Relationship of variables beyond teachers control and teachers effectiveness ratings by students
by Dennis Richard Haley

A thesis submitted in partial fulfillment of the requirements for the degree of DOCTOR OF EDUCATION
Montana State University

Abstract:
This study investigated the relationships' between student ratings of teachers and a variety of variables over which teachers have little or no control. The research was conducted during the 1973-1974 academic year at Montana College of Mineral Science and Technology and utilized as many as 1,800 student ratings of forty-three teachers at that college.

More specifically, the ratings the teachers received were correlated with each of the following: day of the week on which ratings were conducted, period class was held, level of the course, sex of the teacher, number of students in the class, difference between sex of the teacher and the student, cumulative grade point average of the student, number of credits student was taking, number of years student had known the teacher prior-to rating, number of exams the student had taken on day of rating, age of student, whether the course was required or elective, sex of student, the grade the student expected in the course, whether student's major was the same or different than the teacher's, student's attitude toward the classroom, student's attitude toward occupation of teaching, student's attitude toward the hour the class was held, student's attitude toward teacher evaluations, student's attitude toward Montana Tech, student's attitude toward topics of the course, student's attitude toward grade expected in the course, and student's attitude toward the way he felt at time of rating. Correlations between ratings and the above listed variables were found to be significantly different from zero for all variables except: the day of the week on which ratings were conducted, the period during which the class was held, the level of the course, the sex of the teacher, cumulative grade point average of student, the number of years the student had known the teacher, and the number of exams the student had taken on the day ratings were conducted.

In addition a multiple regression analysis was made using the twenty-three variables listed above as independent variables and teacher rating as the dependent variable. From this analysis a multiple R coefficient of correlation of .69 between the twenty-three independent variables and the teacher rating variable was found. The coefficient was found to be significantly different from zero at the α=.01 level.

Using the above mentioned regression formula to predict the portion of each teacher's rating attributable to the twenty-three variables, each teacher's rating was then adjusted by subtraction of this predicted value. Comparison of teacher ranks compiled from unadjusted ratings to teacher ranks compiled from adjusted ratings showed significant changes in rank for one-third of teachers involved.

RELATIONSHIP OF VARIABLES BEYOND TEACHER'S CONTROL AND TEACHER'S EFFECTIVENESS RATINGS BY STUDENTS

by

DENNIS RICHARD HALEY

A thesis submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF EDUCATION

Approved:

Earl N. Ringo
Chairman, Examining Committee

Robert J. Hiferault
Head, Major Department

Henry L. Parsons
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

August, 1974

ACKNOWLEDGEMENTS

I wish to thank my chairman, Dr. Earl Ringo, and the remainder of my graduate committee for their positive support throughout the writing of this paper.

Special thanks is extended to Dr. Ringo's secretary, Mrs. Carolyn Winchester, who provided invaluable assistance in meeting the multitude of departmental and university requirements associated with the writing of this dissertation.

Thanks is also given to the administration of Montana College of Mineral Science and Technology for allowing the study to be conducted in their institution.

I am especially indebted to Dr. Jack Goebel, Chairman of the Department of Mathematics at Montana College of Mineral Science and Technology, whose special consideration in assignment and scheduling of teaching duties allowed me to pursue this degree.

And most of all I wish to thank my wife, Miriam, who in addition to typing the dissertation has given encouragement throughout my graduate work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study investigated the relationships between student ratings of teachers and a variety of variables over which teachers have little or no control. The research was conducted during the 1973-1974 academic year at Montana College of Mineral Science and Technology and utilized as many as 1,800 student ratings of forty-three teachers at that college.

More specifically, the ratings the teachers received were correlated with each of the following: day of the week on which ratings were conducted, period class was held, level of the course, sex of the teacher, number of students in the class, difference between sex of the teacher and the student, cumulative grade point average of the student, number of credits student was taking, number of years student had known the teacher prior to rating, number of exams the student had taken on day of rating, age of student, whether the course was required or elective, sex of student, the grade the student expected in the course, whether student's major was the same or different than the teacher's, student's attitude toward the classroom, student's attitude toward occupation of teaching, student's attitude toward the hour the class was held, student's attitude toward teacher evaluations, student's attitude toward Montana Tech, student's attitude toward topics of the course, student's attitude toward grade expected in the course, and student's attitude toward the way he felt at time of rating. Correlations between ratings and the above listed variables were found to be significantly different from zero for all variables except: the day of the week on which ratings were conducted, the period during which the class was held, the level of the course, the sex of the teacher, cumulative grade point average of student, the number of years the student had known the teacher, and the number of exams the student had taken on the day ratings were conducted.

In addition a multiple regression analysis was made using the twenty-three variables listed above as independent variables and teacher rating as the dependent variable. From this analysis a multiple R coefficient of correlation of .69 between the twenty-three independent variables and the teacher rating variable was found. The coefficient was found to be significantly different from zero at the $\alpha=.01$ level.

Using the above mentioned regression formula to predict the portion of each teacher's rating attributable to the twenty-three variables, each teacher's rating was then adjusted by subtraction of this predicted value. Comparison of teacher ranks compiled from unadjusted ratings to teacher ranks compiled from adjusted ratings showed significant changes in rank for one-third of teachers involved.

CHAPTER I

INTRODUCTION

SETTING

The recent trend toward accountability and efficiency in teaching has produced an increased effort to determine factors for which teachers should be held accountable and accompanying this has been an expanded effort in the process of evaluation of teacher effectiveness. One such evaluation process that has been expanded is the process of teacher ratings. The use of ratings to assess teacher effectiveness is, however, not a new idea. According to Morsh and Wilder (1954) the administrative use of ratings in the measurement of teacher effectiveness dates back as far as 1896 when, in the city of Milwaukee, they were used for the first time. Since that time a variety of rating schemes have been developed. These ratings have been conducted primarily at the secondary and post-secondary levels with the principal raters being: the teacher himself, the administrators, fellow teachers and students (Morsh and Wilder, 1954:13).

Although student ratings of teachers have received their greatest use in recent years, they have been in evidence for over forty years. In fact formal ratings of teachers by students dates back as far as the 1920's when H. H. Remmers began his research

using the Purdue Rating Scale (Eble, 1970:12). The use of student ratings has grown steadily, and as indicated by a survey taken in 1961, these ratings are now being used in a variety of different colleges, universities and junior colleges (Gustad, 1961:10). Because of this growing use of student ratings of teachers there appeared at the time of this study a need to answer the many questions that occur in regard to the use of these ratings. Should ratings be mandatory for all teachers? Who should see these ratings? Should these ratings be used to determine salary and promotion? Are ratings reliable? Are ratings valid?

This study was designed to investigate a particular aspect of validity and reliability questions; namely, the reduction of rating reliability and validity due to the subjectivity of the rater and other variables over which teachers have little control. In this regard the study was restricted to the relationship between student ratings of instructors and various student attitudes and contextual variables present at the time of rating. In no sense other than this did this study attempt to determine whether or not students can validly rate teacher effectiveness. Furthermore, since student ratings of teachers usually have been carried out in a given department, division, or institution, this writer believes that research on these ratings should be applied primarily to the

particular setting in which they are conducted. Because of the diversity of departments, divisions, and institutions, and also of students attending these various departments, divisions, and institutions, this writer also believes that generalizations about student ratings of teachers are of questionable or limited value. And, since the use of these ratings usually has been confined to the given department, division, or institution in which they were conducted, it follows that research on ratings should be of particular value to the institution in which the ratings are conducted. Thus, the principal efforts of this investigation were directed toward student ratings at a specific institution, Montana Tech (Montana College of Mineral Science and Technology). And although this study was designed in part to improve the validity and reliability of ratings at this institution, it was hoped that the approach and the model developed in the study would be of value to other institutions and researches.

Montana Tech (formerly Montana School of Mines) is located in Butte, Montana; and at the time of this study had a student enrollment of 720, had a teaching faculty of 49, and offered Baccalaureate Degrees in mineral engineering, mathematics, chemistry, English, and history. The college provided an ideal location for this research in that formal student ratings had been used at the school for three years prior to the study and had been made mandatory

of all teaching faculty by faculty vote. The faculty had also voted to allow the results of these ratings to be used by the Academic Dean, the President, and the Committee on Salary and Promotion. In addition to this, the Committee on Teacher Effectiveness had been using these student ratings each year for the past three years to determine the three recipients of the Outstanding Teacher Awards, i. e. awards granted to the college by Standard Oil Company, Indiana Foundation, in the amount of $1,000. each. These ratings had, also, in the past been used by the teachers at Montana Tech as a means of improving their teaching. In this regard a great amount of time and money had been spent in developing the rating form and in reporting the results of the ratings to the teachers. In fact the principal reason for the development of the rating process at Tech and for the undertaking of this study was to provide faculty members with a more valid, reliable, and generally useful information on their teaching. A sample copy of the rating report issued to each teacher is exhibited in APPENDIX D. Included in these reports were student comments about things the teacher did "especially well" and about "things that might be done to improve the course." These comments were recorded on the last page of the rating form (APPENDIX A) and were returned to the teachers as part of the rating report.

FORMAT

This paper is divided into the following divisions: CHAPTER I--INTRODUCTION, CHAPTER II--REVIEW OF RELATED LITERATURE, CHAPTER III--PROCEDURES, CHAPTER IV--FINDINGS, CHAPTER V--A MODEL, CHAPTER VI--CONCLUSIONS AND RECOMMENDATIONS, and APPENDIXES.

The remainder of CHAPTER I includes a statement of the problem, a discussion of the need for the study, questions to be answered, a statement of the general procedures followed in the study, a discussion of limitations and definition of terms.

CHAPTER II, REVIEW OF RELATED RESEARCH, investigates research on rating of teacher effectiveness. The chapter reviews only those studies that deal with student ratings of teachers. Only quantitative studies are related.

CHAPTER III, PROCEDURES, presents a detailed discussion of the design of this research. Included is a description of the population and sampling procedure, data collection procedure, design and discussion of rating instrument used, discussion of variables involved and analysis performed, and a report of statistical tests used.

CHAPTER IV, FINDINGS, reports the results and interpretation of this study. The results are presented in table form and interpreted in narrative as they are reported.

CHAPTER V, A MODEL, includes the presentation of a multiple linear regression model designed to detect and eliminate "biases" or undesired linear relationships between student variables and student ratings. Validation of the model using techniques developed by Mosier (1951) is also reported in this chapter.

CHAPTER VI, CONCLUSIONS AND RECOMMENDATIONS, as the title indicates is a report of the writer's conclusions and recommendations in this study.

APPENDIX A of this paper contains a copy of the teacher rating form and data collection instrument used in this research. APPENDIX B shows PART III and PART IV of the rating instrument with the scales of the semantic differential all oriented the same direction. Further explanation of this altered instrument and its relationship to the instrument shown in APPENDIX A is presented in CHAPTER III--PROCEDURES. APPENDIX C contains a copy of the envelope in which rating forms were distributed. This copy exhibits the various data recorded on these envelopes. APPENDIX D contains a sample of the student rating report that was returned to each faculty member participating in the rating process during fall semester, 1973.

STATEMENT OF THE PROBLEM

The problem investigated in this research was what, if any, is the relationship between the rating a student gives his instructor

and certain student and contextual variables. Those variables investigated were the student's grade point average, the number of credits the student is taking, the number of years the student has known the teacher, the number of exams the student has taken on the day the ratings are made, the age of the student, whether the course is required or elective, the sex of the student, the grade the student expects in the course, whether the student's major is the same or different than the teacher's, the student's attitude toward the classroom, the student's attitude toward the hour the class is held, the student's attitude toward teacher evaluations, the student's attitude toward the school, the student's attitude toward the topics in the course, the student's attitude toward the grade he expects to receive in the course, the general way the student feels at the time of rating, the day on which the ratings are held, the time of day the ratings are held, whether the student's sex is the same as the teacher's, the number of students in the class, and the level of the course.

In addition this research also investigated the problem of what, if any, is the relationship between the rating a student gives his teacher and certain combinations of the student and contextual variables listed in the previous paragraph.

The problem of reducing or eliminating these relationships through the use of a statistical model was also investigated.

NEED OR PURPOSE OF THE STUDY

House Bill 109, a bill defeated in the 1973 session of the Montana legislature, would have made teacher evaluations mandatory throughout the state. Furthermore, a meeting held by the Montana Superintendent of Public Instruction in the spring of 1973 was directed exclusively to the topic of teacher evaluations. These were both indicators that uniform teacher evaluations in Montana may be in the offing in the near future. Whether this will happen, or if it does happen, whether student ratings will play a part, is uncertain; but in any case there was a need for this research. At the time of this writing at least three units of the Montana University System were using student ratings in some fashion. Those were The University of Montana, Montana State University, and Montana College of Mineral Science and Technology. In at least one of these institutions, Montana Tech, these ratings were used, in part, as a means of determining salary and promotion. Since these ratings were used in this fashion it was important that more research be done on the rating process to determine what relationships exist between these ratings and the student and contextual variables stated previously.

Other researchers who have investigated the problem of teacher evaluations and in particular student ratings of teachers stress the need for further research on the reliability and validity

of these ratings. "Overall, general ratings of teacher effectiveness have been shown to be, under certain conditions, exceedingly unreliable. Depending upon one's point of view, this unreliability provides a substantial road block or a challenge to the researcher interested in this area of research (Barr, 1961:8)." Specific variables related to reliability and validity of ratings were cited by Gage.

> In appraising teaching to obtain a basis for administrative decisions on academic rank, tenure, and salary, one has to make sure, above all, that the appraisal is fair. This means that teachers should not be penalized because of conditions over which they have no control, such as the level of the course (for example, undergraduate or graduate), the size of the class, whether the course is elective or required, and where it is taught (on campus or off campus) (Gage, 1961:17).

In 1968 N. F. Rayder conducted a research study within the School of Education at Colorado State College. His study found age, grade level, GPA, and course grades of students not to be correlated with their ratings of instructors. He pointed out, however, in his concluding remarks that there were still unanswered questions.

> We need to know what student characteristics, if any, do influence their rating of instructors. We need to know if a parallel study performed within another academic area or at another institution will yield similar results (Rayder, 1968:81).

There was yet a more important need for research of the kind conducted in this study. That need developed from the use of rating techniques in educational research studies. In fact, a review of the literature on teacher effectiveness by Morsh and Wilder (1954)

indicated a wide use of rating techniques in measurement of teacher performance prior to 1952. In discussing this wide use of ratings Morsh and Wilder concluded that "by and large investigators have tended to ignore the problems of correcting for the various sources of error and have worked with ratings as though they were already a perfected criterion (13:1954)." Important recent studies on teacher effectiveness by Ryans (1950) and by Combs (1960) indicated that rating instruments were still being used as research tools. Even though these ratings were often conducted by trained "experts" there appeared to be little research dealing with what might be termed the error of rater subjectivity or bias. In this regard Biddle and Ellena stated:

> Ratings, unfortunately, offer the investigator an unexamined hodgepodge of classroom relationships and confusions stemming from rater's biases and lack of information. It is quite possible, of course, that ratings by a trained and unbiased person sometimes may be valid. However, until careful studies are made of the facts and artifacts involved in the rating process, ratings seem less than useful for research on teacher effectiveness (Biddle and Ellena, 1964:27).

