewhat slow	9	10	7	8	6	9	22	24	10	8	14	16
Very or somewhat fast	26	20	27	23	33	25	20	28	24	30	30	27
2 Level of difficulty of course for students enrolled:												
Very or somewhat elementary	11	13	10	10	9	11	10	7	10	4	8	7
Very or somewhat difficult	31	25	32	21	38	30	26	31	37	37	41	34
3 Work load of course relative to others:												
Lighter	18	22	17	19	18	19	25	24	21	17	14	20
Heavier	20	21	27	29	27	25	35	23	32	32	33	30
4 Extent to which examples and illustrations were used:												
Frequently	60	70	76	67	58	67	88	75	86	82	65	80
Occasionally	28	26	20	26	34	26	12	21	14	18	32	19
Seldom	10	4	4	6	8	6	0	2	0	0	3	1
Never	2	1	1	1	1	1	0	2	0	0	0	1

[a]For items 1-3, the four responses have been collapsed into two categories; the middle response ("about right" or "about the same") is not shown.

A final question regarding individual teacher-class differences was whether those differences were related to instructors of different sexes, with varying amounts of teaching experiences, or those teaching different subject areas. Are the self-ratings for female teachers, for example, more similar to their students' ratings than are those of male teachers? For this analysis, each course was grouped into one of four general subject area categories: natural sciences, humanities, social sciences, and education and applied subjects (e.g., business, home economics, nursing). Teaching experience consisted of three categories: one or two years, three to six years, and seven years or more. Data for 235 teachers were available for this analysis.

The results of the multivariate analysis of variance, in which all 21 items were used as variables, are given in Table 4. There were no differences due to sex or years of teaching experience or for any of the interactions; there was, however, a significant difference ($p < .05$) due to subject area. This difference was largely between natural science courses and those in education and applied subjects. Specifically, teachers in the natural sciences did not think the pace of the course was as fast as their students said it was, and they did not think students put as much effort into the course as students said they did. Conversely, teachers in education and applied subjects reported the course as having a faster pace than their students reported, and thought that students put more effort into the course than students said they did.

Table 4

Summary of MANOVA Results of Instructor-Class Differences by Sex, Subject Area, and Number of Years Teaching (N = 235)

Source[a]	df Hypothesis	df Error	F	p <
Sex	21	192	.34	.99
Years of Teaching	42	384	1.09	.34
Subject Area	63	574	1.33	.05
Sex x Years Teaching	42	384	.86	.72
Sex x Subject Area	63	574	.62	.99
Years Teaching x Subject Area	126	1121	.85	.89

[a]The triple-order interaction was not run because one of the cells was blank.

Summary and Conclusions

A comparison of students' ratings of instruction with teachers' self-reported ratings in over 300 classes at five colleges disclosed a modest relationship between the two sets of evaluations. The median correlation for 17 items was .21, indicating that faculty members generally evaluate or describe their teaching somewhat differently from the way it is evaluated or described by their students. Not surprisingly, the highest correlations occurred for the more factual items, on which there was somewhat less chance for disagreement (e.g., the instructor informs students of how they would be evaluated), while items eliciting opinions (e.g., the instructor is using class time well) resulted in the lowest correlations.

As mentioned earlier, previous studies, in which students and faculty ratings of instruction had been compared, employed a single overall measure of teaching and produced conflicting results: .62 in one instance (Webb & Nolan, 1955) and .19 in the other (Clark & Blackburn, 1971). The latter correlation was reported for college teachers and, of course, was fairly similar to the median correlation for the 17 items used in the five-college study reported here. Webb and Nolan's use of instructors in a military setting may explain the unusually high correlation found in their study; in any event, it does not seem to apply to more typical college teaching situations.

In addition to the general lack of agreement between self and student evaluations, there was also a tendency for teachers as a group to give themselves better ratings than their students did. In a sense this tendency might be viewed as only "human," or certainly not surprising.

As Robert Burns has reminded us, most people do not see themselves as others see them; teachers and the way they see their instruction are apparently no exception.

Comparisons between student and faculty responses were also made across items, and a rank correlation of .77 indicated a good deal of similarity in the way the two groups rank ordered the items. This suggests that instructors are indeed aware of many of their particular teaching strengths and weaknesses, even though they see themselves more favorably in absolute terms. They are also probably more aware of their own relative strengths and weaknesses than they are of the way they might compare to other instructors, as suggested by the previously cited correlational data for each item. An ipsative approach to student rating of faculty, therefore, in which the emphasis is on identifying the specific "good" and "bad" practices of each individual teacher, would not appear to be as informative to instructors as the normative approach, in which comparisons may be made with other relevant groups of instructors.

The discrepancy between individual teacher ratings and the mean rating given by his class was most notable for between a fourth to a third of the 343 instructors in the study, and in particular for items related to student-instructor interaction, course objectives, and the instructor's openness to other viewpoints. These areas of instruction, then, would seem to be particular ones in which a sizable proportion of teachers could profit from student feedback.

Teacher-student discrepancies were about the same for men and women teachers and for the more and less experienced teachers. That there were no sex differences in rating discrepancies is not particularly surprising; but one might have predicted that the self-ratings of more experienced teachers would be closer to student ratings. Since most of the teachers in this study had not made a practice of obtaining systematic feedback from their students, the findings suggest that getting to know student reactions to teaching is not something that comes merely with experience.

Of particular interest, however, were differential discrepancies noted for the subject areas; teachers of natural science subjects underestimated (relative to their students) both the pace of their course and their students' efforts, while teachers of education and applied subjects overestimated the course pace and their students' efforts. These subject area differences might be explained by the differences in the content and in the intended objectives of courses in each area. Instructors of mathematics, physics, biology, and the like may feel that there is so much factual and theoretical material to cover in their courses that a fast pace coupled with a good deal of student effort is a necessity. What teachers in the natural sciences view as an acceptable pace and work load, however, apparently does not coincide with their students, who frequently are using courses in other fields for comparison. In education and applied subject areas, not only might the amount of factual material be less demanding on students, but frequently the major objectives of the courses are to establish particular attitudes or skills with students. Working toward those objectives may result in courses that appear slower paced to students.

