DOCUMENT RESUME

ED 110 113 JC 750 419

AUTHOR Ostrowski, Michael V.
TITLE A Comparison of Grades Students Achieve at William Rainey Harper College and How They Rate the Effectiveness of Their Instructor at Mid-Term During the Spring 1975 Semester.
PUB DATE Jul 75
NOTE 29p.; Ed.D. Practicum, Nova University. Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS Effective Teaching; *Evaluation Criteria; Faculty Evaluation; *Grades (Scholastic); *Junior Colleges; Junior College Students; *Student Opinion; *Teacher Evaluation; Tenure

ABSTRACT

In order to determine whether student evaluation of faculty performance is related to the grade received by the student, a study was conducted in a large lecture class at William Rainey Harper College (Illinois). Of 300 students, 75 were systematically selected for the sample. The students were asked to list all courses for which they were enrolled at midterm, their midterm grade in each, and their rating of each instructor on a scale of 1-5. The investigator found that there was a significant correlation between letter grade received and the faculty rating; ratings go up as grades go up. The implications of this finding in a college where student evaluation is a major part of the faculty retention/tenure/merit process are discussed, as is the potential correlation between alternative teaching methods, such as mastery learning, grades, and consequent student satifaction. The investigator makes several suggestions for further study. (MJK)

A Comparison of Grades Students Achieve At William Rainey Harper College And How They Rate the Effectiveness of Their Instructor At Mid-Term During the Spring 1975 Semester

by

Michael V. Ostrowski, M.S., M.Ed.

A Practicum Presented to Nova University
In Partial Fulfillment of the Requirements for
the Degree of Doctor of Education

Nova University

July 12, 1975

Table of Contents

II. Statement of Problem

The writer was interested in the problem of faculty evaluation by the students. The writer believed that the students' evaluation of their instructors was one very important way of evaluating faculty and the writer believed that the students were in an excellent position to make this evaluation.

The writer was interested in knowing if there were any relationships between the grades a student achieved and how he, the student, rated that instructor.

Instructors' ratings, by their students, are becoming increasingly more important as they are being used in questions of promotion in rank, tenure, retention and salary increases. This was the nature of the problem and the motivation of the writer for undertaking this particular research topic.

III. Hypotheses:

Null Hypothesis (H_o)

There is <u>no</u> significant relationship between the grade a student received in a class and how he rated the instructor in that class.

Alternative Hypothesis (H_a)

There <u>is</u> a significant relationship between the grade a student received in a class and how the student rated the instructor of that class.

IV. Background and Significance of the Study

The writer believed that this study was of vital importance, as student ratings do play an important part in each faculty member's evaluation. The writer has overheard faculty conversation in the offices, lounges, halls, etc., that tended to be disparaging and pejorative towards student rating. Many faculty members feel that the student ratings are worthless and should not be a part of the evaluation process. Typical comments that have passed among faculty include: instructors who gave high grades and who were "soft" or lenient would tend to receive high student ratings. Other faculty members said that the instructor who maintained hard, rigid standards of excellence was punished by this system, as the students would view him as mean and punitive with his tough grading policy and thus lower his rating. Many instructors view the student ratings as a way for the students to gain revenge for a low grade achieved in a class. Some instructors perceived the ratings as a way of forcing them to "water down" the course content. The writer wished to ascertain if, in fact, it was possible for an instructor to "load the deck" in his favor and thus, by giving high grades, insure himself of obtaining a high, favorable rating.

Irene R. Kiernan states in her article that the research evidence shows that faculty who "give" good grades are likely to receive "good" student evaluations and vice versa. She further goes on to state that her data appear to show that present student evaluations measure only one thing, whether or not students like their teacher.[1]

The Rodin and Rodin study indicates that the only effective method of evaluating teacher effectiveness is whether or not the students have learned. They believe that teacher evaluation should be based on how many students complete the course with acceptable grades, an indication of their learning the material. They believe we are equating "liking" with "learning" and this is not valid.[2]

In an article by J.A. Kulik and C.L.C. Kulik, Item 5 in their conclusions indicates a slight tendency for students of highly rated teachers to outscore, on final examinations, the students of low-rated teachers.