This study attempts to determine some of these "facts and artifacts." Thus, even though this research was of particular interest to Montana Tech it might also be of interest to anyone using rating techniques in educational research. Furthermore, the model developed and the general approach to student ratings of instructors exhibited in this study should also be of value to other

institutions and educational researchers.

GENERAL QUESTIONS ANSWERED

The following questions (with reference to students and teachers at Montana Tech) were answered in this study:

1. Is a student's grade point average related to the way he rates his teacher?

2. Is the number of credits a student is taking correlated with the rating he gives his teacher?

3. Is there a relation between the number of years a student has known the teacher and the rating he gives the teacher?

4. Is the number of exams a student has had on the day of ratings related to his rating of the teacher?

5. Does the age of the student relate to the rating of his teacher?

6. Is there a relation between whether a course is required or elective and the rating of the teacher?

7. Is the sex of a student related to the rating of the teacher?

8. Is the grade a student expects to receive in a course related to the rating he gives his teacher?

9. Is there a relation between the major of the student and the major of the teacher and the rating the student gives the teacher?

10. Is there a relation between the student's attitude toward the classroom and the student's rating of his teacher?

11. Is the student's attitude toward teaching in general related to the rating he gives his teacher?

12. Is the student's attitude toward the hour the class is held related to the rating he gives the teacher?

13. Is the student's attitude toward teacher evaluations related to the rating of his teacher?

14. Is the student's attitude toward the school related to the rating he gives his teacher?

15. Is the student's attitude toward the topics of the course related to the rating he gives his teacher.

16. Is the student's attitude toward the grade he expects to receive in the course related to his rating of the teacher?

17. Is the way the student feels at the time of the rating related to the rating?

18. Is the day on which ratings are held related to the ratings?

19. Is the time of day the ratings are held related to the ratings?

20. Is whether or not the student's sex is the same as the teacher's related to the student's rating of the teacher?

21. Is the number of students in the class related to the

rating of the teacher?

22. Is the level of the class related to the rating of the teacher?

A more general and perhaps, more important question investigated in this research is: What, if any, is the relationship between student ratings and certain combinations of the student-contextual variables just listed? In addition this study investigated the possibility of altering this relationship by adjusting the ratings.

GENERAL PROCEDURE

The general procedure of this research consisted of administering the rating form displayed in APPENDIX A to classes being taught at Montana Tech during fall semester, 1973. The data was analyzed and the questions stated in the previous section answered.

The reliability of the rating form was investigated earlier in fall semester, 1973, using a test-retest application of the rating form in a lower division mathematics class, a lower division biology class, an upper division engineering class and an upper division history class at Montana Tech. A period of two weeks between test and retest was used. The validity of certain items (items requesting demographic data) was also investigated using the data collected from the reliability study. This was done by

comparison with the Registrar's records.

A major problem of this research centered around the development of a rating instrument that included the measurement of the student attitude variables stated in the previous section. As shown in the APPENDIX A the final rating form utilized the Semantic Differential for the measurement of these variables. PART IV of the instrument, the teacher rating scales, also utilized the Semantic Differential. A more detailed discussion of the rating instrument is given in CHAPTER III.

The analysis consisted of correlating each of the student and contextual variables listed in this chapter with each of the various teacher rating variables shown in PART IV of the rating form. In addition these student and contextual variables were used in combination as independent variables in a multiple regression analysis. The dependent or criterion variable for this analysis was taken as the combined teacher rating as measured by an average rating score on PART IV of the form exhibited in APPENDIX A. Validation of this regression model was conducted by use of the "cross-validation" techniques of Mosier (1951). An extension of this model was then used in an attempt to reduce the multiple linear relationship just mentioned.

LIMITATIONS

A basic limitation of this study is it's lack of generalizability. Relationships between student-contextual variables and teacher ratings demonstrated at Montana Tech may not exist at another institution. This writer believes, however, that this limitation is no greater for this study than for a study using a larger sample of a more diverse population. This follows since findings from such a generalized study would probably be of no more use to a particular institution than would the findings of this study. The rating bias model developed and the general approach should, however, be of value to any institution or department using formal student ratings.

Of course the regression model developed at Montana Tech without modification would be of limited use to another institution or department since it is probable that the student and contextual variables vary from one setting to another. Also, since the model is dependent upon voluminous amounts of data generated by the measurement of the student and contextual variables, it is apparent that it would be of limited use to departments or institutions lacking access to sophisticated computing facilities and technical resource people.

A delimitation was made in the selection of student and contextual variables for the investigation. Certainly there were many other such variables that could have been investigated. This

delimitation was necessary, however, to keep the length of the rating form manageable and to ensure that excessive class time was not required for the administration of the form. Even with this delimitation the coding and analysis of the data generated required over 80 man hours of keypunch time and over 200 hours of central processor time on the IBM 1620 computer. Many of the variables omitted could be included as a part of the evaluation model in subsequent evaluations. The model developed should not be a static one but should evolve as student and teacher variables are added and as these variables change with time.

DEFINITION OF TERMS

Student Variable

In this paper a student variable is any quantifiable student characteristic, trait, or behavior.

Teacher Variable

A teacher variable is any quantifiable teacher characteristic, trait, or behavior.

Contextual Variable

A contextual variable is any classroom or situational characteristic that is not a student variable and is not controllable by the teacher.

Attitude Variable

"A relatively enduring system of evaluative, affective reactions based upon and reflecting the evaluative concepts or beliefs which have been learned about the characteristics of a social object or class of social objects (Shaw and Wright, 1967:3)."

SUMMARY

Although student ratings of teacher effectiveness have been used in various institutions for a considerable time there are indications that various aspects of this practice are in need of research. If these ratings are to be used as a means of ranking teachers for salary and promotional purposes, then there is a need to investigate what relationships, if any, exist between these ratings and the variables that cannot be controlled by the teacher. This study investigated such relationships.

Since this study was conducted using teachers and students at Montana Tech, the results are of special value to that institution. Also, since certain of the student and contextual variables to be studied have been investigated in other colleges, this research should provide added insight into any general relationships that might exist between these variables and student ratings. In addition this study developed a general approach or model that could be altered from institution to institution and that could, hopefully,

reduce or eliminate certain student and contextual biases in ratings of teachers.

CHAPTER II

REVIEW OF RELATED RESEARCH

INTRODUCTION

Evaluation of teachers and teaching, an ever present process in education, has received growing emphasis in recent years. Unfortunately, this evaluative process often has been unsystematic and lacking in objectivity and has often resulted in decision made through intuition, rationalization, tradition, or prejudice. In discussing this problem, Dressel (1961:6) stated that "such patterns of decision making are not consistent with the aims of education, particularly with those of higher education, which in our culture are based upon the assumption that informed judgements can and should be the wiser judgements."

Student evaluations of teachers and the use of these evaluations also have been unsystematic and lacking in objectivity. How often, in fact, have decisions about a teacher's teaching effectiveness been based entirely on second-hand information, information obtained primarily through informal student opinion? This does not say, however that student evaluation of teachers has been totally without systematic use or that such evaluations have not been the subject of detailed research. Riley et al. (1950) reported that formal student ratings of teachers were used as early as 1922 in the

School of Education of Oklahoma A & M. From that time student evaluation of teachers continued to grow, receiving its greatest impetus in the late nineteen-twenties when H. H. Remmers (1930) initiated his research using the Purdue Rating Scale for Instruction. Since then, formal student ratings have become the topic of heated argument between advocates and opponents. These arguments for the most part have centered about the questions of reliability and validity or, in other words, the student's ability to judge effective teaching. One result of this argumentation has been a steady flow of research on the rating process, especially the reliability and validity aspects.

Because of the voluminous research on student ratings of teachers, this review has been limited to include only the quantitative research studies. More specifically, the chapter concentrates on those studies which have investigated relationships between student ratings and the various student and contextual variables listed in CHAPTER I.

CHAPTER II is divided into the following major catagories: INTRODUCTION, RESEARCH FINDINGS, ANALYSIS OF RESEARCH, and SUMMARY. The remainder of the INTRODUCTION describes the format for CHAPTER II. The section entitled RESEARCH FINDINGS, which gives a detailed account of research related to student ratings of teachers, is further subdivided according to the various student and contextual variables

that have been researched in relation to student ratings. These subdivisions are entitled: Student's Grades, Student's Grade Point Average, Class Size, Class Level, Sex of Student, Required vs. Elective Courses, Student Attitudes, Multiple Relationships, and Other Variables. Research findings presented in each of these subdivisions are presented in chronological order and in the case of long presentations are summarized at the end of the subdivision. It should be noted that many of these research studies involved more than one of the variables just categorized. Thus, to avoid duplication, the research methodology used in these studies is not given detailed discussion in each subdivision. This information, however, is important in judging the validity of the research findings and is contained in Table 1 on page 22. This table summarizes all research related in this review and includes: the name of the investigator, the date of publication, the size of student and teacher sample used, the location of the study, the student and contextual variables investigated, the kind of analysis used, the rating instrument used, and the reliability of the rating instrument.

A general analysis of the research cited is given in the division of CHAPTER II entitled ANALYSIS OF RESEARCH. The ANALYSIS OF RESEARCH section is divided into the following subsections: Representativeness of Samples, Rating Instruments Used, and Time of Study. Included in this section is a discussion

Table 1

Research Related to Student Ratings of Teachers

Researcher & Date	Sample Used	Location of Study	Variables Studied	Analysis Used	Rating Instrument	Reliability of Instrument
Anikeef (1953)	1,500 Students rated 19 Teachers	Oklahoma A & M College	Student Grade	Correlational	Researcher Made; 8 point Graphic	Not Reported
Bausell and Magoon (1972)	500 Students rated 35 Teachers	University of Delaware	Student Grade; Student GPA; Discrepency between Grade and GPA	Analysis of Variance	Modified Purdue Rating Scale; 5 pt.-bipolar; 29 items	.94 Spearman-Brown technique
Bendig (1952)	67 Students rated 2 Psych. Teachers	University of Pittsburg	Class Level; Student Sex	Analysis of Variance	Modified Miami University Scales; 5 pt.; (14 item)	Not Reported
Bendig (1953)	132 Students rated 6 Psych. Teachers	University of Pittsburg	Student Achievement level	Correlational	Purdue Rating Scale; graphic; (10 item)	.60-.85 by various techniques
Bledsoe and Others (1971)	4,368 Students rated 180 Secondary Teachers	9 Secondary Schools in Georgia	Student Grades; Class Level; Student Sex; Student Age	Analysis of Variance	Pupil Observation Report; (38 item)	.66-.84 Test-Retest
Bowman (1934)	X Students rated 30 Student Teachers	Green Castle Indiana	Student Grade	Correlational	3 Different Instruments	.84-.91 Chance Half
Brookover (1940)	X Students rated 5 Teachers	5 Indiana High Schools	Student Sex; Class Level	Not Reported	Modified Purdue Rating Scale	Not Reported
Brookover (1945)	X Students rated 66 Teachers	12 Indiana High Schools	Length of Time Student Knew Teacher	Correlational	Researcher Made; 5 pt.	Not Reported

Table 1 (continued)

Researcher & Date	Sample Used	Location of Study	Variables Studied	Analysis Used	Rating Instrument	Reliability of Instrument
Bryan (1937)	1,500 Students rated 63 Secondary Teachers	Brooklyn and Cincinnati	Student Grade; Student Sex	Correlational	Researcher Made;	.75-.96 Chance Half
Cornwell (1974)	X Students rated 70 Teachers	20 College Chemistry Departments	Class Size; Hour of Class; Age of Teacher; Class Grade	Multiple Regression Analysis	Researcher Made; 5 pt. Likert Scales; (6 items)	Not Reported
Davenport (1944)	1,250 Students rated 51 Secondary Teachers	Middlewest High School	Student Sex; Student Age; Class Level; Semesters with Teacher	Difference of Proportions; Comparison of Means	Researcher Made; 5 pt; Graphic; (25 item)	.86 Half Sample Comparison
Drucker and Remmers (1951)	251 Students and 138 Alumni rated 92 Teachers	Purdue University	Alumni vs. Student Ratings	Comparison of Means	Purdue Rating Scale	.60-.85 by various techniques
Heilman and Armentrout (1936)	2.115 Students rated 46 Teachers	Colorado College of Education	Class Size; Student Grade; Class Level; Required-Elective	Correlation and Comparison of Means	Purdue Rating Scale	.75 Split Half
Jackson and Fuller (1966)	128 Students rated 28 Female Teachers	Secondary Schools	Student Social Class; Student Sex	Analysis of Variance	Pupil Observation Report	Not Reported
Overturf and Price (1966)	10,000 ratings in 500 Classes	St. Johns River Junior College	Student Grade; Student GPA: Hour of Class	Correlational; Comparison of Means	Institution Made; 5 pt.	Not Reported

Table 1 (continued)

Researcher & Date	Sample Used	Location of Study	Variables Studied	Analysis Used	Rating Instrument	Reliability of Instrument
Rayder (1968)	4,285 Students rated 87 Teachers	Colorado State College	Student Age; Student Sex; GPA; Prior Grade from Teacher	Correlational	Researcher Made; 7 pt.	Not Reported
Remmers (1929)	8,609 Students rated 115 Teachers	Purdue University	Class Level	Comparison of Means	Purdue Rating Scale	.60-.85 by various techniques
Remmers (1930)	409 Students rated 11 Teachers	Not Reported	Student Grade	Correlational	Purdue Rating Scale	.60-.85 by various techniques
Remmers and Elliot (1949)	26,014 Students rated 460 Teachers	10 Different Colleges	Student Sex; Class Level; Student Grade; Veteran vs. Non-Veteran	Comparison of Means	Purdue Rating Scale	.60-.85 by various techniques
Remmers and Others (1941)	X Students rated 40 Student Teachers	West Liverty West Virginia	Student's Scholastic Standing	Correlational	Purdue Rating Scale	.60-.85 by various techniques
Remmers and Others (1949)	X Students rated 37 Chemistry Teachers	Purdue University	Student Grades	Comparison of Means	Purdue Rating Scale	.60-.85 by various techniques
Riley et al. (1950)	1,258 Students rated 384 Teachers	Brooklyn College	Student Grade; Class Size; Required vs Elective; Student Sex; Class Level; Student Attitude	Comparison of Means	Researcher Made; 4 pt; (10 item)	Not Reported

Table 1 (continued)

Researcher & Date	Sample Used	Location of Study	Variables Studied	Analysis Used	Rating Instrument	Reliability of Instrument
Starrak (1934)	X Students rated Entire Faculty	Iowa State College	Student Grade; Class Size; Class Level	Correlational	Researcher Made; Graphic; (17 item)	Not Reported
Stewart and Malpass (1966)	1,975 Students rated 67 Teachers	University of South Florida	Student Grade; Class Level	χ^2 Test of Imdependence	Institution Made; 5 pt.; (37 items)	Not Reported
Weaver (1960)	699 Students rated 12 Teachers	Central Michigan University	Student Grade	Comparison of Means	Institution Made; 5 pt.; (19 items)	Not Reported
Weigel and Others (1971)	331 Students rated 4 Teachers	Psych. Dept. Colorado State University	Student Grade	Nonparametric Sign Test	Researcher Made; 5 pt; (23 items)	Not Reported

of the various techniques used in the design of rating instruments, and also, in the determination of the validity and reliability of such instruments. This section also includes a discussion of the various research designs used. Time and location of the studies, representativeness of the samples, and appropriateness of statistics used are the major topics included. The final major division of CHAPTER II is the SUMMARY. In this summarization an attempt is made to draw together the seemingly divergent research findings relative to student ratings, and also, to relate these findings to this present study.