In conclusion the results of this study would argue for the collection of student ratings as a means of providing instructors with information they do not already have about their teaching. As an aid to instructional improvement, teacher self-ratings might in fact be used in conjunction with student feedback as a means of highlighting discrepancies for the individual instructor.

References

Centra, J. A. The student instructional report: Its development and uses. SIR Report Number 1. Princeton, N. J.: Educational Testing Service, 1972.

Clark, M. J., & Blackburn, R. Assessment of faculty performance: Some correlates between self, colleagues, students, and administrators. Center for the Study of Higher Education, University of Michigan, 1971. (Mimeo)

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-285.

Gibb, C. A. Classroom behavior of the college teacher. Educational and Psychological Measurement, 1955, 15, 254-263.

Hodgson, T. F. The general and primary factors in student evaluation of teaching ability. Seattle: University of Washington, 1958.

Isaacson, R., McKeachie, W., Milholland, J., Lin, Y. G., Hofeller, M., Baerwalt, J., & Zinn, K. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.

Webb, W. B., & Nolan, C. Y. Student, supervisor, and self-ratings of instructional proficiency. Journal of Educational Psychology, 1955, 46, 42-46.

Appendix A

Examples of Letters Sent to Colleges and Individual Teachers

Initial letter to some 60 colleges.
Self-addressed postcard was also included.

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

We have received a grant to study the effects of student ratings of instruction. More precisely, we are interested in assessing the impact of formal student feedback on teaching behavior. The study, to be conducted next fall, will provide participating colleges with the free use and scoring of a rating form that we have been developing over the past year. Questionnaire responses will be returned only to the individual instructor.

In order to invite colleges to participate we first need to obtain some general information about the extent to which student rating forms are used on college campuses. Could you or a staff member please respond to the three brief questions pertaining to your institution listed on the enclosed self-addressed postcard? We would appreciate your reply by May 5. Your responses do not commit you to participate in the study; if invited, you will of course be provided with more details.

Thank you.

Sincerely,

John A. Centra
Research Psychologist
Project Director

Invitation to colleges to participate.

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTEST N.J.

Developmental Research Division

A short while ago you received a letter briefly describing a study we will be conducting during the fall, 1971 semester dealing with student ratings of instruction. We would like to invite College to participate in this study. As mentioned in the earlier letter, your faculty will be asked to use a student rating form that we have been developing here at ETS over the past year. Student responses to the questionnaire will be scored and returned to each professor at no cost to the college.

The major purpose of the study is to investigate the effectiveness of formal student reactions in modifying college instruction. Funded by the Esso Education Foundation, the study will hopefully shed some light on whether student ratings lead to positive changes in instruction. The procedure is as follows:

1. We would select a portion of your faculty to use a brief rating form at midsemester in one of their classes. The form will take only 10-15 minutes for students to complete. Half of this group of instructors will receive a summary of student responses a few days later; the other half will not receive these results until after the end of the semester.

2. At the end of the semester we would like the remaining faculty as well as the original group to administer our rating form in one of their classes. For that portion of the faculty that used the form at midsemester, this will represent a second use in the same class; other faculty will be employing it for the first time. The second rating form will be identical to the first except for some additional questions pertinent to an end of course appraisal. Once again the questionnaires will be scored at ETS and sent back to each instructor.

Thus one portion of your faculty will use a student rating form twice during the semester and we will compare end-of-semester responses for those who received their results immediately vs. those who did not. The third group of instructors will simply use the form at the end of the semester which is normally the case. Further analyses with all three groups will also help us learn more about the particular items included in the form.

We plan to include seven or eight colleges in this study and we hope that can be one of these. The information that individual faculty members receive from the rating form should be of interest to them. Furthermore, summary results will be made available for each college, which will provide you with an overall view of student reactions to instruction at If you feel your college can participate, we will need to hear from you as soon as possible. We would also need the name of a person at the college who would act as our liaison.

If you need further information about the study at this time, feel free to call me collect at (609) 921-9000, extension 2793.

I will look forward to hearing from you.

Sincerely,

John A. Centra
Research Psychologist
Project Director

JAC:lil

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE:EDUCTESTSVC

Developmental Research Division

October 14, 1971

Dr. , Vice President for Academic Affairs at , mentioned to you in his memo of September 20 that College along with several other colleges was participating with us in a project dealing with student course evaluation. The project, funded by the Esso Educational Foundation, is investigating what students are able to evaluate in the classroom and how useful this information might be to the individual instructor.

Your participation in the project will be brief and, hopefully, of value to you. There are 30 copies of a mid-semester Student Instructional Report enclosed. These should be used in one of your classes--of your own choosing--which will be meeting on October 18, 19, or 20. If you need additional copies of the questionnaire, they are available in Dr. office. After the forms have been administered, put them back in the enclosed envelope (which has my name on it), seal it (to maintain confidentiality) and put it in interoffice mail. The mail clerk will set them aside for me. They should be returned no later than Wednesday the 20th.

The questionnaires will be scored here at ETS, summarized, and returned to you as soon as possible. Only you will receive this summary of student responses. Because of the number of colleges and faculty members participating in this project and possible delays in processing the forms, we may not be able to get all of the summaries back at the same time.

We will be also asking you to administer an end-of-semester form in the same class. That form is slightly longer than the mid-semester form and allows you to add questions of your own for scoring.