---

[1]Irene R. Kiernan, "Student Evaluations Re-Evaluated," Community College Journal, 45 (April, 1975) pp. 25-27.

[2]M. Rodin and B. Rodin, "Student Evaluations of Teachers," Science, 117, (1972) pp. 1164-1166.

The research results have been very inconsistent in that very high positive, very high negative and inconsistent correlations all have been reported on various studies. They do agree that the highly rated teachers tend to be cultured, sophisticated, expressive, and most of all, enthusiastic.[3]

T. P. Hogan indicates that the factors that measure teaching skill are strongly influenced by the teacher's characteristics.[4]

F. N. Kerlinger believes that possibly much of the hostility that one hears from faculty about student ratings may stem from the fact that teachers see ratings as an alien intrusion into the sanctity of their classroom and the instructional process.[5]

The most damaging material presented about student ratings has been from the Rodins. The sub title to their article is, "Students Rate Most Highly Instructors

---

[3] J.A. Kulik and C.L.C. Kulik, "Student Ratings of Instruction," 1, (December, 1974), Teaching of Psychology, pp. 50-56.

[4] T.P. Hogan, "Similarity of Student Ratings Across Instructors, Courses and Time," Research in Higher Education, 1, (1973) pp. 149-154.

[5] F. N. Kerlinger, "Student Evaluation of University Professors," School and Society, 11, (1971) pp. 353-356.

From Whom They Learn Least." They found a correlation of -.75 between instructor ratings and class examination performance. The instructors with the lowest student ratings taught the three classes with the highest exam scores. The instructor with the highest student rating taught the class with the lowest exam score. They felt that good teaching was not validly measured by student ratings in their present form.[6]

In the Rodin study, it was felt that "effectiveness" is what the student learns from the instructor. In their thinking, the more the student learns, the more effective is the teacher. In many studies, the student's evaluation is rather a subjective measure of the teacher's effectiveness. Rodin further states that student evaluations, to a large extent, reflect the personal, social qualities and warmth of an instructor rather than that which he does or how he teaches his subjects.

Rodin concludes

> If how much students learn is considered to be a major component of good teaching, it must be concluded that good teaching is not validly measured by present student evaluation forms.

---

[6] M. Rodin and B. Rodin, p. 1165.

H. H. Remmers, in two studies, states that grades do not influence students' ratings of their instructors.[7]

W. J. McKeachie feels that the ultimate criterion of good teaching effectiveness should be based on how well the students meet the educational objectives of the course and he does not believe that the present evaluations take into consideration objective mastery.[8] McKeachie also believes that student ratings of instructors should not be published, as this tends to increase instructor anxiety and to prevent them from changing.

Elliott finds that teacher effectiveness is determined, in part, with certain types of students. Those teachers of high ability or highly motivated students tend to receive high ratings from those students and poor ratings from low achievement oriented students.[9]

---

[7] H. H. Remmers, "Teaching Methods in Research on Teaching," in Handbook of Research on Teaching. (Chicago, 1963).

H. H. Remmers and F. D. Martin and D. N. Elliott, "Are Students' Ratings of Instructors Related to Their Grades?" Purdue University Studies in Higher Education, 66 (1949), pp. 17-26.

[8] W. J. McKeachie, "Student Ratings of Faculty," AAUP Bulletin, 55(1969), pp. 439-444.

[9] D. H. Elliott, "Characteristics and Relationships of Various Criteria of College and University Teaching," Purdue University Studies in Higher Education, 70 (1950), pp. 5-61.

He also finds that instructors' ranks tend to affect ratings. For example, professors and associate professors are rated higher than assistant professors and instructors. The B.A. degree instructors are rated lower than those having the M.A. or Ed.D.

P. W. Frey concludes that there is no evidence for a strong positive relationship between final exam grades and the ratings, when the effect of the instructor is removed. He feels that the teacher's performance is an independent variable that is directly related to student evaluations and ratings. Of the thirteen instructors that he observed, he finds that the "type of person" the instructor appears to be and how he related to the students is more important than the grade the student earned in correlation with the rating. The personal element appears to be very important.[10]

Gessner, in 1973 has challenged Rodin's selection of teaching assistants as instructors, as being invalid. Rodin observed graduate assistants who Gessner feels were too homogeneous a group and were not experienced or committed to teaching as other more mature professors would be. He feels that with heterogeneous groups, the students can pick out the more effective teacher.