RESEARCH FINDINGS

Student's Grade in Course

The question of whether or not a student's grade in a course is related to the rating the student gives the teacher has been and still remains the most often argued aspect of student ratings. For over forty years opponents of student ratings of teachers have argued that the teacher who gives "low" grades receives "low" ratings, and likewise, the student who receives a "low" grade gives the teacher a "low" rating. On the other hand, advocates of student ratings have argued that ratings are not related to grades. Researchers, in attempting to answer this question, have approached the problem in a variety of ways: some have studied the relationship

between the final grade the student received and the rating; some have studied the relationship between the grade the student expected at the time of rating and the rating; and some have studied the relationship between grades received in prior courses and the rating.

H. H. Remmers was, apparently, the first researcher to show an active interest in student ratings of teachers, and consequently, was instrumental in much of the research to date. Included in Remmers' work was the development of the Purdue Rating Scale for Instruction, a rating form that has undoubtable become the most widely used and copied instrument of its kind. It was, in fact, a result of Remmers' reliability and validity research on the Purdue Rating Scale for Instruction that research dealing with the relationship between student grades and student ratings began to appear. Remmers (1930) in reporting his first investigation of this relationship computed the biserial-r coefficient of correlation between students' ranks by grade and teachers' ranks as a result of student rating; the relationship was not found to be significant (Remmers, 1930:314). (For further information on this study and other studies in this review see Table 1 on page 22.)

This initial study by Remmers was followed closely with studies by Bowman (1934), Starrak (1934), Heilman (1936), and Bryan (1937). Bowman (1934) reported that secondary students' ratings

of student teachers were not related to the final grade received or to the differential between grade received and previous grades received in the same subject area. Starrak (1934) using ratings performed on the entire faculty at Iowa State College also reported the finding of no significant relationship between final grades received and student ratings. Heilman and Armentrout (1936) obtained similar results when they found the teacher's grading severity not to be correlated with student ratings. The teacher's grading severity was measured by computing the mean of all grades assigned for three quarters by each of the forty-six teachers involved. These researchers (Heilman and Armentrout, 1936:211) did conclude, however, that teachers giving lower grades tended to receive higher ratings on the item dealing with fairness in grading. Bryan (1937), unlike these other early researchers, found the grades received by 2,118 secondary students to be positively correlated with the ratings these students gave their teachers; and as a result he concluded that students receiving higher marks tended to give the teachers higher ratings (Bryan, 1937:70).

During the decade of the 1940's Remmers, as a part of his continued work on the Purdue Rating Scale for Instruction conducted further research on the student grade--student rating relationship. Beginning in 1941 with a correlational study involving secondary student ratings of student teachers conducted by Remmers, Ward and

Schmalzried (1941) further support was given to the argument that grades and ratings are not related. Again in 1949 the relationship between student grades and student ratings of teachers was the topic of research reports. Remmers and Elliot (1949), using ratings conducted in ten different colleges and 460 different instructors, reported that students in the upper scholastic half of their class gave the teacher no different ratings than did students in the lower scholastic half of the class. In yet another study during this same period Remmers, Martin and Elliot (1949) compared the ratings of Purdue University chemistry teachers who gave a mean grade for the class which was higher than the predicted mean grade for the class with chemistry teachers who gave a mean grade for the class which was lower than the predicted mean grade for the class. Having found the ratings of chemistry teachers giving a mean grade higher than the predicted mean grade to be significantly greater than the ratings of those chemistry teachers giving a mean grade lower than the predicted mean, the researcher did not, however, attribute this difference to a defect in the rating process. Instead these researchers proposed that the difference in ratings, was due, perhaps, to a true difference in teaching effectiveness; it was here argued that the better teachers were those giving a mean grade higher than the predicted mean grade (Remmers, Martin and Elliot, 1949:20).

As just related these early studies gave support to the

argument that student grades and student ratings of teachers are not related. Beginning in the 1950's, however, a substantial amount of research reporting quite different findings began to appear. Riley et al. (1950) in reporting the results of an extensive study conducted at Brooklyn College concluded that superior students tended to give teachers higher ratings than did average students, and that average students tended to give higher ratings than did poor students. In this study superior students were defined as those who received a B or better in the course; average students were defined as those who received between a C plus and a B minus; and poor students were defined as those who received a C or less (Riley et al., 1950:85). Anikeef (1953) divided the ratings of 19 college teachers into two groups: 1) ratings by freshmen and sophomores and 2) ratings by juniors and seniors. From this grouping he reported a highly significant correlation between student grades and student ratings for the freshman-sophomore group. In addition he concluded that at least 53% of the rating variance for freshman-sophomore students was attributable to grades received. No relationship was found between grades and ratings for the junior-senior group (Anikeef, 1953:459). Bendig (1953), using standardized achievement tests for certain psychology courses, concluded that a student's level of achievement at the time of rating was not related to the student's rating of the teacher. The student's achievement level was, however, found to be

related to his rating of the course. In further discussing the research Bendig (1953) pointed out that no significant relationship was found between achievement level of the student and the student's rating of the teacher when all classes were pooled; there was, however, a great difference in the strength of the relationship from class to class. This finding clearly emphasized the possibility that such a relationship may exist for some classes and not for others (Bendig, 1953:446).

Following these initial efforts came studies of a somewhat different and in many ways more sophisticated nature. Beginning with Weaver (1960), for instance, researchers began to study not only the relationship between final grades and ratings, but also, the relationship between expected grades and ratings, and also, the relationship between prior grades and ratings. Weaver (1960), using 699 student ratings, found significant differences between the mean teacher rating of students expecting to receive A's and B's in a course and the mean teacher rating of students expecting C's and D's. Following the lead of Weaver, Stewart and Malpass (1966) also studied the relationship between student ratings and student expected grade. This research, utilizing an χ^2 analysis of independence and involving 1,975 student ratings of 67 teachers, supported Weaver's contention that students expecting high grades in a course rate their teachers higher than students expecting low

grades (Stewart and Malpass, 1966:348). Expanding the work of Weaver (1960) and Stewart and Malpass (1966), Bausell and Magoon (1972) pinpointed the shortcoming of research dealing with the relationship between final grade received and ratings: "One obvious problem with this approach is that at the time of his rating the student may have expected a grade other than the one he received (Bausell and Magoon, 1972:1014)." Thus, these two researchers also studied the relationship between expected grade and rating. The resulting study conducted at the University of Delaware utilized over 12,000 student ratings, employed an analysis of variance design, and produced significant rating differences between students expecting different grades. These researchers argued, in fact, that as much as 44% of rating variability was accounted for by expected grade. Bausell and Magoon (1972) further reported a strong relationship between ratings and grade discrepancy (difference between grade point average and expected grade).

The relationship between prior grades and ratings was studied by Rayder (1968) and was not found to be significant. In this study grades that a student received in prior courses from the same teacher were correlated with the student's rating of the teacher (Rayder, 1968:78).

The studies by Weaver (1960); Stewart and Malpass (1966); Bausell and Magoon (1972); and Rayder (1968) did not, however,

bring a termination of research dealing with the relationship between final grade received and ratings. In fact, recent studies by Overturf and Price (1966); Bledsoe, Brown and Strickland (1971); Weigel, Oetting and Tasto (1971); and Cornwell (1974) all investigated the relationship between final course grade received and ratings. Bausell and Magoon (1972) did study the relationship between expected grades and ratings. Overturf and Price (1966) conducted a rank correlation of final course grades of students at St. Johns River Junior College with the ratings these students gave their teachers, and as a result, concluded that the grades were independent of rankings. On the other hand, the recent research by Bledsoe, Brown, and Strickland (1971) resulted in the conclusion that students receiving the highest grades in a course gave higher ratings than students receiving the lowest grades. As can be seen in Table 1 on page 22 this study utilized analysis of variance and involved high school teachers in Georgia (Bledsoe, Brown, and Strickland, 1971:121). Weigel, Oetting and Tasto (1971) reported similar findings when they compared final course grades of 331 students in seven psychology courses at Colorado State University with the ratings these students gave their instructors. In pooling the ratings of all the instructors these researchers found that ratings by students receiving A's were higher than either the ratings of B-C students or D-F students. In addition these researchers drew attention to a possible shortcoming of the many large scale

correlational studies that have been conducted on the relationship between final grades and ratings:

> Finally even though large correlational studies indicate that students' grades and evaluations of teachers are not importantly related, this relationship should not be dismissed lightly. The effect is likely to be idiosyncratic to both the teacher and the course and should be considered in planning or interpreting teacher evaluations (Weigel, Oetting, and Tasto, 1971:62).

Cornwell (1974) further substantiated the claim of rating idiosyncracy when he found the average ratings of chemistry teachers in twenty different institutions to be significantly related to the average grade received by the class.

In summarizing it is apparent that generalization about the existence of a global relationship between the rating a student gives the teacher and grades (expected, prior, or final) of the student has not been demonstrated through the research cited. Some conclusions and recommendations for future research do, however, seem warranted. The early studies of the 1930's and 1940's were for the most part extensive and correlational in nature; and generally speaking they all supported the conclusion that no relationship exists between student ratings of teachers and student grades. Subsequent studies, however, did not substantially verify this conclusion; many recent studies did, in fact, report a strong relationship between these two variables. A close inspection of the research presents two possible reasons for the seemingly contradictory findings: 1) The early

studies being correlational in nature and in some cases involving large and diverse samples, may have disguised relationships that were present for some teachers and classes, but not for others; and 2) The early studies, for the most part, investigated the relationship between the grade the student received in the course and the rating, while certain recent studies have investigated the relationship between expected grades and ratings.

Each of the studies dealing with the relationship between expected grade and ratings, namely; Weaver (1960), Stewart and Malpass (1966), and Bausell and Magoon (1972) supported the conclusions that the grade a student expects to receive is related to the rating he gives his teacher. Thus, if a cause-and-effect relationship is to be inferred for the relationship between grades and ratings, it follows that expected grade at the time of rating (rather than final grade received) should be the grading variable investigated.

Even using this variable it is necessary to assume, as has been done by past researchers, that students expecting a "low" grade in a course are necessarily unhappy and that students expecting a "high" grade in a course are necessarily happy. Is it not possible that some students might be unhappy with a B while others might be very happy with a C or even in some cases a D? In this regard it is this writer's contention that neither expected

grade nor final grade is the proper variable to be investigated if, indeed, a cause-and-effect relationship is to be inferred. A better variable, it seems, would be the student's attitude or disposition toward the grade he is expecting (at the time of rating).

Student's Grade Point Average

The relationship between a student's grade point average and the rating the student gives the teacher has been researched by Overturf and Price (1966), Rayder (1968), and Bausell and Magoon (1972). By comparing the ratings honor students gave their teachers with the ratings of other students, Overturf and Price (1966) concluded that the honor students tended to be more sensitive in their ratings in that they rated the "high" instructors higher and "low" instructors lower than did the other students. Actual instructor ranking, however, was not found to be different for honor students' ratings than for other students' ratings. Rayder (1968) also found no relationship between reported grade point average of 4,032 college students and the ratings these students gave their teachers. In the same study Rayder (1968) found by comparison of reported grade point averages with registrar's records that for the most part the students reported their grade point averages accurately. The findings of Bausell and Magoon

(1972:1021) prompted them to conclude that "the relationship between grade point average and ratings alone is negligible, and should not be considered an important source of bias." (Additional information on these studies is contained in Table 1, page 22.)

Although the three studies cited in this section all reported similar findings, the extent to which the relationship between students' grade point averages and students' ratings may exist in a given class, department or even college has not been adequately demonstrated. It would appear that more research, conducted in a variety of different contexts, is necessary.

Class Size

Class size, a much debated topic in other areas of education, has not escaped notice in student rating research. Starrak (1934), Heilman and Armentrout (1936), Riley et al. (1950), and Cornwell (1974) all have studied the relationship between student ratings of teachers and class size. The findings of Starrak (1934) indicated only a slight relationship between class size and ratings. Starrak (1934) did conclude, however, that ratings given by classes of less than seven students and more than fifty students tended to be lower than ratings given by classes whose size fell between these two extremes. From a sample of fifty classes ranging in size

from seventeen to 120, Heilman and Armentrout (1936) found no correlation between class size and student ratings of teachers. The extensive Brooklyn College study substantiated these earlier studies. In reporting the results of the Brooklyn study Riley et al. (1950) contended that the overall relationship between class size and ratings is slight. Riley et al. (1950) did report, however, that classes with fewer than twenty students produced high ratings for art instructors in the area of speaking ability and that science classes with large enrollments tended to give higher ratings in the area of organization of material and knowledge of the subject. In summarizing this relationship Riley et al. (1950:87) stated that "with these few exceptions--and most of them seem reasonable and understandable--size of class appears to have little to do with student estimates." The research by Cornwell (1974) gave added support to these previous studies. Cornwell (1974), using student ratings conducted in the chemistry department of the University of Wisconsin--Madison, also concluded that class size was a weak factor in those ratings. Classes with enrollments of less than twenty were found, however, to give somewhat higher ratings than did larger classes. No rating differences were found among classes having enrollments of more than twenty. This same study, when extended to include seventy different chemistry lecturers from twenty different institutions, resulted in similar findings:

classes with twenty students or less again gave higher ratings. The relationship was, however, much weaker for the extended sample (Cornwell, 1974:158).

The similarity of the findings of each of these studies supports the argument that class size is not strongly related to student ratings. Even had the relationship been found to be a strong one, it could have been argued that the relationship may have been attributable to differences in teaching effectiveness among different size classes rather than to a defect or bias in the ratings. This difficulty which has also occured in other areas of research on student ratings of teachers, emphasizes the need for additional research with better controls.

Class Level and Student's Age

In many ways research on the relationship between class level and student ratings of teachers has paralleled the research on the relationship between student grades and ratings. The early studies, dating back to 1930's, generally supported the argument that class level is independent of student ratings while later studies indicate that a relationship may exist in certain contexts.

Studies reported in the 1930's, namely, Starrak (1934), Heilman and Armentrout (1936), and Remmers (1939) all supported the same conclusion: Student ratings of teachers are not related

to the class level of the student. All three of these studies dealt with ratings conducted at the undergraduate college level.

Research during the 1940's and 1950's, however, began to show that a relationship between class level and student rating may exist for comparisions between certain class levels, but not between others. Supporting the research of the 1930's were the studies by Brookover (1940), Riley et al. (1950), and Drucker and Remmers (1951). Brookover (1940) found no differences among various classes of high school students in Indiana. Riley et al. (1950:88) reported similar findings: "While some differences appear between the ratings of various classes, these are small and erratic fluctuations." The research of Drucker and Remmers (1951) was different from prior research in that it included ratings of teachers by college alumni. In this research the ratings of Purdue teachers by certain Purdue alumni were compared with current student ratings of those same teachers. No significant difference was found between the mean ratings of these two groups (Drucker and Remmers, 1951:137). On the other hand, research during this period by Davenport (1944), Remmers and Elliot (1949), and Bendig (1952) gave indications that class level may, in some instances, be related to ratings. Davenport (1944), after grouping students into fifteen, sixteen, seventeen, eighteen, and nineteen year-old categories, found the ratings of the eighteen

year olds to be significantly different from the ratings of other groups. Remmers and Elliot (1949) included the teacher ratings by graduate students in the study and in so doing found that the graduate students rated their teachers higher than did the undergraduates. Research by Bendig (1952) further indicated that, although the relationship between class level and student ratings may not be strong for large samples from whole colleges, it may, nevertheless, exist in some classes. In Bendig's study comparisons by class level of the ratings of two University of Pittsburg psychology teachers revealed that "juniors and seniors rated both instructors more unfavorable than did freshmen and sophomore students (Bendig, 1952:26)."