There is one last request. One of the questionnaires is marked "Instructor's Form." Would you respond to the items according to what you feel you deserve--that is, according to how you would describe or rate yourself. While this may be difficult to do for a few of the items, we have good reason to ask for the information: we would like to know the extent to which some of the items may be tapping information that most instructors are already aware of. The questionnaire with your responses should be enclosed along with those filled out by students.

Thank you for your cooperation. You will receive the end-of-semester questionnaire around December 1.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921 - 9000
CABLE-EDUCTESTSVC

Developmental Research Division

October 14, 1971

Dr. , Vice President for Academic Affairs at , mentioned to you in his memo of September 20 that College along with several other colleges was participating with us in a project dealing with student course evaluation. The project, funded by the Esso Educational Foundation, is investigating what students are able to evaluate in the classroom and how useful this information might be to the individual instructor.

Your participation in the project will be brief and hopefully of value to you. I would like you to have your students complete an "Instructional Report" questionnaire in one of your classes at the end of the fall semester. This questionnaire consists of 39 items which elicit student perceptions and ratings of the course and instruction; it should take students only 10-15 minutes to complete. The form also allows you to add questions of your own for scoring. Student responses will be scored, summarized by us, and sent back to each instructor. Only the individual instructor will receive these summaries.

You will receive a supply of questionnaires in your mailbox along with additional instructions around December 1.

Thank you for your assistance. I hope you will find your participation worthwhile.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist and
Project Director

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

October 14, 1971

The enclosed material deals with the project on student ratings of instruction which Dean has mentioned to you and which I described briefly at the first faculty meeting. The project, you may recall, is investigating what students are able to evaluate in the classroom and how useful this information might be to the individual instructor.

There are 30 copies of a mid-semester Student Instructional Report enclosed. These should be used in one of your classes which will be meeting on October 18, 19, or 20. If you need additional copies of the questionnaire, they may be obtained in Dean office. After the forms have been administered, put them back in the envelope in which they were received, seal it (to maintain confidentiality) and return it to one of the mailboxes in the General Office which has been reserved for that purpose. You should return the forms no later than Wednesday the 20th.

The questionnaires will be scored here at ETS, summarized, and returned to you as soon as possible. Only you will receive this summary of student responses. Because of the number of colleges and faculty members participating in this project and possible delays in processing the forms, we may not be able to get all of the summaries back at the same time.

As you may recall, we will be also asking you to administer an end-of-semester form in the same class. That form is slightly longer than the mid-semester form and allows you to add questions of your own for scoring.

There is one last request. One of the questionnaires is marked "Instructor's Form." Would you respond to the items according to what you feel you deserve--that is, according to how you would describe or rate yourself. While this may be difficult to do for a few of the items, we have good reason to ask for the information: we would like to know the extent to which some of the items may be tapping information that most instructors are already aware. The questionnaire with your responses should be enclosed along with those filled out by students.

Thank you for your cooperation. You will receive the end-of-semester questionnaire around December 9.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist and
Project Director

Letter to feedback group, sent with midsemester class results.

EDUCATIONAL TESTING SERVICE PRINCETON, N. J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

To the Instructor

Enclosed is an IBM "printout" summarizing your students' responses to the mid-semester Student Instructional Report. Your name, course identification, and the number of students responding should appear at the top.

The printout shows the percentage of students who chose various options to each question and the average (mean) response to each. A standard deviation is also provided for those of you who would like some indication of variation in responses.

Also enclosed is a response summary based on 2960 students in 75 classes. This information was collected for 17 of the 23 items at one four-year institution last Spring. It may help you interpret your own responses, although it is by no means offered as ideal normative data. It is, however, all that is now available.

We hope you will find the information provided you at this time useful. As mentioned in previous correspondence, only you have received these results. The end-of-semester form, which is slightly longer and provides you with the opportunity to add questions of your own for scoring, will be sent to you in early January.

Thank you for participating in the project thus far.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist and
Project Director

JAC:mh

Enclosures

EDUCATIONAL TESTING SERVICE PRINCETON, N. J. 08540

Area Code 609
921 - 9000
CABLE-EDUCTESTSVC

Developmental Research Division

November 29, 1971

To the Instructor

To complete the second and final phase of our project dealing with student ratings of instruction and to give you the opportunity to obtain additional feedback from your students, you will find enclosed some end-of-semester Student Instructional Report forms. The form contains 39 items plus a section for up to 10 supplementary questions that you may add.*

The questionnaire should be administered in the same class and section in which you administered the mid-semester forms. That is imperative. I have consequently enclosed the same number of questionnaires that you used at mid-semester (plus a few extra). You should administer the forms during any class period of this last week of the semester. After they have been administered, put them back in the envelope in which they were received and seal it to maintain confidentiality. Then you should cross your name off the address label, add mine, and put it in interoffice mail where the mail clerk will set them aside for me (if you prefer you can drop it off personally to the mail clerk).

You should return them on or before the last day of class (December 8). We will score them and send you a summary of responses as soon as possible. Once again only you will receive that summary.

Also enclosed is a green-printed form titled Instructor's Cover Sheet. We have already written or coded much of the necessary information on the left side of the sheet, including your course, but we need some additional information from you for questions A through I. They will take only a minute to complete. The cover sheet should be returned in the same envelope as the forms filled out by students. It is important that you complete and return this cover sheet so that we can identify your class. It is also important that you direct your students to write the ETS report number at the top of their answer sheet. That five digit number is written at the top of your cover sheet.

Thank you once again for your help with this project. I hope you will find your results helpful. A report of this five college study will be available next Spring; let me know if you would like a copy.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist
Project Director

*Make sure students respond to both sides of the questionnaire.

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTEST SVC

Developmental Research Division

December 3, 1971

To the Instructor

To complete the second and final phase of our project dealing with student ratings of instruction and to give you the opportunity to obtain additional feedback from your students, you will find enclosed some end-of-semester Student Instructional Report forms. The form contains 39 items plus a section (IV) for student responses to up to 10 supplementary questions that you may add.