---

[10]P. W. Frey, "Student Ratings of Teaching: Validity of Several Rating Factors," Science, 192 (1973), pp. 83-85.

Frey also supports Gessner's study; they both used older, more mature, seasoned professors, whereas Rodin used a homogeneous group of teaching assistants.

Factors that were associated with professors' high ratings were: 1)clarity of presentation 2) always well prepared 3) visual presentations 4) task oriented 5) specific objectives were presented 6) highly enthusiastic and 7) ability to develop interest in their first year students which was necessary to motivate them to go on to advanced courses. There was also an expectation that one would master a given amount of material and be held accountable.

McKeachie and Lin found these factors significant in relation to high student ratings: 1) skill of the teacher 2) structure 3) feedback 4) group interaction 5) rapport or warmth 6) sensitivity 7) emotional stability. [11]

Costin, Greenough and Menges feel that high ratings are not based on entertainment value, but rather on substance teaching, especially if this was from teachers who themselves were highly interested, motivated and capable of expressing enthusiasm about their work.[12]

---

[11] W. J. McKeachie and Yi-Guan Lin, "Student Ratings of Teacher Effectiveness; Validity Studies," American Educational Research Journal, 8 (May, 1971), pp. 425-445.

[12] Frank Costin, William Greenough and Robert J. Menges, "Student Ratings of College Teaching; Reliability, Validity, and Usefulness," Review of Educational Research, 41 (December, 1971), pp. 511-535.

In the most recent study published on faculty rating, Murray shows that student rating of teachers is based on personality traits of the instructor. Four personality traits appear to be related to high student ratings such as 1) leadership 2) extroversion 3) objectivity and 4) lack of, or low anxiety. Murray also shows that peer teachers can rate new teachers on how well they will do, when using these personality traits, before they teach. This can also be used for inservice training to help teachers if the emphasis is placed on affective and interpersonal traits rather than on information or cognitive elements. It appears that students may rate the instructor's personality more than his ability to teach. The low anxiety is also related to some classroom teaching skills such as: speaking clearly and coherently. The low anxiety would also imply more self confidence.[13]

Murray also mentions that only one of these traits would be significant in a popularity contest - extroversion. This instructor must be: friendly, fair, flexible, have definite goals and take the initiative in the class-

---

[13]Harry Murray, "Predicting Student Ratings of College Teaching from Peer Ratings of Personality Types," Teaching of Psychology, 2 (April, 1975), pp. 66-69.

room situation. Initiative was very close to authoritarianism - on the benevolent side. Students do not seem comfortable with teachers they can manipulate, "psych - out" or dominate.

In conclusion, the correlations between grades and ratings may be due to greater interest in the course by students receiving the better grades. It is possible that certain instructors can so stimulate interest and excite the students that they cause the students to want to come to class, do the assignments and in turn, receive high grades. This in turn, makes the students feel good about themselves. Once a student feels "good about himself" he can afford to feel happy about someone else, especially a teacher who so clearly sets out objectives that allow the student to master the subject matter and to learn. The writer believes that not all teachers allow their students to learn. The writer believes that the idea of student evaluations is good, but we must obtain some definitive, reliable studies to support their use, as we are using the ratings for very important purposes. If it is true that grades received affect teachers' ratings, these independent variables, the grades, should be identified so that they can be properly weighted or controlled.

V. Definition of Terms

1. William R. Harper is a two year community college offering both career and transfer courses. It is located in the northwestern suburban area of Chicago and is made up mainly of middle class students, having a few upper-lower class students, as well as a few lower-upper class students.

2. Grades that were used in this study were the ones that the students received at mid-term for the Spring 1975 semester. These were chosen as the administration required the faculty evaluation forms, course evaluation questionnaire, to be distributed after the mid-term grades were sent to the students. If grades did have an effect on the ratings, this would have been the time when the effects of the grades would have had the greatest impact on the students' feelings and thus the rating they assign to their instructors.