Recent studies have produced added emphasis to the argument that the relationship between class level and student ratings (as was the case for the grade-rating relationship) may be idiosyncratic to the class, teacher, and school involved. Stewart and Malpass (1966) reported to have found college class level independent of teacher ratings. Bledsoe, Brown and Strickland (1971), however, in a study of ratings of secondary school teachers found both student age and grade level to be related to teacher ratings. These researchers reported that twelfth and tenth graders gave higher ratings than did ninth and eleventh graders and that eighteen year-olds gave higher ratings than did seventeen year-olds. Nineteen

year-olds gave significantly higher ratings than did any other age group (Bledsoe, Brown and Strickland, 1971:121).

Generally speaking the research to date on the relationship between class level and student rating has shown a development similar to research on the relationship between student grade and student rating. While the early studies of the 1930's indicated the absences of a relationship between class level and student rating, more recent studies have shown the relationship to be present in some contexts but not in others.

Student's Sex

Research findings on the relationship between the sex of the student and student ratings of teachers seem to be equally divided between those indicating such a relationship exists and those indicating such a relationship does not exist.

The differences in research findings on this relationship became evident as early as the 1940's with Bryan (1937) and Davenport (1944) claiming that the sex of the student is related to the student's rating of the teacher; and Brookover (1940), and Remmers and Elliot (1949) claiming it is not. Bryan's study (1937) indicated that boys tended to rate men teachers higher than did girls and that girls tended to rate women teachers higher than did boys. Davenport's study (1944), as did Bryan's (1937), involved

secondary students and teachers and resulted in showing that ratings (of both male and female teachers) by female students were higher than ratings by boys. Boys and girls alike tended to rate male teachers higher than female teachers (Davenport, 1944:37). On the other hand, neither Brookover (1949) nor Remmers and Elliot (1949) found the student's sex to be related to the rating of the teacher. The study by Remmers and Elliot (1949) unlike these others, was conducted using college students and teachers.

Subsequent studies did not settle the question. Bendig (1952) and Jackson and Fuller (1966) found student's sex related to teacher ratings; Riley et al. (1950) and Bledsoe, Brown and Strickland (1971) did not. Bendig, as related earlier, investigated student ratings of two psychology teachers at the University of Pittsburg. In this study Bendig (1952:26) found that "women tended to rate both instructors more unfavorably than did men students." Jackson and Fuller (1966) found, using a sample of 28 female teachers and 128 students, that female students tended to rate the teachers as being more confident and poised than did male students. Producing different results were the studies of Riley et al. (1950) conducted a at Brooklyn College, and Bledsoe, Brown and Strickland (1971) conducted in Georgia secondary schools. Neither of these two studies found the student's sex to be related to the rating of the teacher.

The divergence of results obtained in the studies conducted on the relationship between student's sex and student rating make it difficult to summarize the findings. Apparently this relationship like the relationship between class level and student ratings, and also, the relationship between student grade and student ratings, is dependent upon the setting in which the ratings are performed. Certainly no generalizations can be made based on the research investigated here.

Required vs. Elective Courses

At Montana Tech the question of whether or not teachers of elective courses receive higher ratings than teachers of required courses has been one of the most debated of the questions regarding student ratings of teachers at that school. If the amount of published research related to this question is indicative it would appear, however, that other faculties in which ratings have been used do not share this concern.

Of the research considered in this review only two studies included an investigation of the question of student rating differences in required and elective classes; Heilman and Armentrout (1936), and Riley et al (1950). With the exception of the "slightly higher" ratings received by sociology instructors of elective courses in the study by Riley et al. (1950), neither of the two studies found that

teachers of required courses received ratings different than teachers of elective courses.

Certainly the lack of published research on this question makes generalization dangerous. The questions of whether or not teachers of required courses and elective courses at various class levels, in various subject areas, at various institutions, at various schools, and in various locations still appear unanswered.

Student's Attitudes

The relationship between a student's attitudes and ratings of teachers, a topic of central importance to the present study, has received very little past research effort. In fact, of all the research reviewed, only that conducted by Riley et al. (1950) included consideration of the relationship between student attitudes and student ratings of teachers. Riley et al. (1950) reported that students in the arts possessed the most favorable attitude toward teaching of ideas while students in the social scienced followed closely. Science students tended to be more favorably disposed toward factual teaching. "Superior" students were found to be more favorably disposed toward the teaching of ideas than were "poor" students. Riley et al. (1950) also found no significant difference between class (freshman, sophomore, junior, senior) with respect to student attitudes toward the teaching of ideas vs. facts. In

addition Riley et al. (1950) found that less than half of the student body at Brooklyn College felt that there were too many required courses. The study also pointed out that over two thirds of the student body felt that the principal purpose of a college education is to prepare the student for professional work (Riley et al., 1950:79).

Although the findings of Riley et al. (1950) are of interest, it is unfortunate that no attempt was made to determine whether or not these various student attitudes were related to the ratings the students gave their teachers.

Much of the published research on student ratings of teachers has been directed toward the determination of whether or not student ratings are equitable for all teachers. In this regard the research has, by and large, been confined to the study of variables which teachers, administrators, students and researchers have contended, make these ratings unfair. This being the case it is somewhat curious that virtually no research to date has attempted to study the relationship between student ratings and student attitudes toward such things as: teachers in general, teacher ratings in general, the hour the class is held, the classroom in which the class is held, the topics of the course, the grade the student expects to receive, and the school.

Multiple Relationships

Most of the research on student ratings has been confined to studies of relationships between pairs of variables, one of which always was the rating the teacher received. In fact, only two of the studies reviewed were concerned with relationships involving more than two variables.

The first of these studies, conducted by Rayder (1968) at Colorado State College, was a multiple linear regression analysis using student sex, student age, class level, and major of the student as independent variables and teacher rating as the dependent variable. (This writer is curious as to how a student's major could be included in a multiple regression analysis.) This analysis resulted in an R^2 (coefficient of determination) value between .006 and .019. From this Rayder (1968) concluded that these variables demonstrate low predictability of teacher ratings. The other study, conducted by Cornwell (1974), also used the average ratings of chemistry teachers as the dependent variable. Independent variables for the analysis were: number of students enrolled in the class, hour class met, age of the lecturer, and average grade for entire class. The coefficient of determination (R^2) for this analysis was found to be .17. Thus, it was concluded that only a "small fraction" of the variance of the ratings was attributable to these dependent variables (Cornwell, 1974:159).

The scarcity of multivariate research on student ratings of teachers is not surprising due to the fact that, for the most part, past investigations have tended to produce "small" pairwise correlations. In fact, many of these coefficients were found to be significant partly because of the large samples involved. It does, nevertheless, appear that additional multivariate analyses, involving many of the variables researched in the past are needed.

Other Variables

Limited research has been directed toward the relationship between student ratings and each of the following: length of time the student has known the teacher, social class of the student, ability grouping of the students, and the hour the class was held.

Davenport (1944) and Brookover (1945) both studied the relationship between student ratings and the length of time the student had known the teacher. Davenport (1944) found no difference between mean ratings of teachers with respect to the number of semesters the students had known the teacher. Brookover (1945) did, however, find a significant correlation between the rating a high school student gave his teacher and the length of time the student had known the teacher.

Only one study was found that investigated the relationship between student ratings and student social class. This study by

Fuller and Jackson (1966) resulted in the conclusion that pupils tended to give higher ratings to teachers of a different social class than their own. These ratings were higher in every area except "effective communication" where it was found that a teacher of the same social class as the student received the higher rating. The social classes of the teachers were determined by use of the McGuire-White Index of Social Status; students' social classes were determined by inspection of home neighborhoods.

Differences in the teacher ratings by students in different ability groups has been investigated by Bledsoe, Brown and Strickland (1971). This research revealed that students not grouped by ability tended to give lower ratings than did students in the low and average ability groups. The ratings of students in the high ability group were not, however, found to be different from the ratings of students not grouped by ability. This study was conducted in secondary schools in Georgia and utilized the POSR (Pupil Observation Survey Report) as the rating instrument.

Of the studies reviewed only two dealt with the relationship between student ratings and the hour of the class. By using faculty and student opinion to classify a 1:00 pm class as meeting at a bad hour and a 9:00 am class as a good or preferred hour, Overturf and Price (1966) found the difference in the ratings of teachers teaching at the "preferred" hour not to be significantly different

from the ratings of teachers teaching at the "bad" hour. Similar findings by Cornwell (1974) indicated that the hour a class met was not significantly related to student ratings of chemistry teachers.

Because of the limited number of studies related to the relationships discussed in this section generalizations seem unwarranted. It should be noted, however, that in the research by Overturf and Price (1966) it was implicitly assumed that students uniformly preferred a 9:00 am class. Since each student might, conceivably, prefer a different class hour, a better approach might have been to measure each student's disposition toward the hour the class was actually held and then to determine whether or not these student dispositions were related to ratings. Along these same lines, much of the past research on student ratings has been directed toward strictly "mechanical" variables such as: the student's grade in the course, the number of students in the class, the level of the class, the sex of the student, etc. In that it is the student who does the rating, it seems that this approach has been somewhat misdirected. A better approach, perhaps, would be to study the relationships between student ratings and student disposition or attitude toward these "mechanical" variables.

ANALYSIS OF RESEARCH

Representativeness of Samples

A discussion of the representativeness of the samples used in the research cited in this chapter must, it seems, include comments about the intent of such research. If research on student ratings of teachers is intended to determine global relationships involving all levels of education and covering wide geographical and cultural settings, then the only valid sample would be one that involves a wide variety of teachers and students, and that is taken from a large cross-section of cultural and geographical areas. Not only does such a procedure appear impractical; it also appears improper, especially since the use of student ratings is for the most part confined to a given course, department, or school. And in fact, the many divergent findings of the research cited is a strong indication that global relationships may not even exist. If this is indeed the case then research such as that of Bendig (1952) which involved only two psychology teachers cannot be interpreted as being any less representative than research such as that of Remmers and Elliot (1949) which involved over 460 teachers in ten different colleges.

In this regard, it is this writer's contention that the findings of research studies on student ratings of teachers should

not be extrapolated beyond the specific setting in which they are conducted. Also, since the research to date has indicated that relationships found to exist in one setting may not exist in another, it is a further contention of this writer that research on these relationships should be a part of the institutional research of any institution using student ratings of teachers.

Rating Instruments Used

The Purdue Rating Scale for Instruction has, undoubtedly, become the most widely used and studied instrument of its kind; and as such, is deserving of special consideration. As can be seen in Table 1, on page 22, over one third of studies reviewed used this instrument or a modification of it. This instrument, developed by H. H. Remmers in the 1920's, contains ten rating items and utilizes 100-division graphic scales. Research by Brandenburg and Remmers (1928) and Remmers (1929) has shown that the instrument is reliable and absent of any great halo effect.

For the most part, however, the rating instruments used in research dealing with student ratings of teachers were constructed either by the researcher or by the institution in which the ratings were performed. If the research included in this review is indicative, it appears that researchers using these "home-made" rating forms have failed to ascertain the reliability of these instruments;

or if they have, they have failed to report it. Failure to report reliability estimates of teacher ratings forms used in research on student ratings is a serious shortcoming of much of this research. This shortcoming is emphasized by the findings of Morsh and Wilder (1954) (a review of teacher effectiveness studies conducted prior to 1952) which indicated that student ratings tend to show a wide range of reliabilities.

Time of Study

As indicated in this review there appears to be some differences between early findings (primarily in the 1930's) and recent findings in the area of student rating research. The reasons for these differences are not easily established, however. One possible reason for the differences lies in the gradual increase in "sophistication" of research designs and statistical analyses used. In this regard it may be argued that the large sample correlational analyses, prevalent in the 1930's tended to hide significant relationships in particular classes or for particular teachers.

To determine these class to class and teacher to teacher differences many of the recent studies (See Table 1 on page 22) have employed statistical comparisons of mean teacher ratings and also analysis of variance designs. It is the opinion of this writer, however, that data generated from teacher ratings scales cannot be

considered as other than ordinal data; a case in which many of these "sophisticated" techniques also appear inappropriate.

On the other extreme, certain of the recent researchers, Stewart and Malpass (1966) and Weigel et al. (1971), have resorted to the weaker nonparametric statistical analysis. This latter approach may, in fact, be the most realistic.

SUMMARY

Student ratings of teachers and research on student ratings of teachers have been in evidence for over forty years. The pioneering studies by Remmers and others in the 1930's tended to support the argument that student ratings are reliable and valid; more recent studies, however, have pointed to possible shortcommings of the process. In this regard the main arguments have centered around the relationships between student ratings of teachers and the following: the student's grade in the course (prior, expected, and final), the student's grade point average, size of the class, class level and student's age, sex of the student, and whether the course is required or elective. Research during this forty year period has, unfortunately, failed to establish many generalizations about these relationships. This research does, in fact, seem to indicate that these relationships are idiosyncratic to a given teacher, class, school or setting; and implicit in this finding, then,

is the need for further research on these relationships, especially by those institutions currently using formalized student ratings of teachers.

CHAPTER III

PROCEDURES

INTRODUCTION

As stated previously this study was designed to determine what, if any, relationships exist between certain student and contextual variables and the ratings the students give their teachers. The specific student and contextual variables selected are of a kind that tend to make these ratings invalid if significantly related to the ratings. In this regard the study attempted to determine only the linear relationships between student ratings and the student and contextual variables listed in CHAPTER I. This was achieved by correlating the student and contextual variables with the student ratings. These variables were correlated individually using the Pearson product-moment coefficient of correlation, and also, in combination using the multiple R coefficient of correlation. No curvilinear relationships were investigated. In addition a multiple linear regression model was developed using the arithmetic mean of items on PART IV of the rating instrument (shown in APPENDIX A) as the dependent variable. The independent variables for this model were chosen from the student and contextual variables listed in CHAPTER I; variables chosen were those over which teachers have limited control. As indicated in the previous

chapter, certain of these independent variables have received investigation in the past; none of the research cited, however, dealt with the measurement of student attitudes and the relationship of these attitudes to student ratings. Because it was necessary to measure these attitudes, the rating instrument used in this research appears quite different from teacher rating forms used in the past.

The remainder of this chapter presents a description of the procedures followed in constructing the rating instrument, administering the instrument and analyzing the data generated from it. The chapter is divided into the following major divisions: POPULATION, DESCRIPTION AND SAMPLING PROCEDURE, AREA OF INVESTIGATION, DATA COLLECTION AND VALIDATION, ORGANIZATION OF DATA, ANALYSIS OF DATA, PRECAUTIONS TAKEN FOR ACCURACY, and SUMMARY.

POPULATION DESCRIPTION AND SAMPLING PROCEDURE

In studies dealing with student ratings of teachers, two populations must be considered: 1. The students who rate the instructors, and 2. The instructors who are rated. Since a basic intent of this research was to improve the rating process at Montana Tech, it follows that in this sense the two specific populations under investigation in this study were: 1. All the teaching faculty at Montana Tech, and 2. All the students attending Montana Tech.