The questionnaire should be administered in the same class and section in which you administered the mid-semester forms. That is imperative. I have consequently enclosed the same number of questionnaires that you used at mid-semester (plus a few extra). You should administer the forms during any class period before December 17. After they have been administered, put them back in the enclosed envelope and seal it to maintain confidentiality. Then return the envelope to Mr. office (Institutional Research).

You should return them to Mr. before the Christmas holiday. We will score them and send you a summary of responses as soon as possible. Once again only you will receive that summary.

Also enclosed is a green-printed form titled Instructor's Cover Sheet. We have already written or coded much of the necessary information on the left side of the sheet, including your course designation, but we need some additional information from you for questions A through I. They will take only a minute to complete. The cover sheet should be returned in the same envelope as the forms filled out by students. It is important that you complete and return this cover sheet so that we can identify your class. It is also important that you direct your students to write the ETS Report Number at the top of their answer sheet. That five-digit number is written at the top of your cover sheet.

Thank you once again for your help with this project. I hope you will find your results helpful. A report of this five-college study will be available late next Spring; let me know if you would like a copy.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist
Project Director

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTEST

Developmental Research Division

December 3, 1971

You will undoubtedly recall that you administered a Student Instructional Report in one of your classes at mid-semester as part of our project dealing with the effectiveness of student ratings of instruction. Enclosed is a computer printout summarizing your students' responses.

Although I could not say so at the time you were initially contacted, it was necessary to withhold these responses from you (and many of your faculty colleagues) in order to carry out the project. Briefly, we are trying to find out in what way, if any, student feedback may modify instructional procedures. Part of the faculty at were given a summary of student responses a few days after administering the forms, and the other part, which included you, were not given that information until now. Both faculty groups were determined on a strictly random basis.

To complete the project and to give you the opportunity to obtain additional information from your students, I've enclosed a batch of end-of-semester Student Instructional Report forms. The form contains 39 items plus a section for student responses to up to 10 supplementary questions that you may add (Section IV).

The forms should be administered in the same class and section in which you administered the mid-semester forms. That is imperative. Consequently I've enclosed the same number of questionnaires that you used at mid-semester (plus a few extra). You should administer the questionnaire during any class period before December 17. After they have been administered, put them back in the enclosed envelope and seal it to maintain confidentiality. Then return the envelope to Mr. office (Institutional Research).

You should return the forms to Mr. before the Christmas holiday. We will score them and send you a summary of responses as soon as possible. Only you will receive that summary.

Finally, you will find enclosed a green-printed form titled Instructor's Cover Sheet. We have already written or coded much of the necessary information on the left side of the sheet, including your course designation, but we need your responses to questions A through I. They will take only a minute to complete. The cover sheet should be enclosed in the same envelope as the student forms. It is important that you complete and return this cover sheet since it is the only way that we can identify the information for your course. It is also important that you instruct your students to write the ETS Report Number at the appropriate place on their answer sheet. That five-digit number is written toward the top of your cover sheet.

Thank you once again for your help on this project. If you would like additional details about it, I would be happy to provide them. A report of this five-college study will be available late next Spring; let me know if you would like a copy.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist
Project Director

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

December 3, 1971

To the Instructor

You will recall receiving a letter from me in October indicating that you would receive a supply of questionnaires from me at this time dealing with our project on student ratings of instruction. The project is funded by the Esso Educational Foundation and is investigating what students are able to evaluate in the classroom and how useful this information might be to the individual instructor.

Enclosed are 35 copies of the Student Instructional Report which I would ask you to use in one of your classes (of your own choosing). If you need additional forms, Mr. has a limited supply. You should administer the forms during any class period before December 17. After they have been administered, put them back in the enclosed envelope and seal it to maintain confidentiality. Then return the envelope to Mr. office (Institutional Research).

You should return them to Mr. before the Christmas holiday. We will score them and send you a summary of responses as soon as possible. Only you will receive that summary.

You will note that the form contains 39 items plus a section (IV) for student responses to up to 10 supplementary questions that you may add. Make sure students respond to both sides of the questionnaire.

Also enclosed is a green-printed form titled Instructor's Cover Sheet. We have already written or coded much of the necessary information on the left side of the sheet, but we need some additional information from you for questions A through I. They will take only a minute to complete. The cover sheet should be returned in the same envelope as the forms filled out by students. It is important that you complete and return this cover sheet so that we can identify your class. It is also important that you direct your students to write the ETS Report Number at the top of their answer sheet. That five-digit number is written at the top of your cover sheet.

Thank you for your help on this project. I hope you will find your results helpful. A report of this five-college study will be available late next Spring; let me know if you would like a copy.

Sincerely,

John A. Centra

John A. Centra
Research Psychologist
Project Director

Comparison data for the feedback group

STUDENT INSTRUCTIONAL REPORT

Responses Based on 2960 Students in 75 Classes at One Institution, Spring 1971

Item, Percentage Responding, Mean (Average) Response, and Standard Deviation of the Responses[1]

1. For me, the pace at which the instructor covers the material is:

		%	
1	Very slow	2.3	
2	Somewhat slow	14.1	Mean 3.05
3	Just about right	62.1	s.d. .73
4	Somewhat fast	18.7	
5	Very fast	2.7	

2. For my preparation and ability, the level of difficulty of this course is:

		%	
1	Very elementary	3.7	
2	Somewhat elementary	15.9	Mean 3.13
3	About right	49.6	s.d. .87
4	Somewhat difficult	25.3	
5	Very difficult	5.5	

3. The work load for this course in relation to other courses of equal credit is:

		%	
1	Much lighter	4.0	
2	Lighter	15.6	Mean 3.19
3	About the same	46.3	s.d. .94
4	Heavier	25.6	
5	Much heavier	8.5	