3. Effectiveness was equated to section nine of the course evaluation questionnaire, where the student rated the instructor in one of five categories: excellent, above average, average, below average, poor. The descriptive terms were translated into a number system assigning five points to the highest, to one for the lowest. The faculty was then ranked on the total mean ratings such as 4.86, 4.20 down to 1.00 or the lowest.

4. The instructions to the students were: "In

relation to other college instructors that you have had, you will rate this instructor as:" (One of the above five categories.)

VI. Limitations of the Study

The exceptions that the writer must make about this study are the following: 1) This referred only to students at William R. Harper College for the Spring, 1975 semester, who were enrolled in the Introduction to Psychology class 101-001. 2) The sample, although adequate and chosen on a systematic random basis, was taken from the population of the Psychology 101-001 large lecture which consisted of approximately 300 students. 3) The student composition of this class appeared to be a representative cross section of the school, so the writer did not see any reason why one could not generalize the findings of this paper to other classes within the institution.

4) The other limitation was that the writer relied on the memory and honesty of the students to recall how they rated their instructors after the mid-term. 5) The writer was not able to ascertain if any instructor gave all C's, etc. to his students at mid-term rather than attempt to make an honest evaluation of each student's work. 6) It was also possible that some students would not accurately remember all their grades that they received at mid-term.

This would be a point to consider in future follow up studies.

7) The writer did not have to conduct a second follow up of questionnaires as they were distributed in the lecture hall and collected the same day. The writer thus received a 100% recovery rate. (However, one questionnaire was spoiled; 75 were distributed and 74 were usable.)

8) The writer requested that the responses be anonymous so as to insure a high degree of confidentiality in response.

The writer was aware that there were other independent variables which the writer had no control over, such as: 9) status of the instructor 10) the instructor's enthusiasm or lack of it 11) his structure of the class presentation 12) text material or other similar factors that other researchers have found to have an effect on student ratings.

VII. Basic Assumptions

1) The students in the population and sample were typical of William Rainey Harper and were not a biased group. 2) There was no bias in the distribution of the I.Q. between the subjects and the population of William

Rainey Harper College 3) There was an equal race and ethnic distribution for college freshmen in Psychology 101 4) There were equal age and sex distributions and 5) Every student had an equal chance to be selected into this sample.

In the absence of any evidence to the contrary, and in view of what is generally assumed about the composition of a general freshman level course at Harper, the writer felt that the students were truly representative of the total Harper population.

VIII. Procedures for Collecting Data

The methodology that was used in this investigation was the submitting of questionnaires to the sample of students of the Psychology 101-001 lecture class on Wednesday, May 21, 1975.[14] (Prior to class exposure, the questionnaire was read by several other students for clarity and ease of comprehension.)

In order to obtain an adequate sample of the population the writer had decided on a 25% systematic sample of the approximately 300 students present in the lecture.[15] A 25% sample of the 300 students was 75 students. The writer selected them by numbering the names on the alpha-

---

[14] See sample questionnaire at end of report

[15] R. F. Mines and George Barton, Applied Educational Research and Evaluation (1973), p. 42.

betical class roster from 1 to 300. The writer then selected every 25th student until the sample of 75 students was obtained. (A systematic sample can be considered to satisfy all requirements of a random sample.)

The exact data that was collected consisted of the students listing, in Column 1, all the courses in which they were registered at mid-term during the Spring, 1975 semester. The listed, in Column 2, the mid-term grade received and in the third column, they rated the instructor in each class on a basis of 1 to 5 as follows: 1, poor; 2, below average; 3, average; 4, good; 5, excellent. The students listed their grades at mid-term as letters, A, B,C,D,F. (I or incompletes were not considered in this study.) The letter grades were transposed into a numerical equivalent of 5 for an A; 4 for a B; 3 for a C; 2 for a D; and 1 for an F.

The time for the collection of the data was 8:00 a.m. on Wednesday, May 21st while the students were in the lecture hall, E-106.