Since the majority of the variables investigated were student variables rather than teacher variables, the population of greatest concern in this study is the student population.

Just as there were two populations to be considered in this study, there were also two samples: 1. The faculty at Montana Tech who participated in the rating process during fall semester, 1973, and 2. The students who participated in the rating process during that same semester. The faculty sample consisted of 46 of the 49 regular teaching faculty (93%). However, since these teachers were asked to conduct ratings in each class they taught, and since it is conceivable that teachers behave differently in different classes, a more basic faculty sampling unit is a particular teacher of a particular class. Considering this as the faculty sampling unit, the sample consisted of 144 out of 153 classes taught (94%) during fall semester, 1973.

Since each student attending Montana Tech during fall semester, 1973, had the opportunity to evaluate each teacher from whom he took a course, the basic student sampling unit for this course was not a particular student at Montana Tech, but rather any student in any course at Montana Tech during fall semester, 1973, who filled out a rating form on the teacher of the course. And, since students were not required to respond to every item on the rating form, the number of responses on the various items differed.

(The number of responses is reported in CHAPTER IV.) Thus, there was actually a different student sample for each item on the form. This sampling procedure was justifiable in that, although a given student may have provided more than one response to each item, each of these responses constituted a different sample point in the sense that the student and contextual variables and the teacher being rated varied from response to response. In fact, no controls were placed on the sampling process other than those that had been used in the past. This was necessary since the study endeavored to investigate relationships in the rating process as it existed at the time of the study. Any sampling controls other than those used in the rating process as it existed at that time might have altered the relationships under investigation.

AREA OF INVESTIGATION

As stated previously the variables investigated are those listed in CHAPTER I. The relationship between these variables and student ratings was the major focus.

Since this research was designed to investigate these relationships in the context of the rating process as it existed at Montana Tech at the time of the study, no attempt was made to control any variables that had not been controlled prior to the study.

In regard to certain of the student and contextual variables and the relationships between these variables and student ratings (e. g. students' age, grade point average of student, expected grade of student, level of the class, class size, and sex of the student) this study was somewhat of a replication of many prior studies. And, although these variables were not of principal interest in this research, they were, because of their importance, included. The variables and relationships of these variables to student ratings that were of primary interest to this investigation were those dealing with student attitudes toward: the classroom, the occupation of teaching, the hour the class was held, teacher evaluations, Montana Tech, topics taught in the course, grade the student expected in the course, and the way the student felt at the time of rating.

DATA COLLECTION AND VALIDATION

Data Collection Instruments and Procedure

Data for this research was collected by means of the rating form exhibited in APPENDIX A and also by means of the instrument shown in APPENDIX C. Both instruments were designed specifically for this study by this researcher.

As shown in APPENDIX A the rating form is divided into four parts, each part requesting different kinds of data. PART I

of the instrument requested the student's cumulative grade point average, the number of credits the student was taking at the time of rating, the number of years the student had known the teacher prior to the rating, the number of exams the student had taken on the day of rating and the student's age. In coding the data for this part, the student's reported cumulative grade point average was rounded to the nearest tenth; the number of credits was recorded to the nearest integer; x, the number of years the student had known the teacher, was recorded to the nearest integer less than or equal to x (i. e. Many students responded to this item in terms of fractions of a year.); the number of exams the student had taken on the day of the rating was recorded to the nearest integer; and, the age of the student was recorded to the nearest integer. PART II of the rating form requested the student to record whether or not the given course was required or elective, the sex of the student, the student's major, the grade the student expected in the course, and whether the student's major field of study was the same or different than the teacher's major field of study. Although items 1, 2, 3, and 5 of this part deal with attribute data, for convenience in data reduction and analysis, they were coded numerically. A required course was coded as a 1 and an elective course as a 2, male students were coded as a 1 and female students as a 2, a chemistry major was coded as a 1, an engineering major as a 2, an English major as a 3, a geology major

as a 4, a history major as a 5, a mathematics major as a 6, and any other major as a 7 (Student's major was coded for institutional research and was not used in any analysis in this study.); an expected grade of A was coded as a 1, a B as a 2, a C as a 3, a D as a 4, and an F as a 5; a student whose major was the same as the teacher's was coded as a 1 and a student whose major was different than the teacher's was coded as a 2. The attribute data recorded on item 3 of this part (major of student) was collected for institutional research purposes and was not used in this study. PART III of the rating form was designed to measure the student's attitudes toward: the classroom in which the class was held, the occupation of teaching in general, the hour the class was held, teacher evaluations in general, Montana Tech, topics of the course, the grade the student expected to receive in the course, and the way the student felt on that day. As can be seen each of these variables was measured using the following seven "evaluative" scales of the semantic differential, namely:

GOOD	____:____:____:____:____:____:____	BAD
CRUEL	____:____:____:____:____:____:____	KIND
CLEAN	____:____:____:____:____:____:____	DIRTY
REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
REPUTABLE	____:____:____:____:____:____:____	DISREPUTABLE
SICK	____:____:____:____:____:____:____	HEALTHY
	____:____:____:____:____:____:____	

Each student was asked to respond to each of the seven scales of

each of the eight variables by placing an x in one of the seven spaces allotted to each scale. In coding the data for this part each of the spaces on each of the semantic differential scales was assigned an integer value from 1 to 7. The extreme left space on each scale was assigned a value of 1; the second space from the left was assigned a value of 2; the third space from the left was assigned a value of 3; etc. Thus, a response on any of these scales was coded as an integer greater than or equal to 1 and less than or equal to 7 depending upon where the student placed the x. (e. g. An x placed in the fifth space from the left of a given scale was coded as the integer 5.) It should be noted that these scales have their positive side alternately on the left and right. Thus, before any analysis was performed the data from these scales were transformed so that all scales could be interpreted with the negative side on the left and the positive side on the right. This is explained more fully on page 80. In PART IV of the rating form, as in Part III," the semantic differential was used. The twenty "evaluative" scales shown in this part constitute the measurement criteria by which the student rated his teacher. The teacher attributes listed at opposite ends of the semantic differential scales in item 2 through 13 were selected from a collection of such attributes studied by Ryans (1960) and by Combs (1969). The reasons for this choice of items as criteria of teacher effectiveness; and also, the

validity and reliability of the scales used in the measurement of these items is discussed further in a subsequent section of this chapter. The remainder of the items, with the exception of item 20, are items that had been used previously on rating forms at Montana Tech. Item 20, the smile and frown faces, was used in an attempt to solicit a non-verbal and, perhaps more general, response from the student regarding the effectiveness of the teacher. The data for this part was coded in the same manner as was the data in PART III. None of the information on the comment sheet (last page of the rating form), other than the item requesting comments concerning the rating form and evaluation procedure, was used in this study. The comments on the rating form and evaluation procedure were considered in the determination of the validity and reliability of the form. Any item left blank on any part of the rating form was coded as a blank and was not used in any analysis.

The rating form just described was administered during the week starting Monday, January 14, 1973, and ending Friday, January 18, 1973; this was the week prior to final exam week in fall semester. During this week each instructor was asked to conduct student ratings in each class he was teaching. The procedure for this was the same as it had been in the past. At the beginning of a normal class period a student was chosen (by the teacher) to go to the Office of the Dean of Academic Affairs, pick up an envelope

containing the rating forms for the rest of the class, pass out the forms, collect the completed forms, and return them to the Office of the Dean of Academic Affairs. These rating packets or envelopes were coded by instructor code and also by class code and were made up prior to rating time, one packet per class taught. The students who distributed the rating forms gave no directions since directions were explicitly stated on each rating form. For the classes taught during evening hours, the teachers themselves were requested to collect and return the rating packets.

A sample rating packet envelope is exhibited in APPENDIX C and, as can be seen, these envelopes were also used as a data collection instrument. Data collected here consisted of instructor and class code, date rating took place, period class was normally held, course name and section, teacher's name, sex of teacher, number of students enrolled in the class, and major field of specialization of the teacher. Each instructor's code number (CODE) was coded as a three digit integer with the two high order digits designating the instructor code and the low order digit designating the course code. These codes were used for the purpose of making up the teacher evaluation reports shown in APPENDIX D. The DATE indicated on each rating envelope was used to indicate the day of the month on which the ratings were conducted. This information was coded by the Dean's secretary at the time the student picked up

the rating envelope. The PERIOD the class was held was coded using class schedule record in the Registrar's office. Class periods were coded as integers from 1 through 9 (normal daytime class periods) according to the period during which a class began. Class period data was not used for evening classes. COURSE--SEC, and INSTRUCTOR were coded respectively according to the course name and number, the section of the course and number, and the name of the teacher of the course. This information was used primarily in preparing the evaluation report for the faculty. The level of the course, 100, 200, 300, or 400 corresponding respectively to the freshman level, sophomore level, junior level, and senior level was recorded from the course numbers designated in the COURSE blank. These levels were coded respectively as 1, 2, 3, and 4. The sex of the teacher MALE or FEMALE was coded as a 1 or a 2, respectively. NUMBER OF STUDENTS was used to designate the number of students enrolled in the class and was coded using the registrar's finalized class lists. The TEACHER'S MAJOR was recorded as one of the seven categories shown in APPENDIX C according to the major area of specialization of the teacher; and it was included for institutional research purposes outside this study.

On PART III and PART IV of the rating form the positive sides of the semantic differential scales were alternated from left to right. This was done to reduce the occurence of response sets. To

reduce halo effect error the student and contextual variables (PART I, PART II, AND PART III) were placed on separate pages from the teacher rating variables (PART IV). In this respect it would also have been desirable to place each of the student attitude variables (PART III) on separate pages. This was not done since it would have made the rating form excessively long and cumbersome.

Reliability

A measure of the reliability of the rating instrument was determined by a test-retest application of the instrument in each of the following fall semester classes: Humanities and Social Science 283 (American National Government), Engineering Science 353 (Engineering Analysis of Electrical Machines), Biology 221 (General Botany), and Mathematics 107 (Introductory Algebra and Trigonometry). The test-retest procedure consisted of having each student present in each of these classes complete two rating forms on his teacher. For the purposes of this test the initial teacher rating in Humanities 283 was conducted on November 7, 1973, and the second on November 28, 1973. The initial teacher rating for Engineering Science 353 was conducted on November 9, 1973, and the second on November 30, 1973. The initial teacher rating in Biology 221 was conducted on November 13, 1973, and the second on November 29, 1973. The initial teacher rating for Mathematics 107 was conducted on

November 11, 1973, and the second on November 27, 1973. Reliability measures for each item on the form were determined by computing the Pearson product-moment coefficient of correlation between a student's first response to an item and the student's second response to that item. Tables 2, 3, 4, and 5 show these correlations.

Table 2

Reliability Coefficients for Items on PART I of Rating Instrument (see APPENDIX A)

	Item Number				
	1	2	3	4	5
r	.98	.97	.96	.15	.99
n	49	78	78	78	77

As can be seen from Table 2 the items in PART I (excepting item 4) all show a high degree of test-retest reliability as evidenced by correlations greater than .95. Item 4 (The number of exams I have taken today is:) obtained a low correlation coefficient of .15. However, in this item, unlike the other four items in PART I, the criterion being measured was a time dependent variable and since the test-retest method of determining reliability of a measuring

device assumes that the object being measured does not change between the initial test and the final retest, the low correlation coefficient does not necessarily mean the measurement is unreliable. The items of PART II as shown in Table 3 all show a high degree of test-retest reliability.

Table 3

Reliability Coefficients for Items on PART II of Rating Instrument (See APPENDIX A)

	Item Number				
	1	2	3	4	5
r	.95	1.00	.97	.88	1.00
n	75	78	78	77	77

Table 4 shows the reliability coefficients for the various attitude scales of PART III. As expected, these coefficients are lower and show greater variation than the coefficients of PART I and PART II. The interpretation of the reliability coefficients of PART III is once again subject to the possibility of a change in the criteria being measured during the period between test-retest. Since the way a person feels may change from day to day, the reliability coefficients for this item (item 8 in Table 4) may not be a meaningful

Table 4

Reliability Coefficients for Variables on PART III of Rating Instrument

Attitude Variable*	Reliability Coefficients (r)								Composite n
	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6	Scale 7	Composite Attitude (Average)	
1	.77	.51	.54	.42	.58	.52	.38	.75	75
2	.49	.38	.22	.26	.54	.51	.28	.70	75
3	.80	.79	.61	.73	.68	.74	.66	.80	74
4	.47	.68	.67	.69	.49	.56	.50	.66	75
5	.75	.70	.70	.54	.36	.54	.58	.74	76
6	.61	.46	.49	.52	.51	.46	.57	.64	74
7	.58	.47	.55	.64	.66	.57	.53	.55	75
8	.25	.32	.24	.34	.35	.27	.26	.42	76

* Attitude Variables: Variable 1. THIS CLASSROOM, Variable 2. OCCUPATION OF TEACHING, Variable 3. THE HOUR THE CLASS IS HELD, Variable 4. TEACHER EVALUATIONS, Variable 5. MONTANA TECH, Variable 6. TOPICS OF THIS COURSE, Variable 7. GRADE I EXPECT IN THIS COURSE, Variable 8. TODAY I FEEL. (Scales 1 through 7 are shown in APPENDIX A.)

measure of the consistency of measurement. Also since the occurance of an examination during the period between test-retest may have changed the attitude of the student toward the grade he expected in the course, the same argument can be made for this item (item 7, Table 4) as was made for item 8. For these two items it would have been better, perhaps to use a shorter period between applications of the rating instrument. Shortening the period between applications of the rating instrument does, however, increase the possibility of remembering the initial response to a given item, and thus, responding in the same way on the second application. The period between applications should be long enough to minimize memorization, but short enough to prevent a change in the criteria being measured. "Unfortunately this ideal interval is unknown, but most investigators use intervals ranging from two to six weeks (Shaw and Wright, 1967:17)." As stated earlier a period of approximately two weeks was used in this reliability test. According to Shaw and Wright (1967) greater reliability is found for attitude measurements when the attitude score is computed by using the average of all scales involved. Examination of Table 4 shows this also to be the case for the attitude measurements of PART III of this instrument. The reliability coefficients listed under the heading composite (average of all nonzero scales for the given attitude variable) for a given variable are all higher than coefficients on the majority

of individual scales for that variable. Table 5 shows the reliability coefficients for the teacher rating variables of PART IV of the rating instrument. As was the case for the measurement of composite attitudes, the composite teacher rating (average of all nonzero rating scales) rating shows a higher reliability coefficient than do the majority of individual rating scales.