4. To what extent does the instructor use examples or illustrations to help clarify the material?

1	Frequently	Responses
2	Occasionally	not available
3	Seldom	
4	Never	

Items	Percentages Strongly Agree (1)	(2)	(3)	Strongly Disagree (4)	Mean (Average)	Standard Deviation
5. The instructor's objectives for the course have been made clear	36.5	49.2	11.8	2.5	1.80	.74
6. There was considerable agreement between the announced objectives of the course and what is being taught	not available					
7. The instructor is using class time well	38.7	42.2	15.2	4.0	1.84	.82
8. The instructor is readily available for consultation with students	43.8	47.6	6.9	1.7	1.66	.68
9. The instructor seems to know when students didn't understand the material	18.0	54.0	23.4	4.6	2.15	.76
10. Lectures are too repetitive of what is in the textbook(s)	4.5	12.9	56.4	26.2	3.04	.75
11. The instructor encourages students to think for themselves	29.2	50.5	17.5	2.8	1.94	.76
12. The instructor seems genuinely concerned about whether students learn and is actively helpful	35.4	51.4	11.6	1.6	1.79	.70
13. The instructor makes helpful comments on papers or exams	19.7	40.9	29.7	9.7	2.30	.89
14. The instructor raises challenging questions or problems for discussion	24.4	46.2	25.1	4.4	2.09	.81
15. In this class I feel free to ask questions or express my opinions	48.6	45.4	5.1	0.9	1.58	.63
16. The instructor is well-prepared for each class	61.1	33.5	4.1	1.3	1.46	.64
17. The instructor has informed students how they would be evaluated in the course	43.0	44.4	10.7	1.9	1.71	.73
18. The instructor summarizes or emphasizes major points in lectures or discussions	35.5	45.6	16.5	2.5	1.86	.77
19. My interest in the subject area is being stimulated by this course	35.6	38.9	17.9	7.6	1.97	.92

Responses to items 20 to 23 were not available.

[1]Not Applicable (NA) and omits were excluded in computing percentages, means, and standard deviations.

Appendix C

Midsemester and End-of-Semester

Student Instructional Report

Midsemester
STUDENT INSTRUCTIONAL REPORT

Name of Instructor _____

Name of Course _____

This questionnaire gives you an opportunity to express anonymously your perceptions of this course and the way it has been taught. Responses will be summarized and given only to your instructor.

SECTION I. Items 1-4. *Directions* Circle one response number for each question. Use any pen or pencil you have handy.

1. For me, the pace at which the instructor covers the material is.
   1 Very slow
   2 Somewhat slow
   3 Just about right
   4 Somewhat fast
   5 Very fast

2. For my preparation and ability, the level of difficulty of this course is:
   1 Very elementary
   2 Somewhat elementary
   3 About right
   4 Somewhat difficult
   5 Very difficult

3. The work load for this course in relation to other courses of equal credit is
   1 Much lighter
   2 Lighter
   3 About the same
   4 Heavier
   5 Much heavier

4. To what extent does the instructor use examples or illustrations to help clarify the material?
   1 Frequently
   2 Occasionally
   3 Seldom
   4 Never

SECTION II. Items 5-23. *Directions*: Circle the number that represents the response closest to your opinion.

NA (0) = *Not Applicable or don't know*. The statement does not apply to this course or instructor, or you simply are not able to give a knowledgeable response

SA (1) = *Strongly Agree*. You strongly agree with the statement as it applies to this course or instructor.

A (2) = *Agree*. You agree more than you disagree with the statement as it applies to this course or instructor.

D (3) = *Disagree*. You disagree more than you agree with the statement as it applies to this course or instructor.

SD (4) = *Strongly Disagree*. You strongly disagree with the statement as it applies to this course or instructor.

		NA	SA	A	D	SD
5.	The instructor's objectives for the course have been made clear	0	1	2	3	4
6.	There has been considerable agreement between the announced objectives of the course and what is being taught	0	1	2	3	4
7.	The instructor is using class time well	0	1	2	3	4
8.	The instructor is readily available for consultation with students	0	1	2	3	4
9.	The instructor seems to know when students don't understand the material	0	1	2	3	4
10.	Lectures are too repetitive of what is in the textbook(s)	0	1	2	3	4
11.	The instructor encourages students to think for themselves	0	1	2	3	4
12.	The instructor seems genuinely concerned about whether students learn and is actively helpful	0	1	2	3	4
13.	The instructor makes helpful comments on papers or exams	0	1	2	3	4
14.	The instructor raises challenging questions or problems for discussion	0	1	2	3	4
15.	In this class I feel free to ask questions or express my opinions	0	1	2	3	4
16.	The instructor is well-prepared for each class	0	1	2	3	4
17.	The instructor has informed students of how they would be evaluated in the course	0	1	2	3	4
18.	The instructor summarizes or emphasizes major points in lectures or discussions	0	1	2	3	4
19.	My interest in the subject area is being stimulated by this course	0	1	2	3	4
20.	The scope of the course is too limited; not enough material is being covered	0	1	2	3	4
21.	I have been putting a good deal of effort into this course	0	1	2	3	4
22.	The instructor is open to other viewpoints	0	1	2	3	4
23.	In my opinion, the instructor is accomplishing his objectives for the course	0	1	2	3	4

THANK YOU

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION

This questionnaire gives you an opportunity to express anonymously your views of this course and the way it has been taught. Indicate the response closest to your view by blackening the appropriate oval. Use a soft lead pencil (preferably No. 2) for all responses to the questionnaire. Do not use an ink or ball point pen.

ETS Report Number

SECTION I Items 1-20. Blacken one response number for each question.

NA (0) = Not Applicable or don't know. The statement does not apply to this course or instructor, or you simply are not able to give a knowledgeable response.