The students were read a passage which was as follows:

> Your instructor is interested in what courses you were enrolled in this semester, the grades you received at mid-term and how you rated the instructors that you had in those courses. On the forms before you, will you please give that information. List your courses in Column 1, grade for each of those courses at mid-term in Column 2. If you felt the instructor was excellent, rate a 5; very good, a 4; average, a 3; below average a 2; and poor, a 1. This rating goes in Column 3.
>
> If you have any questions, please indicate them to your instructor.

IX. Procedures for Treating Data

The writer subjected the raw data to the statistical treatment of a oneway analysis of variance.[16] This was done to ascertain if there was a significant relation between the grades the students earned and the rating given to their instructors.[17] The specific formula that was used was obtained from Hays' Statistics.[18]

Of the 25%, systematic random sample of the Psychology 101-001 population, all but one of the 75 questionnaires were used (25% of 300 is 75). One questionnaire was completed inappropriately and was discarded. Thus, the grand N for the study was actually 74 rather than 75 as reported in the proposal.

The writer had originally established five treatment cells (A,B,C,D, and F). However, due to the insignificant number of F grades at mid-term in Cell #5 (one grade) the writer, under the direction of Dr. G. Rankin, collapsed the

---

16 William L. Hays, Statistics (New York, 1963), p. 677.

17 The significance level was .05.

18 Dr. Gary Rankin, Dean of Students at Harper College, Reader for Nova's Ed.D. program for Community College personnel.

the D and F cells into one, thus increasing the N.

X. Data Resulting From the Study

Of the 74 student questionnaires returned, there were 328 grades reported or approximately an average of 4.43 grades per student.

Of the 328 grades, 104 were A's, 116 were B's, 90 were C's, 18 were D's and 1 was an F.

A preliminary analysis that was made prior to the analysis of variance was to sum the ratings that each student gave to the instructor for each grade given. Each student was instructed to rate each instructor on a scale from 1-5, poor to excellent.

1. Sum of ratings from students who received A grades: 438 divided by 104 = a mean instructor rating for the A grades of 4.211.
2. Sum of ratings from students who received B grades: 446 ÷ 116 = a mean instructor rating for the B grades of 3.844,
3. Sum of ratings from students who received C grades: 310 ÷ 90 = a mean instructor rating for the C grades of 3.444.
4. Sum of ratings from students who received D and F grades: 50 ÷ 18 = a mean instructor rating for the D and F grades of 2.777.

From a cursory observation, we find a declining progression in mean faculty ratings as the grades decline in a corresponding fashion.

The overall mean was $\frac{328 \text{ observations}}{1{,}244 \text{ total ratings}}$ or 3.8

which placed the instructors between the high end of the "average" range towards the low end of the "good" range.

The actual steps involved in the "one way analysis of variance" were as follows:

1. $\sum XA = 104$
   $\sum X^2A = 1,928$

of the students (104) who received an A grade,

49 rated their instructor as a 5

33 rated their instructor as a 4

18 rated their instructor as a 3

3 rated their instructor as a 2

1 rated his instructor as a 1

2. $\sum XB = 116$
   $\sum X^2B = 1,849$

of the students who received a B grade (116)

38 rated their instructor as a 5

41 rated their instructor as a 4

22 rated their instructor as a 3

10 rated their instructor as a 2

5 rated their instructor as a 1

3. $\sum XC = 90$
   $\sum X^2C = 1,178$

of the students who received a C grade (90)

14 rated their instructor as a 5

32 rated their instructor as a 4

31 rated their instructor as a 3

8 rated their instructor as a 2

5 rated their instructor as a 1

4. $\sum XDF = 18$
$\sum X^2DF = 172$

of the students who received a D or F grade,

2 rated their instructor as a 5

4 rated their instructor as a 4

5 rated their instructor as a 3

2 rated their instructor as a 2

5 rated their instructor as a 1

In essence, the writer is comparing the mean ratings of four groups of students.