Validity

Concurrent validity of items 1, 2, 3, and 5 of PART I and items 1, 2, 3, and 5 of PART II of the rating instrument was determined using data collected in the initial application of the reliability test. In this validation the following external criterion measures were used: For items 1, 2, and 5 of PART I (student's cumulative grade point average, number of credits student was taking at time of rating, and student's age) the external criteria were taken from the Registrar's records; for item 3 of Part I (the number of years the student had known the teacher) the external criterion was selected by asking the teacher how long he had known the student; for item 1 of PART II (whether course was required or elective) the external criterion was selected by checking the Registrar's records to determine the student's major and then checking the college catalog to determine whether or not the given course was required or elective for the given major; for items 2

Table 5

Reliability Coefficients for Teacher Rating Variables on PART IV of Rating Instrument

	Scale Items*																				Composite Rating
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
r	.61	.55	.54	.49	.58	.21	.30	.56	.65	.70	.58	.41	.50	.64	.80	.54	.66	.32	.65	.60	.77
n	73	74	73	74	74	74	74	73	75	74	73	75	70	70	72	74	74	75	74	73	74

*1. EXCELLENT TEACHER --POOR TEACHER
2. STRESSES FACTS--STRESSES CONCEPTS
3. STUDENT CENTERED--SUBJECT CENTERED
4. UNINVOLVED--INVOLVED
5. CONTROLS STUDENTS--FREES STUDENTS
6. FRIENDLY--UNFRIENDLY
7. HINDERING--HELPFUL
8. ALOOF--RESPONSIVE
9. STIMULATING--DULL
10. UNCERTAIN--CONFIDENT
11. SYSTEMATIC--DISORGANIZED
12. INFLEXIBLE--ADAPTABLE
13. BROAD--NARROW
14. PARTIAL GRADER--FAIR GRADER
15. NEAT--SLOPPY
16. CONFUSED OBJECTIVES--CLEAR OBJECTIVES
17. PREPARED--UNPREPARED
18. BAD COMMUNICATOR--GOOD COMMUNICATOR
19. I LEARNED LITTLE--I LEARNED MUCH
20. SMILE FACE--FROWN FACE

Composite Rating is the average of all twenty items (non zero items). (See APPENDIX A.)

and 3 of PART II (sex of student, and major of student) the external criteria were also selected from the Registrar's records; for item 5 of PART II (whether student's major was same or different than teacher's major) the external criterion was determined by comparing the student's major as stated in the Registrar's records with the teacher's major as stated in the faculty directory. The validation procedure, then, for these items consisted of correlating the student's response to each of these items with the corresponding external criteria just named. The Pearson product-moment correlation was used in this correlation. The results of this validation as shown in Table 6 indicate a high degree of concurrent validity for each of these items. Item 3 of PART I (the number of years the student had known the teacher prior to rating) shows a lower concurrent validity than the other items. This may, however, be attributable to a less perfect choice of an external criterion; i.e. it is possible for a student to "know" a teacher a different length of time than the length of time the teacher "knew" the student. Estimation of the concurrent validity of items 4 of PART I (number of exams the student had taken on day of rating) and item 4 of PART II (the grade the student expected in the course at the time of rating) was not attempted due to the inaccessability of suitable external criteria. It may be noted that the coefficient for item 1 of PART I (students cumulative grade point average) is comparable to the

coefficient of .96 obtained by Rayder (1968) for the same item, and thus, substantiates his claim that students are capable of remembering and accurately reporting their grade point averages.

Table 6

Validity Coefficients for Items of PART I and PART II of Rating Instrument (See APPENDIX A)

	Item							
	PART I				PART II			
	1	2	3	5	1	2	3	5
r	.96	.97	.87	.99	.94	1.00	.95	.95
n	52	93	93	92	65	92	93	88

Validation of the measurement device used in PART III of the rating form was a more difficult task than was the case for PART I and PART II. The instrument used in PART III was designed to measure attitudes which, being theoretical constructs, require precise definition. In CHAPTER I attitudes were defined as "a relatively enduring system of evaluative, affective reactions based upon and reflecting the evaluative concepts or beliefs which have been learned about the characteristics of a social object or class of social

objects (Shaw and Wright, 1967:3)." Although attitudes resemble other personality constructs, they are, nevertheless, unique.

> To the extent that attitudes are considered to fall within the realm of personality, they are differentiated from other personality constructs on several bases. First, attitudes are relational, and second, their referents are specific. That is, an attitude is a characteristic which implies a type of relationship between the person and specific aspects of his environment. Third, attitudes differ from many other personality constructs in their possession of an evaluative function. Fourth, attitudes, rather than being overt responses, serve as predispositions to respond overtly. Therefore, as with any mediating variable, it is necessary to measure them indirectly (Shaw and Wright, 1967:4).

As can be seen in APPENDIX A the semantic differential was used in PART III of the rating form to measure the various student attitudes. The semantic differential, unlike Thurstone and Guttman attitudinal measurement scales, is a generalized instrument; that is "the same instrument may be used to measure attitude toward any object (Shaw and Wright, 1967:568)." In this regard the semantic differential scales were more easily constructed than would have been the case for other scales, especially since the study required measurement of attitudes toward eight different objects (classroom, occupation of teaching, hour class was held, teacher evaluations, Montana Tech, topics of the course, grade expected in the course, and feeling state of individual). Of course, as is the situation with other attitude scales, only attitude direction and intensity were measured with these scales. The word opposite semantic differential scales

used were taken from Osgood (1957). Osgood, with the use of factor analysis, has classified many such word opposite pairs into three principal factors: "evaluation," "potency" and "activity." The word opposite pairs used in PART III (good--bad, cruel--kind, clean--dirty, regressive--progressive, reputable--disreputable, sick--healthy) were all chosen from the evaluative factor catagory. According to Shaw and Wright the number of such word opposite pair scales commonly used in attitude measurement research varies from 3 to 15 (Shaw and Wright, 1967:30). As shown this research utilized seven such scales for each item of PART III.

In reference to the validity of these bipolar scales Osgood (1957) has stated that "the evaluative dimension of the semantic differential displays reasonable face-validity as a measure of attitude." Furthermore, attitudes toward a variety of different concepts have been measured by the evaluative factor of the semantic differential and compared to measurement taken from Guttman and Thustone scales, scales specifically designed to measure these attitudes. Because these comparisons yielded correlations of the magnitude .74 to .82 Osgood maintained that the semantic differential does provide a valid index of attitude (Osgood, 1957:195). The smile-frown faces used as scale opposites, of course, are not word opposites; but they do connote opposites in the semantic space. These symbols were used to provide a more general measure of the

attitude in question than the measurements obtained from the specific word opposites. Although these symbols were not included in Osgood's study, the closely associated word opposite pair happy--sad was included and was found to have a high evaluative factor loading.

In discussing the use of attitude measurement scales (including the semantic differential) Shaw and Wright (1967) contended that although these scales are not necessarily an adequate assessment of attitudes of a given individual, they are adequate for purposes of research.

> This point needs clarification. In most experimental investigations comparisons are made between attitudes held by groups of individuals. Errors of measurement may be assumed to be randomly distributed about the mean, so that with a sizable number of subjects in each group the obtained mean attitude score approximates the true mean of the population. Consequently, an attitude scale that measures individual attitude imperfectly may yield a reliable and valid measure of the mean attitude held by the group (Shaw and Wright, 1967:565).

All analysis conducted in this research which utilized attitude measurements dealt with the combined group of all students rating teachers at Montana Tech. The possibility that the scales used do not adequately assess a given individual's attitudes does, however, have validity implications for the regression model developed in this study. These implications are discussed in a subsequent chapter.

The scale items of PART IV of the rating instrument

supposedly measures effective teaching. However, the content validity of these scales or any scales presently in use in teacher effectiveness assessment is questionable since it is apparent that research to date has failed to provide general criteria of effective teaching. Ryans clearly summarized this problem:

> It is not surprising, then, to note the difficulties that have confronted those seeking to establish criteria of teaching effectiveness, the dearth of testable hypotheses produced in such research as has been undertaken, and the general lack of understanding of the problem of characteristics of effective teachers. One very important reason why effective or ineffective teachers cannot be described with any assurance is the wide variation that exists in tasks performed by teachers and in value concepts of what constitutes desirable teaching objectives (Ryans, 1960:4).

Research by Combs (1969) proposed that good teachers are those who stress concepts rather than facts, are student centered rather than subject centered, are involved rather than uninvolved, free students rather than control students, are friendly rather than unfriendly and are helpful rather than hindering. Scales 2 through 7 of PART IV use these bipolar attributes. Research by Ryans (1960) indicated that effective teachers are those who are responsive rather than aloof, stimulating rather than dull, confident rather than uncertain, systematic rather than disorganized, adaptable rather than inflexible and broad rather than narrow. Scales 8 through 13 of PART IV utilize these bipolar attributes. As stated earlier in this chapter the remainder of the scale items

are carry-overs from previous rating forms used at Montana Tech. Although it was tacitly assumed that these scales do measure teaching effectiveness, the question of validity of measurement obtained in PART IV was not of great concern to this research. This follows since the scores obtained from these scales, no matter what they measured, were the scores used to rate and rank the teachers at Montana Tech. In fact this research was designed to determine relationships between rating scores (on the existing rating instrument) and the various other variables outlined earlier. In this regard, any rating form would have been appropriate providing it was the form commonly used at the institution.

ORGANIZATION OF DATA

The data and results presented in this paper are presented in table form with explanation of tables and significant results described in the text. Results are presented in the form of correlations between variables recorded on the rating packet, PART I, PART II, PART III of the rating instrument and the rating variables on PART IV of the rating form. When interpreting these tables it should be noted that the bipolar scales of PART III and PART IV of the rating instrument (as shown in APPENDIX A) appear with the positive side alternately on the left and right. In APPENDIX B of this paper these scales are all reoriented so that the negative end of the

scales always appears on the left and the positive end on the right. All analyses dealing with scales on PART III and PART IV were performed using the reoriented scales exhibited in APPENDIX B. Thus, all interpretations of correlation coefficients dealing with these scales should be made with reference to APPENDIX B rather than APPENDIX A. All null hypotheses tested in this research were hypotheses of form $\rho=0$ (population correlation coefficient). Each r (sample correlation coefficient statistic) found to be sufficiently large to warrant rejection of a null hypothesis is exhibited with an * superscript. This superscript indicates that the corresponding ρ has been found to be significantly different from 0 at the .01 level of significance.

ANALYSIS OF DATA

As mentioned earlier in this chapter, and as can be seen in APPENDIX A, many of the bipolar scales of PART III and PART IV are oriented with the negative side of the scales alternating from left to right. Thus, since each of these scales was coded with 1 representing the left most space on the scale and 7 representing the right most space, it was necessary to reorient some of the scales so that all scales were oriented with negative items on the left and positive items on the right (as shown in APPENDIX B). To be more precise this was accomplished by letting the reoriented scale value

equal 8 minus the originally coded value for each of the following scales: all GOOD--BAD, CLEAN--DIRTY, REPUTABLE--DISREPUTABLE scales of PART III, those smile--frown scales with the smile on the left in both PART III and PART IV, and the EXCELLENT TEACHER--POOR TEACHER, STUDENT CENTERED--SUBJECT CENTERED, FRIENDLY--UNFRIENDLY, STIMULATING--DULL, SYSTEMATIC--DISORGANIZED, BROAD--NARROW, NEAT--SLOPPY, PREPARED--UNPREPARED, and I LEARNED MUCH--I LEARNED LITTLE scales of PART IV. For the remainder of this paper all items refered to in PART III and PART IV are oriented as indicated above, and as shown in APPENDIX B.

As a matter of convenience and as a convention to be followed in the remainder of this paper, the several variables under investigation are referenced as indicated in Table 7. Table 8 shows statistical hypotheses that were tested in this study. In this table $\rho(X;Y)$ is used to designate a population correlation coefficient between random variable X and random variable Y. The heading H_0 refers to null hypotheses and H_1 refers to the corresponding alternate hypothesis. The heading, Variables Included, indicates which variables were tested. The variables listed in this table refer to those named in Table 7. Each of the hypotheses tested in Table 8 was tested using the test statistic:

$$t = r\sqrt{(n-2) / (1-r^2)}$$

Table 7

Variable Names for Various Items on Rating Form and Rating Packet

Variable Name	Variable
V_1	Day of week rating conducted
V_2	Period class was held
V_3	Level of course
V_4	Sex of teacher
V_5	Number of students in class
V_6	Major of teacher
V_7	Difference between sex of teacher and sex of student
W_i (i = 1-5)	Item i, PART I of rating form
Z_i (i = 1-5)	Item i, PART II of rating form
$X_{i,j}$ (i = 1-8; j = 1-7)	Scale j of attitude object i, PART III of rating form
Y_i (i = 1-20)	Rating scale i, PART IV of rating form
S_i (i = 1-8)	Arithmetic mean of non zero responses on scales 1-7 of attitude object i, PART III of rating form
T	Arithmetic mean of non zero responses to scales 1-20, PART IV of rating form

Table 8

Hypotheses Tested

Null H_0	Alternate H_1	Variables Included
$\rho(V_i; Y_k) = 0$	$\rho(V_i; Y_k) \neq 0$	(i = 1,2,3,4,5,7), (k = 1-20)
$\rho(W_i; Y_k) = 0$	$\rho(W_i; Y_k) \neq 0$	(i = 1-5), (k = 1-20)
$\rho(Z_i; Y_k) = 0$	$\rho(Z_i; Y_k) \neq 0$	(i = 1,2,4,5), (k = 1-20)
$\rho(X_{i,j}; Y_k) = 0$	$\rho(X_{i,j}; Y_k) \neq 0$	(i = 1-8), (j = 1-7), (k = 1-20)
$\rho(S_i; Y_k) = 0$	$\rho(S_i; Y_k) \neq 0$	(i = 1-8), (k = 1-20)
$\rho(V_i; T) = 0$	$\rho(V_i; T) \neq 0$	(i = 1,2,3,4,5,7)
$\rho(W_i; T) = 0$	$\rho(W_i; T) \neq 0$	(i = 1-5)
$\rho(Z_i; T) = 0$	$\rho(Z_i; T) \neq 0$	(i = 1,2,4,5)
$\rho(X_{i,j}; T) = 0$	$\rho(X_{i,j}; T) \neq 0$	(i = 1-8), (j = 1-7)
$\rho(S_i; T) = 0$	$\rho(S_i; T) \neq 0$	(i = 1-8)

Note: See Table 7 for definitions of: V_i; W_i; Z_i; $X_{i,j}$; S_i; and Y_i.

where n is the sample size and r is the sample Pearson product-moment correlation coefficient. To determine significance, the computed t's were compared to a student's t-distribution with n-2 degrees of freedom. All hypotheses were tested at the α=.01 level of significance.

In addition to this pairwise correlation analysis, a multiple linear regression was also conducted using T (the combined teacher rating) as the dependent variable and the following twenty-three variables as independent variables: V_i (i = 1, 2, 3, 4, 5, 7), W_i(i = 1-5), Z_i (i = 1, 2, 4, 5) and S_i (i = 1-8). From this regression analysis the sample multiple correlation coefficient between T and the V_i, W_i, Z_i, and S_i variables was calculated. The null hypothesis that the population R = 0 was tested using an F-test on the statistic:

$$F = \frac{R^2 \ / \ p}{(1-R^2) \ / \ n_1}$$

where $df_1 = p$ and $df_2 = n_1 = n-p-1$ are the degrees of freedom (n = sample size and p = number of independent variables). The test was made at the P = .01 probability level. A measure of the predictive validity of the regression model was determined using a double cross-validation technique developed by Mosier (1951). To do this it was necessary to randomly divide the regression data into two separate samples and to compute multiple R's for each sample. Each

of these samples was also used to determine separate multiple regression equations involving the variables just named. Each of these regression equations was then applied to the data of the other sample. An estimate of the predictive validity of the regression model (model developed from the complete set of data) was then made by computing the Pearson product-moment correlations between the measured value of the independent variable of one sample and the value predicted for that independent variable from the regression equation of the other sample. This procedure is explained in greater detail in CHAPTER V.