SA (1) = Strongly Agree. You strongly agree with the statement as it applies to this course or instructor.

A (2) = Agree. You agree more than you disagree with the statement as it applies to this course or instructor.

D (3) = Disagree. You disagree more than you agree with the statement as it applies to this course or instructor.

SD (4) = Strongly Disagree. You strongly disagree with the statement as it applies to this course or instructor.

		NA	SA	A	D	SD
1.	The instructor's objectives for the course have been made clear					
2.	There was considerable agreement between the announced objectives of the course and what was actually taught					
3.	The instructor used class time well					
4.	The instructor was readily available for consultation with students					
5.	The instructor seemed to know when students didn't understand the material					
6.	Lectures were too repetitive of what was in the textbook(s)					
7.	The instructor encouraged students to think for themselves					
8.	The instructor seemed genuinely concerned with students' progress and was actively helpful					
9.	The instructor made helpful comments on papers or exams					
10.	The instructor raised challenging questions or problems for discussion					
11.	In this class I felt free to ask questions or express my opinions					
12.	The instructor was well-prepared for each class					
13.	The instructor told students how they would be evaluated in the course					
14.	The instructor summarized or emphasized major points in lectures or discussions					
15.	My interest in the subject area has been stimulated by this course					
16.	The scope of the course has been too limited; not enough material has been covered					
17.	Examinations reflected the important aspects of the course					
18.	I have been putting a good deal of effort into this course					
19.	The instructor was open to other viewpoints					
20.	In my opinion, the instructor has accomplished (is accomplishing) his objectives for the course					

SECTION II Items 21-31. Blacken one response number for each question.

21. For my preparation and ability, the level of difficulty of this course was:

- Very elementary
- Somewhat elementary
- About right
- Somewhat difficult
- Very difficult

22. The work load for this course in relation to other courses of equal credit was:

- Much lighter
- Lighter
- About the same
- Heavier
- Much heavier

23. For me, the pace at which the instructor covered the material during the term was:

- Very slow
- Somewhat slow
- Just about right
- Somewhat fast
- Very fast

24. To what extent did the instructor use examples or illustrations to help clarify the material?

- Frequently
- Occasionally
- Seldom
- Never

572MRC91P50

Princeton, N.J. 08540

25 Was class size satisfactory for the method of conducting the class?

- Yes, most of the time
- No, class was too small
- No, class was too large
- It didn't make any difference one way or the other

26 Which one of the following best describes this course for you?

- Major requirement or elective within major field
- College requirement but not part of my major or minor field
- Minor requirement or required elective outside major field
- Elective not required in any way
- Other

27 Which one of the following was your most important reason for selecting this course?

- Friend(s) recommended it
- Faculty advisor's recommendation
- Teacher's excellent reputation
- Thought I could make a good grade
- Could use pass/no credit option
- It was required
- Subject was of interest
- Other

28. What grade do you expect to receive in this course?

- A
- B
- C
- D
- Fail
- Pass
- No credit
- Other

29. What is your approximate cumulative grade-point average?

- 3.50-4.00
- 3.00-3.49
- 2.50-2.99
- 2.00-2.49
- 1.50-1.99
- 1.00 1.49
- Less than 1.00
- None yet--freshmen or transfer

30. What is your class level?

- Freshman
- Sophomore
- Junior
- Senior
- Graduate
- Other

31. Sex:

- Female
- Male

SECTION III Items 32-39. Blacken one response number for each question.

		Not applicable, don't know, or there were none.	Excellent	Good	Satisfactory	Fair	Poor
32.	Overall, I would rate the textbook(s)						
33.	Overall, I would rate the supplementary readings						
34.	Overall, I would rate the quality of the exams						
35.	I would rate the general quality of the lectures						
36.	I would rate the overall value of class discussions						
37.	Overall, I would rate the laboratories						
38.	I would rate the overall value of this course to me as						

39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course? (Blacken one response number.)

One of the most effective (among the top 10%)	More effective than most (among the top 30%)	About average	Not as effective as most (in the lowest 30%)	One of the least effective (in the lowest 10%)

SECTION IV Items 40 49. If the instructor provided supplementary questions and response options, use this section for responding. Blacken only one response number for each question.

	NA									
40.										
41.										
42.										
43.										
44.										

	NA									
45.										
46.										
47.										
48.										
49.										

SECTION V Students' Comment Section

If you would like to make additional comments about the course or instruction, use a separate sheet of paper. You might elaborate on the particular aspects you liked most as well as those you liked least. Also, how can the course or the way it was taught be improved? PLEASE GIVE THESE COMMENTS TO THE INSTRUCTOR.

If you have any comments or suggestions about this questionnaire (for example, the content or responses available), please send them to: Student Instructional Report, Educational Testing Service, Princeton, New Jersey 08540.

Appendix D

Reliability of the Student Instructional Report Items[1]

Item Number (SIR End-of-Semester Form)	Estimated reliability for the following number of individuals in each class					
	5	10	15	20	25	35
1	.39	.56	.66	.72	.76	.82
2	.37	.54	.64	.71	.75	.81
3	.50	.67	.75	.80	.83	.88
4	.49	.66	.74	.79	.83	.87
5	.56	.72	.80	.84	.87	.90
6	.50	.66	.75	.80	.83	.87
7	.47	.64	.72	.78	.81	.86
8	.69	.82	.87	.90	.92	.94
9	.39	.56	.66	.72	.76	.82
10	.37	.54	.64	.71	.75	.81
11	.50	.67	.75	.80	.83	.88
12	.49	.66	.74	.79	.83	.87
13	.56	.72	.80	.84	.87	.90
14	.50	.66	.75	.80	.83	.87
15	.47	.64	.72	.78	.81	.86
16	.69	.82	.87	.90	.92	.94
17	.71	.83	.88	.91	.93	.95
18	.57	.73	.80	.84	.87	.90
19	.52	.68	.76	.81	.84	.88
20	.58	.74	.81	.85	.88	.91
21	.31	.47	.57	.64	.69	.76
22	.45	.63	.71	.77	.81	.85
23	.57	.72	.80	.84	.87	.90
24	.28	.44	.54	.61	.66	.73
32	.66	.79	.85	.88	.91	.93
34	.59	.74	.81	.85	.88	.91
35	.58	.74	.81	.85	.87	.91
36	.59	.74	.81	.85	.88	.91
38	.55	.71	.79	.83	.86	.90
39	.65	.78	.85	.88	.90	.93

*Based on analysis of variance (see Winer, 1962, pp. 124-132). Fifteen students were randomly selected from each of 28 classes for this analysis. Estimated reliabilities for the various numbers of individuals were calculated by Spearman-Brown.