G 1	G 2
$N_1 = 104$	$N_2 = 116$
$\sum X_1 = 438$	$\sum X_2 = 446$
$\overline{X} = 4.2115$ (Mean)	$\overline{X_2} = 3.8448$ (mean)
$\sum X^2_1 = 1,928$	$\sum X_2^2 = 1,849$
$(\sum X_1)^2 = 191,844$	$(\sum X_2)^2 = 198,916$
$\frac{(\sum X_1)^2}{N} = 1,844.65$	$\frac{(\sum X_2)^2}{N} = 1,714.79$

G 3

$N_3 = 90$

$\Sigma X_3 = 310$

$\overline{X_3} = 3.4444$ (Mean)

$\Sigma X_3^2 = 1,178$

$(\Sigma X_3)^2 = 96,100$

$\frac{(\Sigma X_3)^2}{N} = 1,067.78$

G 4

$N_4 = 18$

$\Sigma X_4 = 50$

$\overline{X_4} = 2.7778$ (Mean)

$\Sigma X_4 = 172$

$(\Sigma X_4)^2 = 2,500$

$\frac{(\Sigma X_4)^2}{N} = 138.89$

$A = 5,127$

$B = 1,244$

$C = 4,766.11$

$$SS_t = A - \frac{(B^2)}{N}$$

$$SS_t = 5,127 - \frac{1,547,536}{328}$$

$SS_t = 5,127 - 4,718.1$

$SS_t = 408.9$

$$SS_b = C - \frac{(B^2)}{N}$$

$$SS_b = 4,766.11 - \frac{1,547,536}{328}$$

$SS_b = 48.01$

$$SS_w = A-C$$

$$SS_w = 5,127 - 4,766.11$$

$$SS_w = 360.89$$

$$MS_b = \frac{SS_b}{(G-1) = 3}$$

$$MS_b = \frac{48.01}{3}$$

$$MS_b = 16$$

$$MS_w = \frac{SS_w}{N\text{-groups}}$$

$$MS_w = \frac{360.89}{328-4}$$

$$MS_w = \frac{360.89}{324}$$

$$MS_w = 1.11$$

$$F = \frac{MS_b}{MS_w}$$

$$F = \frac{16}{1.11}$$ or the calculated F is 14.414

$df = \frac{3}{324}$. The critical F according to the tables[19] is 2.6 at .05 level of significance.

---

[19] Hays, p. 677.

Having obtained a calculated F of 14.414, the null hypothesis can be rejected and the alternative hypothesis can be accepted. The study indicated that there is a significant relationship between the grades a student achieves in a course and how the student rates the effectiveness of his instructor.

Table 1

One Way Analysis of Variance

Summary

Source	S.S.	dF	MS	F
Treatments between groups	48.01	3	16	14.414
Error within groups	360.89	324	1.11	
Totals	408.9	327		

XI. Conclusions

The conclusions that can be drawn from this study include: there is a significant relationship between the grades a student receives and the rating which the student gives the instructor of the course in which the grade was earned. We cannot assume a cause and effect relationship, but there is certainly a

positive correlation: student ratings of instructors tend to go upward as grades in the course move upward and conversely, as grades move downward.

The implications for Harper college and other institutions that tend to tie student evaluations with faculty retention, promotion, rank, etc. is that if an instructor wants to insure himself a high rating, he may do this by making certain that the majority of his students receive high or higher grades than might otherwise be distributed.

This study has implications for the instructors who use the mastery level concept which allows a higher percentage of students, by taking successive exams, to obtain A - B grades, as well as the instructor who grades on a curve with a one-time exam. This would also have implications for the instructors who use the mastery workbook technique to base future exam questions upon. Again, this technique allows for a larger percentage of students to obtain higher grades and this increases the likelihood of the instructor's obtaining a higher rating.

When the administration is reviewing the rating of instructors, does it also consider the method of presentation or the delivery system of the instructor in the classroom? Do the instructor ratings obtained in different

delivery system classes have the same weight? Can they be comparable or should this factor be weighted so that a more equitable evaluation can be made?

Is it possible that the higher ratings are due to a mastery system of teaching that allows more students to learn, thus helping them enjoy the class more and rating the instructor higher as a result? Perhaps it should be so if we are really interested in having students learn.

The writer will present this paper to Dr. D. Carlson, Chairperson, Social Science Division, for discussion within the division; Dr. J. Lucas, Director of Institutional Research for publication in the house publication, Heuristic; and Mr. Robert Powell, President of the Faculty Senate.