Using the regression model just described, another model was developed. This model was designed to adjust or transform the teacher rating variable T in such a way that any multiple linear relationship found using the above procedure would be eliminated. The effect of this transformation on teacher ranking was determined by comparison of teacher rankings using the unadjusted ratings with teacher rankings using the adjusted ratings. This model is further discussed in CHAPTER V.

PRECAUTIONS TAKEN FOR ACCURACY

All computations in this study were done using the modern digital computers at Montana Tech and Montana State University. All keypunched data was verified.

SUMMARY

The procedures followed in this study resulted from the correlational design of the research. Since this research attempted to determine and alter relationships present in student ratings of teachers at Montana Tech and to develop a new evaluation model from these relationships, it was necessary to use only those rating controls that had existed prior to the study. This resulted in a sample that consisted of 93% of the teaching faculty as ratees and those students present at the time of rating as the raters.

Reliability of the instrument was measured by means of a test-retest application of the rating form. PART I and PART II of the form were found to possess a high degree of test-retest reliability while PART III and PART IV were found to be somewhat less reliable. The composite attitude measurements of PART III and the composite teacher rating of PART IV were, however, found to possess acceptable test-retest reliability. Concurrent validity of various items of PART I and PART II was ascertained by comparing student responses to items of this part with the Registrar's records. Validity of attitude measurement of PART III was verified by reference to the works of Osgood (1957) and Shaw and Wright (1967). Validity of the teacher rating items of PART IV was verified by reference to the works Ryans (1960) and Combs (1969).

Results of this research appear, primarily, in the form of correlation coefficients and linear regression equations; these results are displayed in table form. Because of the many variables under investigation, these variables were assigned letter names (Table 7) and are referenced accordingly in the remainder of this paper.

Analysis performed included the testing of statistical hypotheses dealing with pairwise and multiple coefficients of correlation between student-contextual variables and teacher ratings. Development and validation of a regression model was also included in the analysis.

CHAPTER IV

FINDINGS

INTRODUCTION

CHAPTER IV, a report and interpretation of the findings of this study, is organized according to the various variables involved, and thus, is divided into the following three divisions: INTRODUCTION, FINDINGS, and SUMMARY.

The INTRODUCTION describes the organization of the chapter. Included here is a listing of the major and minor subdivisions of the chapter; also included is a description of the way in which the findings are related.

FINDINGS reports and interprets the research findings of this study. For the most part these findings appear in the form of correlation coefficients that were computed in testing the hypotheses listed in CHAPTER III. These coefficients, organized with respect to the variables involved in the various hypotheses of CHAPTER III, all appear in table form and are interpreted in the text. FINDINGS is itself divided into subdivisions entitled: Contextual Variables, Non-Attitudinal Student Variables, Student Attitude Toward the Classroom, Student Attitude Toward the Occupation of Teaching, Student Attitude Toward The Hour the Class is Held, Student Attitude Toward Teacher Evaluations, Student Attitude Toward Montana Tech,

Student Attitude Toward Topics of the Course, Student Attitude Toward Grade Expected in Course, Student Attitude Toward the Way He Feels, and Multiple Regression Findings. Each of these subdivisions is again subdivided into the following two sections: Results of hypotheses tested and Interpretation of results.

The SUMMARY brings together the major findings of the study, and also, answers the questions listed in CHAPTER I. The SUMMARY does not attempt to generalize these findings nor does it attempt to qualify them since this is done in CHAPTER VI.

Throughout this chapter repeated references are made to Tables 9 through 18; and as can be seen, these tables all contain correlation coefficients and all have, essentially, the same format. The Y_i variables listed in the heading of each table (under Rating Variables) correspond respectively to each of the twenty scales used on the teacher rating form (PART IV) shown in APPENDIX B. The variable T, as defined in CHAPTER III, is the composite teacher rating (average of nonzero responses to twenty scales just mentioned). The extreme left column of each of these tables contains a list of the various contextual, student, or student attitude variables referenced in the table. The body of each of these tables consists of correlation coefficients between each of the rating variables listed in the heading and each of the contextual, student, or student attitude variables listed in the extreme left column. Each of these

correlation coefficients resulted in the rejection or acceptance of one of the null hypotheses stated in CHAPTER III. Any of these coefficients that were found to be sufficiently large to warrant the rejection of the corresponding null hypothesis is designated with an *. The n in each table refers to the number of paired observations used in the computation of correlation coefficients involving T, the composite teacher rating. The various letter names for the variables named in these tables and used throughout this chapter are defined in Table 7 on page 83.

FINDINGS

Contextual Variables

Results of hypotheses tested. As shown, Table 9 contains correlation coefficients computed in testing null hypotheses involving contextual variables V_1, V_2, V_3, V_4, V_5, and V_7. From these correlation coefficients the following null hypotheses were rejected at the $\alpha = .01$ level of significance:

$H_0: \rho(V_1; Y_k) = 0$ for k = 9 and 19

$H_0: \rho(V_3; Y_k) = 0$ for k = 3, 5, 9, and 17

$H_0: \rho(V_4; Y_k) = 0$ for k = 3, 4, 6, 13, 15, and 18

$H_0: \rho(V_5; Y_k) = 0$ for all k except k = 2, 11, and 17

$H_0: \rho(V_5; T) = 0$

Table 9

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Certain Contextual Variables

Context. Variable	Rating Variable																				n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
V_1	.03	-.01	.02	-.02	-.03	-.07	-.02	-.02	.01	.05	.03	-.04	.00	-.01	-.01	.05	.03	.03	.07*	.02	.02	1.925
V_2	.02	-.01	-.01	.01	-.01	.00	.01	-.01	.03	.03	.00	.01	.01	.01	.00	.01	.02	-.01	.02	.02	.01	1,850
V_3	.05	.06	.07*	.04	.13*	.03	.02	.02	.09*	-.01	-.04	.06	.03	-.01	.04	.00	-.09*	.06	.03	.01	.05	1,925
V_4	.03	-.01	.09*	.09*	.01	.09*	-.02	.05	.03	.03	-.04	.05	.07*	-.03	.07*	.03	.00	.07*	.02	.04	.06	1,925
V_5	-.16*	-.02	-.15*	-.13*	-.08*	-.13*	-.14*	-.12*	-.16*	-.10*	-.05	-.13*	-.13*	-.07*	-.07*	-.12*	-.03	-.16*	-.14*	-.15*	-.17*	1,925
V_7	.06	.01	.02	.07*	.03	.07*	.05	.06	.06	.07*	.07*	.06	.07*	.06	.12*	.10*	.11*	.05	.07*	.10*	.09*	1,925

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

V_i (i=1,2,3,4,5,7) refers to the various contextual variables as defined in Table 7.

$H_0: \rho(V_7; Y_k) = 0$ for k = 4, 6, 10, 11, 13, 15, 16, 17, 19, 20

$H_0: \rho(V_7; T) = 0$

Interpretation of results. Contextual variable V_1 (the day of the week on which rating took place) was found to be significantly related to only two of the rating variables Y_9 (DULL--STIMULATING) and Y_{19} (I LEARNED LITTLE--I LEARNED MUCH). Both correlation coefficients were positive indicating that in the later part of the week the students rated the teachers as somewhat more stimulating and also rated their own learning somewhat higher than they did earlier in the week. V_1 was also found not to be significantly related to T (the composite teacher rating). The coefficient of determination (r^2) associated with variable V_1 also indicates that less than .01 of the variance of any rating variable was accounted for by V_1.

Contextual Variable V_2 (the period during which the class was held) was not found to be significantly related to any of the rating variables.

Contextual variable V_3 (the level of the course) was found to be significantly related to rating variables Y_3 (SUBJECT CENTERED--STUDENT CENTERED), Y_5 (CONTROLS STUDENTS--FREES STUDENTS), Y_9 (DULL--STIMULATING), and Y_{17} (UNPREPARED--PREPARED). The signs on these coefficients indicate that teachers of upper level courses tended to

be rated as being somewhat more student centered, student freeing, and stimulating than were teachers of lower level courses. Teachers of lower level courses were, however, rated as being more prepared than were teachers of higher level courses. Because V_3 was not found to be significantly related to T, (the composite rating) and because r^2 values were all less than .01 it appears that very little of the rating variance was accounted for by course level.

Contextual variable V_4 (sex of the teacher) although not found to be significantly related to the composite teacher rating, was found to be significantly related to: Y_3 (SUBJECT CENTERED--STUDENT CENTERED), Y_4 (UNINVOLVED--INVOLVED), Y_6 (UNFRIENDLY--FRIENDLY), Y_{13} (NARROW--BROAD), Y_{15} (SLOPPY--NEAT), and Y_{18} (BAD COMMUNICATOR--GOOD COMMUNICATOR). The signs on the coefficients indicate that female teachers tended to be rated as being somewhat more student centered, involved, friendly, broad, and neat than were male teachers. Female teachers were also rated as somewhat better communicators than were male teachers. Again, however, r^2 values were all found to be less than .01.

Contextual variable V_5 (number of students in the class) was found to be significantly related to every rating variable except Y_2 (STRESSES FACTS--STRESSES CONCEPTS), Y_{11} (DISORGANIZED--ORGANIZED), and Y_{17} (UNPREPARED--PREPARED). All coefficients were found to have negative signs indicating that teachers with larger

classes tended to receive lower ratings than teachers of the smaller classes. The coefficient of determination (r^2) for V_5 (with respect to the composite rating T) was found to be .03; thus indicating that as much as .03 of the variance of T could be attributed to the number of students in the class.

Contextual variable V_7 (the difference between the sex of the teacher and the sex of the student) was found to be significantly related to over half of the rating variables, including T, the composite rating. The sign of these coefficients indicate that teachers tended to be rated higher by students of the opposite sex. The coefficients of determination again indicated, however, that less than .01 of the rating variance was accounted for by this variable.

It is of interest to note that the rating variable Y_2 (STRESSES FACTS--STRESSES CONCEPTS) was not found to be related to any of the contextual variables.

Non-Attitudinal Student Variables

Results of hypotheses tested. As shown, Table 10 contains correlation coefficients computed in testing null hypotheses involving student variables W_1, W_2, W_3, W_4, W_5, Z_1, Z_2, Z_4, and Z_5. From these correlation coefficients the following null hypotheses were rejected at the $\alpha = .01$ level of significance:

Table 10

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Certain Student Variables

Student Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
W_1	.00	.09*	-.07*	-.01	.01	.05	.03	.03	-.01	.06	.05	-.04	.00	.05	-.02	.05	.03	.01	.05	.02	.03	1,497
W_2	-.10*	.03	-.07*	-.07*	-.01	-.08*	-.08*	-.08*	-.11*	-.08*	-.03	-.09*	-.07*	-.07*	-.11*	-.05	-.07*	-.07*	-.06	-.08*	-.10*	1,911
W_3	.00	-.03	.03	.01	.06	.03	.03	.01	.04	.00	-.03	.01	.01	-.01	.01	.04	-.01	.03	.00	.01	.03	1,241
W_4	-.03	-.04	-.06	.00	-.05	-.01	-.02	-.01	-.02	-.02	.04	-.05	-.01	.04	-.02	-.03	.03	-.03	.01	.00	-.02	1,925
W_5	.01*	.02	.05*	.08*	.09*	.11*	.14*	.10*	.19*	.10*	.01	.12*	.11*	.08*	.10*	.07*	.04	.11*	.11*	.08*	.14*	1,894
Z_1	.11*	-.01	.07*	.10*	.06	.09*	.10*	.13*	.11*	.06	.03	.10*	.07*	.06	.05	.10*	.09*	.11*	.11*	.10*	.13*	1,918
Z_2	.05	-.01	.02	.06	.07*	.10*	.07*	.07*	.08*	.06	.03	.08*	.08*	.06	.15*	.12*	.12*	.05	.07*	.10*	.11*	1,916
Z_4	-.24*	-.07	-.20*	-.22*	-.10*	-.22*	-.22*	-.22*	-.21*	-.12*	-.05	-.19*	-.20*	-.16*	-.09*	-.22*	-.07*	-.23*	-.31*	-.27*	-.28*	1,882
Z_5	-.12*	.02	-.09*	-.09*	-.06	-.08*	-.08*	-.07*	-.12*	-.04	-.01	-.13*	-.06	-.04	-.06	-.05	.00	-.10*	-.09*	-.11*	-.11*	1,906

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

W_i (i=1-5) and Z_i (i=1,2,4,5) refer to various student variables as defined in Table 7.

$H_0:\rho(W_1;Y_k) = 0$ for k = 2 and 3

$H_0:\rho(W_2;Y_k) = 0$ for all k except k = 2, 5, 11, 16, and 19

$H_0:\rho(W_2;T) = 0$

$H_0:\rho(W_5;Y_k) = 0$ for all k except k = 2, 8, 11, and 17

$H_0:\rho(W_5;T) = 0$

$H_0:\rho(Z_1;Y_k) = 0$ for all k except k = 2, 5, 10, 11, 14, and 15

$H_0:\rho(Z_1;T) = 0$

$H_0:\rho(Z_2;Y_k) = 0$ for all k except k = 1, 2, 4, 10, 11, 14, and 18

$H_0:\rho(Z_2;T) = 0$

$H_0:\rho(Z_4;Y_k) = 0$ for all k except k = 2 and 11

$H_0:\rho(Z_4;T) = 0$

$H_0:\rho(Z_5;Y_k) = 0$ for k = 1, 3, 4, 6, 7, 8, 9, 12, 18, and 19

$H_0:\rho(Z_5;T) = 0$

Interpretation of results. Student variable W_1 (the cumulative grade point average of the student) was found to be significantly related to only two of the rating variables, Y_2 (STRESSES FACTS--STRESSES CONCEPT) and Y_3 (SUBJECT CENTERED--STUDENT CENTERED). These coefficients indicate that the students with higher grade point averages tended to rate their teachers as stressing concepts and being more subject centered than did students with lower grade point averages. All r^2 values, however, were again found to be less than .01. The relationship between student grade point

average and the overall or composite rating was also not found to be significant.

Student variable W_2 (the number of credits the student was taking) was found to be negatively and significantly related to over three fourths of all the rating variables. Thus, although all r^2 values were again less than or equal to .01 there was an indication that students taking a larger number of credits gave lower teacher ratings than students taking fewer credits.

Student variable W_3 (the number of years the student had known the teacher) was not found to be significantly related to any of the rating variables.

Student variable W_4 (the number of exams the student had taken on the day of rating prior to rating the teacher), like variable W_3, was also not found to be significantly related to any of the rating variables.

Student variable W_5 (the age of the student) was found to be positively and significantly related to all rating variables except: Y_2 (STRESSES FACTS--STRESSES CONCEPTS), and Y_{17} (UNPREPARED--PREPARED). With these two exceptions, it appears that older students tended to give somewhat higher ratings than did younger students, a surprising result considering class level was not found to be significantly related to the overall rating. Apparently students at Montana Tech are not uniformly distributed by age among the various

class levels. The coefficient of determination for variable W_5 (with respect to T) was found to be .02.