Appendix E

Midsemester-End of Semester Correlations

for Twenty-Three SIR Items[1]

Item Number (mid-semester form)	Feedback Group 1 N ≅ 137 teachers	No-Feedback Group 2 N ≅ 159 teachers
1	.70	.77
2	.73	.75
3	.83	.80
4	.74	71
5	.66	.72
6	.58	.68
7	.78	.76
8	.61	.68
9	.64	.69
10	.76	.65
11	.73	.75
12	.71	.72
13	.67	.67
14	.79	.79
15	.77	.74
16	.80	.69
17	.55	.73
18	.71	.69
19	.79	.72
20	.65	.60
21	.72	.64
22	.73	.67
23	.69	.71

[1]These correlations could be viewed as test-retest reliability.

Appendix F

INSTRUCTOR'S FORM

(To be completed by the instructor)

Please respond to each item according to how you would describe this course, your teaching, or the students enrolled. The items parallel those in the student form and will give us the opportunity to see if some of the items are tapping information already known to most instructors.

Your Name_____ Name of Course:_____

SECTION I. Items 1-4. Directions: Circle one response number for each question. Use any pen or pencil you have handy.

1. For the students enrolled, the pace at which the material in this course is being covered is:
   1 Very slow
   2 Somewhat slow
   3 Just about right
   4 Somewhat fast
   5 Very fast

2. For the students enrolled, the level of difficulty of this course is:
   1 Very elementary
   2 Somewhat elementary
   3 About right
   4 Somewhat difficult
   5 Very difficult

3. In my opinion the work load for this course in relation to other courses is probably:
   1 Much lighter
   2 Lighter
   3 About the same
   4 Heavier
   5 Much heavier

4. The extent to which I have been using examples and illustrations to help clarify the material of this course is:
   1 Frequently
   2 Occasionally
   3 Seldom
   4 Never

SECTION II. Items 5-21. Directions: Circle the number that represents the response closest to your opinion.

NA (0) = Not Applicable or don't know. The statement does not apply to this course or your teaching, or you simply are not able to give a knowledgeable response.
SA (1) = Strongly Agree. You strongly agree with the statement as it applies to this course or your teaching.
A (2) = Agree. You agree more than you disagree with the statement as it applies to this course or your teaching.
D (3) = Disagree. You disagree more than you agree with the statement as it applies to this course or your teaching.
SD (4) = Strongly Disagree. You strongly disagree with the statement as it applies to this course or your teaching.

	NA	SA	A	D	SD
5. I feel my objectives for the course have been made clear to students..	0	1	2	3	4
6. There has been considerable agreement between the announced objectives of the course and what is being taught.....	0	1	2	3	4
7. I feel that I have been using class time well.....	0	1	2	3	4
8. I have been readily available for consultation with students.....	0	1	2	3	4
9. I feel I know when students don't understand the material.....	0	1	2	3	4
10. I encourage students to think for themselves in this course.....	0	1	2	3	4
11. I have been genuinely concerned about whether students learn and try to be actively helpful.....	0	1	2	3	4
12. I make a point of adding helpful comments on students papers or exams...	0	1	2	3	4
13. I have been raising challenging questions or problems for discussion..	0	1	2	3	4
14. In this class, students are free to ask questions or express their opinions.....	0	1	2	3	4
15. I think that I have been well-prepared for each class.....	0	1	2	3	4
16. I have informed students of how they would be evaluated in the course.	0	1	2	3	4
17. I have summarized or emphasized major points of lectures or discussions..	0	1	2	3	4
18. I feel that students' interest in the subject area is being stimulated by this course.....	0	1	2	3	4
19. Students seem to be putting a good deal of effort into this course....	0	1	2	3	4
20. I feel that I am open to other viewpoints.....	0	1	2	3	4
21. I feel that I am accomplishing my objectives for the course at this point..	0	1	2	3	4

THANK YOU

INSTRUCTOR'S COVER SHEET

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION

ADMIN DATE: MONTH | YEAR

CLASS ENROLLMENT

NUMBER OF ANSWER SHEETS RETURNED

COLLEGE CODE

ETS REPORT NUMBER

College Name _____

Instructor's Name _____

Department Name _____

Course Name _____

Course Number _____

Section Number _____

Optional section below may be used by instructor to specify summary report heading. See *Instructions* for further information

Report Heading-Alphabetical

Report Heading-Numerical

A. From the subjects listed on the back, record at the right the two digit number representing the field to which this course most appropriately belongs. Blacken the corresponding oval in each column. If the most appropriate fine category is not listed, indicate the most similar general category. Only one category should be chosen.