XII. Residual Findings

The one bit of data that the writer found especially impressive was the fact that out of 74 students with 328 grades at mid-term, only one student received the lone F grade. If this is from a sample (Psychology 101-001 with 300 students) representative of the entire population, there would only have been four students with one F grade. This impressed the writer as an extremely low figure.

XIII. Further Studies:

1. The writer would like to recommend that a follow up study be conducted, using a larger number of subjects, the entire 101-001 population or a random sample of students from all the freshman level courses at the institution.

2. The writer would also recommend a study of the sophomore level courses to determine if there is any difference between students ratings of instructors at the freshman and sophomore levels.

3. If there is a difference between the two levels, does the administration take this into consideration when ranking two instructors - one above the other by, at times, a few hundredths of a decimal point.

4. Is that type of difference really significant, and is it valid to compare the ratings of a freshman level instructor to that of a sophomore level instructor?

5. The writer would also like to know the specific reasons for a student's rating an instructor low, especially when the student receives a high grade, as was evidenced in some cases of this study.

6. The writer would like to isolate the traits of instructors that tend to be disagreeable to students and result in lower ratings. He would also seek to determine which teaching techniques tend to be disliked by students and produce lower ratings.

7. Having rejected the null hypothesis, the writer would recommend further study of the alternative hypothesis, not the null hypothesis.

8. The writer would recommend a study at the end of the semester to see if these results verify the mid-term results. A correlation could be done between a mid-term grade and a semester's end grade, a study conducted to determine if there is reliability between the two.

Bibliography

1. Costin, Frank; William Greenough and Robert J. Menges. "Student Ratings of College Teaching: Reliability, Validity and Usefulness." Review of Educational Research, 41 (December, 1971), 511-535.

2. Doyle, Kenneth O. and Susan Whitely. "Student Ratings as Criteria for Effective Teaching." American Educational Research Journal, 11 (Summer, 1974), 259-274.

3. Elliott, D. H. "Characteristics and Relationships of Various Criteria of College and University Teaching." Purdue University Studies in Higher Education, 70

4. Frey, P. W. "Student Ratings of Teaching: Validity of Several Rating Factors." Science, 182 (1973), 83-85.

5. Hogan. T. P. "Similarity of Student Ratings Across Instructors. Courses and Time." Research in Higher Education, I (1973), 149-154.

6. Isaacson, Robert L., Wilbert McKeachie and John Milholland. "Correlation of Teacher Personality Variables and Student Ratings." Journal of Educational Psychology, 54 (April 1963), 110-117.

7. Kerlinger, F. N. "Student Evaluation of University Professors." School and Society, IC (1971), 353-356.

8. Kiernan, Irene R. "Student Evaluations Re-Evaluated." Community College Journal, VL (April, 1975), 25-27.

9. Kulick, J. A. and C.L.C. Kulik. "Student Ratings of Instruction." Teaching of Psychology, I (December, 1974), 50-56.

10. McKeachie, W. J. "Student Ratings of Faculty." AAUP Bulletin, 55 (1969), 439-444.

11. McKeachie, W. J. and Yi-Guang Lin. "Student Ratings of Teacher Effectiveness; Validity Studies," American Educational Research Journal, 8 (May, 1971), 435-445.

12. Mines, R. F. and George Barton. Applied Educational Research and Evaluation. Ft. Lauderdale; Nova University, 1973.

13. Murray, Harry. "Predicting Student Ratings of College Teaching from Peer Ratings of Personality Types." Teaching of Psychology, 2 (April, 1975), 66-69.

14. Naftulin, Donald; John E. Ware Jr. and Frank Donnelly. "The Doctor Fox Lecture: A Paradigm of Educational Seduction." Journal of Medical Education, 48 (July, 1973), 630-635.

15. Rodin, M. and B. Rodin. "Student Evaluations of Teachers." Science. CLXXVII (1972), 1164-1166.

16. Remmers, H. H. "Teaching Methods in Research on Teaching," in Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

17. Remmers. H. H. and Martin, F. D. and Elliott, D. N. "Are Students' Ratings of Instructors Related to Their Grades?" Purdue University Studies in Higher Education, 66 (1949), 17-26.

18. Hays, William L. Statistics, New York: Holt, Rinehart, Winston, 1963.