FIRST DIGIT

SECOND DIGIT

B. Indicate your academic rank.
- Teaching Assistant
- Instructor
- Assistant Professor
- Associate Professor
- Professor
- Other

C. Indicate number of years you have been teaching.
- Less than 1 full year
- 1 or 2 years
- 3 to 6 years
- 7 to 12 years
- More than 12 years

D. Indicate your credit-hour teaching load this term.
- 3 or fewer credit-hours
- 4 to 6
- 7 to 9
- 10 to 12
- 13 or more

E. In general, how was this class conducted?
- Lecture- little or no discussion
- Lecture and discussion combined
- Discussion mainly
- Lecture and laboratory
- Laboratory
- Other

F. Indicate approximate level of course
- Freshman-Sophomore
- Junior-Senior
- Graduate
- Other

G. Have there been major revisions in the teaching methods used for this course during the last
- 1 year
- 2 years
- 3 years
- 4 or more years
- There have been no changes

H. Where did students complete the questionnaire?
- In the classroom
- Out of the classroom
- Some in and some out of the classroom
- Other

I. When was the questionnaire completed?
- At or around the midpoint of the course
- Toward the end of the course

Princeton, N.J. 08540.

572MRC91P8

Appendix H

Student Instructional Report Summary Data

Mean of Mean Student Responses for Each of Five General Subject Areas and for the Total (N=436 Instructors)

See questionnaire for wording of items and responses. Note: For most items, a lower mean value is a better (i.e., more positive) response.

Item	Biological Sciences Teachers N=29	Humanities N=146	Math, Physical Sciences Engineering N=67	Social Sciences N=73	Education and Applied Areas N=121	Total N=436
1	1.67	1.69	1.70	1.70	1.63	1.68
2	1.72	1.78	1.71	1.78	1.71	1.74
3	1.66	1.70	1.64	1.73	1.71	1.70
4	1.62	1.60	1.61	1.57	1.58	1.59
5	2.09	1.93	2.04	1.94	1.88	1.94
6	2.94	3.01	3.00	3.04	2.98	3.00
7	1.87	1.70	1.79	1.75	1.63	1.71
8	1.82	1.75	1.80	1.74	1.72	1.75
9	2.25	1.90	2.09	2.12	1.99	2.01
10	2.04	1.85	1.93	1.86	1.76	1.85
11	1.77	1.69	1.74	1.62	1.55	1.65
12	1.48	1.51	1.50	1.54	1.57	1.53
13	1.66	1.84	1.80	1.66	1.65	1.74
14	1.69	1.69	1.70	1.66	1.67	1.68
15	2.01	2.02	2.20	1.92	1.84	1.98
16	3.27	3.19	3.25	3.19	3.13	3.19
17	2.04	2.07	1.88	2.01	2.14	2.05
18	1.76	1.92	1.89	1.93	1.77	1.86
19	1.80	1.78	1.85	1.71	1.71	1.76
20	1.64	1.69	1.75	1.68	1.64	1.68
21	3.53	3.27	3.61	3.28	3.18	3.32
22	3.43	3.11	3.30	3.12	3.13	3.17
23	3.46	3.18	3.37	3.16	3.11	3.20
24	1.38	1.49	1.31	1.40	1.39	1.41
32	2.31	2.41	2.81	2.49	2.38	2.47
33	2.52	2.43	2.86	2.45	2.40	2.51
34	2.55	2.49	2.48	2.49	2.46	2.49
35	2.21	2.09	2.12	2.17	2.10	2.12
36	2.63	2.39	2.55	2.42	2.19	2.38
37	2.75	2.64	2.37	2.57	2.13	2.44
38	2.36	2.22	2.46	2.19	2.04	2.21
39	2.23	2.15	2.20	2.15	2.08	2.14

Appendix I

Student Instructional Report Summary Data

Mean of Mean Student Responses for Each of Five Colleges and for the Total (N=436 Instructors)

See questionnaire for wording of items and responses. Note: For most items, a lower mean value is a better (i.e., more positive) response.

Item	College 1 N=69	College 2 N=113	College 3 N=65	College 4 N=137	College 5 N=52	Total N=436
1	1.67	1.53	1.66	1.74	1.88	1.68
2	1.80	1.58	1.75	1.79	1.87	1.74
3	1.67	1.58	1.62	1.77	1.88	1.70
4	1.64	1.46	1.63	1.61	1.70	1.59
5	1.97	1.78	1.92	2.01	2.12	1.94
6	2.87	3.11	2.94	3.02	2.97	3.00
7	1.75	1.58	1.70	1.75	1.88	1.71
8	1.76	1.57	1.74	1.86	1.88	1.75
9	1.92	1.86	1.94	2.14	2.25	2.01
10	1.83	1.71	1.84	1.87	2.15	1.85
11	1.65	1.52	1.66	1.68	1.86	1.65
12	1.61	1.43	1.53	1.55	1.58	1.53
13	1.71	1.59	1.73	1.77	2.03	1.74
14	1.63	1.57	1.70	1.74	1.81	1.68
15	2.01	1.74	2.02	2.07	2.20	1.98
16	3.09	3.21	3.13	3.24	3.22	3.19
17	1.96	2.06	2.03	2.08	2.10	2.05
18	1.89	1.73	1.90	1.89	1.99	1.86
19	1.74	1.60	1.81	1.82	1.91	1.76
20	1.73	1.50	1.65	1.75	1.85	1.68
21	3.23	3.21	3.35	3.39	3.44	3.32
22	3.01	3.09	3.18	3.27	3.26	3.17
23	3.15	3.12	3.22	3.26	3.29	3.20
24	1.46	1.34	1.39	1.43	1.50	1.41
32	2.29	2.26	2.61	2.64	2.56	2.47
33	2.41	2.27	2.64	2.68	2.62	2.51
34	2.33	2.30	2.49	2.65	2.64	2.49
35	2.08	1.88	2.09	2.25	2.37	2.12
36	2.24	2.09	2.26	2.51	2.99	2.38
37	2.52	2.16	2.51	2.54	2.87	2.44
38	2.18	1.94	2.20	2.32	2.55	2.21
39	2.18	1.95	2.02	2.26	2.35	2.14