Student variable Z_1 (whether the course was required or elective for the student) was found to be significantly related to all rating variables except: Y_2 (STRESSES FACTS--STRESSES CONCEPTS), Y_5 (CONTROLS STUDENTS--FREES STUDENTS), Y_{10} (UNCERTAIN--CONFIDENT), Y_{11} (DISORGANIZED--SYSTEMATIC), Y_{14} (PARTIAL GRADER--FAIR GRADER), and Y_{15} (SLOPPY--NEAT). Correlation coefficients indicate that students in elective courses tended to give teachers higher overall ratings (including composite rating T) than did students in required courses. Again the coefficient of determination was (with respect to T) found to be .02.

Student variable Z_2 (sex of the student) was found to be significantly related to T (composite rating) and also to over half of the Y_k's. Rating items showing the largest significant correlation with Z_2 were Y_{16} (CONFUSED OBJECTIVES--CLEAR OBJECTIVES) and Y_{17} (UNPREPARED--PREPARED). For these items the female students tended to give higher ratings than did the male students. A coefficient of determination of .01 indicates, however, that less than .01 of the composite rating variance was attributable to student sex.

Student variable Z_4 (the grade the student expected to receive in the course) was found to be significantly related to

every rating variable except: Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_{11} (DISORGANIZED--SYSTEMATIC). The negative correlation coefficients all indicate that students expecting the higher grades tended to give the teachers higher ratings than did the students expecting lower grades. It should be noted that an A was coded as a 1, B as a 2, C as a 3, D as a 4, and F as a 5. As can be seen in Table 10 the variable Z_4 has a coefficient of determination of .08 (with respect to T), the largest such reported in either Table 9 or Table 10.

Student variable Z_5 (whether the student's major was the same or different than the teachers) was found to be significantly related to T, the composite rating, and also to half of the Y_K rating variables. The negative signs on all significant correlation coefficients indicate that students having the same major as that of the teacher tended to give the teacher higher overall ratings than did the students whose major was different than the teacher's. Again, however, the coefficient of determination (with respect to T) was only .01.

As shown in Table 10 neither of the rating variables Y_2 (STRESSES FACTS--STRESSES CONCEPTS) nor Y_{11} (DISORGANIZED--SYSTEMATIC) was found to be significantly related to any of the student variables shown in that table.

Student Attitude Toward the Classroom

Results of hypotheses tested. Table 11 contains correlation coefficients computed in testing hypotheses involving student attitude toward the classroom. The first seven entries of column one of Table 11 correspond respectively to each of the seven semantic differential scales used in measuring the attitude. The eighth entry in column one, S_1, represents the composite attitude measurement (average of all seven scales). The correlation coefficients in the body of the table were computed in testing hypotheses involving the various attitude scales and the various teacher rating scales. With the exception of different attitude variables involved, Tables 12 through 18 follow the same format as Table 11.

With reference to the correlation coefficients of Table 11 the following null hypotheses were rejected at the $\alpha = .01$ level of significance:

$H_0: \rho(X_{1,1}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{1,1}; T) = 0$

$H_0: \rho(X_{1,2}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{1,2}; T) = 0$

$H_0: \rho(X_{1,3}; Y_k) = 0$ for all k except k = 2

$H_0: \rho(X_{1,3}; T) = 0$

$H_0: \rho(X_{1,4}; Y_k) = 0$ for all k except k = 2

$H_0: \rho(X_{1,4};T) = 0$

$H_0: \rho(X_{1,5};Y_k) = 0$ for all k except k = 2

$H_0: \rho(X_{1,5};T) = 0$

$H_0: \rho(X_{1,6};Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{1,6};T) = 0$

$H_0: \rho(X_{1,7};Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{1,7};T) = 0$

$H_0: \rho(S_1;Y_k) = 0$ for all k except k = 2

$H_0: \rho(S_1;T) = 0$

Interpretation of results. Student attitude toward the classroom, as shown by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS). All significant coefficients, with the exception of those associated with rating variable Y_5 (CONTROLS STUDENTS--FREES STUDENTS), indicate that students who had a positive or favorable attitude toward the classroom tended to rate the teacher higher than did students who had an unfavorable attitude toward the classroom. With reference to rating variable Y_5, the students who had a more favorable attitude toward the classroom tended to rate the teacher as being more controlling than did the students with less favorable attitudes toward the classroom.

Table 11

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Classroom

Attitude Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{1,1}$	.26*	.00	.12*	.19*	-.05	.22*	.22*	.20*	.24*	.15*	.17*	.15*	.18*	.19*	.21*	.19*	.17*	.21*	.25*	.23*	.26*	1,882
$X_{1,2}$	.30*	-.01	.13*	.25*	-.05	.24*	.27*	.27*	.26*	.16*	.17*	.20*	.22*	.20*	.24*	.22*	.18*	.24*	.29*	.27*	.31*	1,735
$X_{1,3}$	.22*	.02	.09*	.17*	-.08*	.22*	.21*	.21*	.20*	.14*	.17*	.14*	.14*	.17*	.23*	.17*	.16*	.17*	.20*	.19*	.24*	1,797
$X_{1,4}$	.29*	-.01	.12*	.24*	-.06*	.22*	.25*	.28*	.27*	.19*	.20*	.20*	.22*	.20*	.23*	.24*	.18*	.24*	.28*	.28*	.31*	1,763
$X_{1,5}$	.30*	-.02	.15*	.25*	-.06*	.26*	.27*	.27*	.28*	.20*	.20*	.21*	.26*	.18*	.24*	.24*	.19*	.24*	.28*	.29*	.33*	1,742
$X_{1,6}$	.30*	-.01	.14*	.25*	-.05	.24*	.27*	.29*	.28*	.20*	.19*	.19*	.22*	.19*	.24*	.24*	.19*	.25*	.28*	.29*	.33*	1,747
$X_{1,7}$	.33*	-.01	.20*	.27*	-.04	.29*	.29*	.30*	.33*	.22*	.21*	.24*	.26*	.21*	.25*	.26*	.21*	.27*	.33*	.36*	.36*	1,726
S_1	.33*	.00	.16*	.27*	-.07*	.28*	.30*	.29*	.30*	.20*	.22*	.21*	.25*	.22*	.26*	.26*	.22*	.27*	.31*	.31*	.35*	1,899

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{1,j}$ (j=1-7) refers to individual attitude scales on PART III of rating form.

S_1 refers to the composite attitude for attitude variable 1 on PART III of the rating form.

Student Attitude Toward Occupation of Teaching

Results of hypotheses tested. Table 12 contains correlation coefficients found in testing hypotheses involving rating variables and various attitude scales used in measuring student attitude toward the occupation of teaching. With reference to the correlation coefficients of Table 12, the following null hypotheses were rejected at the $\alpha = .01$ level of significance:

$H_0: \rho(X_{2,1}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,1}; T) = 0$

$H_0: \rho(X_{2,2}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,2}; T) = 0$

$H_0: \rho(X_{2,3}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,3}; T) = 0$

$H_0: \rho(X_{2,4}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,4}; T) = 0$

$H_0: \rho(X_{2,5}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,5}; T) = 0$

$H_0: \rho(X_{2,6}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,6}; T) = 0$

$H_0: \rho(X_{2,7}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{2,7}; T) = 0$

$H_0: \rho(S_2; Y_k) = 0$ for all k except k = 2 and 5

Table 12

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Occupation of Teaching

Attitude Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{2,1}$	.48*	-.03	.22*	.36*	.02	.40*	.43*	.39*	.42*	.31*	.28*	.34*	.35*	.30*	.36*	.41*	.31*	.41*	.41*	.44*	.51*	1,865
$X_{2,2}$	.40*	-.01	.19*	.35*	.02	.39*	.39*	.37*	.39*	.30*	.26*	.31*	.32*	.27*	.37*	.37*	.28*	.37*	.37*	.40*	.47*	1,760
$X_{2,3}$	.39*	-.01	.19*	.30*	-.01	.36*	.36*	.35*	.35*	.31*	.27*	.27*	.32*	.27*	.36*	.35*	.30*	.33*	.34*	.37*	.44*	1,758
$X_{2,4}$	.40*	.01	.16*	.31*	-.01	.33*	.36*	.35*	.37*	.29*	.23*	.28*	.30*	.27*	.33*	.34*	.26*	.35*	.35*	.38*	.43*	1,765
$X_{2,5}$	.40*	.03	.17*	.31*	.03	.34*	.38*	.34*	.38*	.32*	.27*	.29*	.33*	.28*	.31*	.38*	.31*	.34*	.35*	.38*	.46*	1,757
$X_{2,6}$	.42*	.01	.20*	.34*	.03	.37*	.41*	.38*	.38*	.31*	.25*	.33*	.33*	.32*	.36*	.33*	.28*	.38*	.36*	.43*	.48*	1,745
$X_{2,7}$	.34*	.03	.20*	.31*	.01	.31*	.36*	.33*	.31*	.27*	.19*	.28*	.26*	.24*	.30*	.33*	.23*	.32*	.32*	.42*	.41*	1,718
S_2	.48*	-.01	.24*	.39*	.02	.43*	.46*	.42*	.44*	.36*	.29*	.36*	.38*	.33*	.40*	.44*	.35*	.42*	.42*	.47*	.54*	1,881

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{2,j}$ (j=1-7) refers to individual attitude scales on PART III of rating form.

S_2 refers to the composite attitude for attitude variable 2 on PART III of the rating form.

$H_0: \rho(S_2;T) = 0$

Interpretation of results. Student attitude toward the occupation of teaching, as indicated by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). With these two exceptions the correlation coefficients all indicate that students who had a more favorable attitude toward the occupation of teaching tended to give their teachers higher ratings than did the students with a less favorable attitude toward the occupation of teaching.

For the composite attitude S_2 and the composite teacher rating T, the coefficient of determination was found to be .29 indicating that as much as .29 of the variance of the composite teacher rating was attributable to the student's attitude toward the occupation.

Student Attitude Toward the Hour the Class Was Held

Results of hypotheses tested. Table 13 contains correlation coefficients found in testing hypotheses involving rating variables and the various attitude scales used in measuring student attitude toward the hour the class was held. With reference to these correlation coefficients, the following null hypotheses were rejected

at the $\alpha = .01$ level of significance:

$H_0: \rho(X_{3,1}; Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0: \rho(X_{3,1}; T) = 0$

$H_0: \rho(X_{3,2}; Y_k) = 0$ for all k except k = 2,3, and 5

$H_0: \rho(X_{3,2}; T) = 0$

$H_0: \rho(X_{3,3}; Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0: \rho(X_{3,3}; T) = 0$

$H_0: \rho(X_{3,4}; Y_k) = 0$ for all k except k = 2 and 3

$H_0: \rho(X_{3,4}; T) = 0$

$H_0: \rho(X_{3,5}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{3,5}; T) = 0$

$H_0: \rho(X_{3,6}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{3,6}; T) = 0$

$H_0: \rho(X_{3,7}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{3,7}; T) = 0$

$H_0: \rho(S_3; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(S_3; T) = 0$

Interpretation of results. Student attitude toward the classroom, as indicated by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). The significant correlation coefficients all indicate that the

Table 13

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Hour the Class was Held

Attitude Variable	Rating Variable																				n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{3,1}$	.17*	-.03	.05	.12*	-.03	.17*	.14*	.14*	.17*	.13*	.14*	.10*	.12*	.15*	.17*	.14*	.18*	.18*	.19*	.21*	.20*	1,882
$X_{3,2}$	.19*	-.01	.04	.15*	-.05	.16*	.15*	.17*	.19*	.17*	.17*	.13*	.16*	.16*	.21*	.14*	.18*	.18*	.20*	.21*	.22*	1,730
$X_{3,3}$	.20*	-.01	.05	.14*	-.04	.20*	.18*	.17*	.20*	.18*	.21*	.14*	.16*	.19*	.23*	.18*	.21*	.21*	.22*	.23*	.25*	1,687
$X_{3,4}$	.21*	-.01	.05	.15*	-.07*	.20*	.20*	.21*	.21*	.17*	.18*	.16*	.19*	.18*	.22*	.17*	.22*	.22*	.24*	.24*	.26*	1,710
$X_{3,5}$	.21*	.03	.07*	.15*	-.06	.20*	.19*	.20*	.21*	.18*	.20*	.18*	.21*	.19*	.21*	.19*	.21*	.21*	.21*	.24*	.26*	1,703
$X_{3,6}$	.24*	.01	.07*	.17*	-.04	.24*	.24*	.23*	.24*	.13*	.21*	.19*	.20*	.21*	.23*	.19*	.23*	.23*	.26*	.29*	.29*	1,706
$X_{3,7}$	.19*	.04	.08*	.14*	-.04	.20*	.17*	.19*	.20*	.27*	.18*	.16*	.17*	.17*	.21*	.18*	.19*	.19*	.22*	.29*	.24*	1,697
S_3	.22*	-.01	.06*	.16*	-.03	.22*	.19*	.20*	.23*	.17*	.19*	.16*	.18*	.19*	.23*	.23*	.18*	.24*	.25*	.27*	.27*	1,892

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{3,j}$ (j=1-7) refers to individual attitude scales on PART III of rating form.

S_3 refers to the composite attitude for attitude variable 3 on PART III of the rating form.

students who had a more favorable attitude toward the classroom tended to give the teacher a higher rating than the students who had a less favorable attitude. The correlation coefficient of .06 between the composite attitude S_3 and the rating variable Y_3 (SUBJECT CENTERED--STUDENT CENTERED), although significant, is only one fourth to one half the magnitude of the other coefficients.

For the composite attitude variable S_3 and the composite teacher rating T the coefficient of determination was found to be .07 indicating that as much as .07 of the variance of T was attributable to student attitude toward the classroom.

Student Attitude Toward Teacher Evaluations

Results of hypotheses tested. Table 14 contains correlation coefficients found in testing hypotheses involving rating variables and the various attitude scales used in measuring student attitude toward teacher evaluations. These correlation coefficients resulted in the rejection of the following null hypotheses at the $\alpha = .01$ level of significance:

$H_0: \rho(X_{4,1}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,1}; T) = 0$

$H_0: \rho(X_{4,2}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,2}; T) = 0$

$H_0: \rho(X_{4,3}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,3}; T) = 0$

$H_0: \rho(X_{4,4}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,4}; T) = 0$

$H_0: \rho(X_{4,5}; Y_k) = 0$ for all k except k = 2

$H_0: \rho(X_{4,5}; T) = 0$

$H_0: \rho(X_{4,6}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,6}; T) = 0$

$H_0: \rho(X_{4,7}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{4,7}; T) = 0$

$H_0: \rho(S_4, Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(S_4; T) = 0$

Interpretation of results. Student attitude toward teacher evaluations, as indicated by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). Again the positive correlation coefficients indicate that students with a more favorable attitude toward teacher evaluations tended to give the teacher a higher rating than did students with a less favorable attitude toward teacher evaluations.

The portion of the composite rating variance accounted for by student attitude toward teacher evaluations, as measured by

Table 14

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Teacher Evaluations

Attitude Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{4,1}$	.40*	-.04	.18*	.33*	-.02	.33*	.35*	.36*	.35*	.27*	.22*	.25*	.28*	.23*	.24*	.31*	.25*	.34*	.33*	.31*	.40*	1,874
$X_{4,2}$	.32*	-.02	.16*	.29*	-.04	.30*	.30*	.31*	.27*	.20*	.17*	.26*	.26*	.19*	.25*	.24*	.20*	.25*	.24*	.24*	.34*	1,783
$X_{4,3}$	.29*	-.01	.11*	.25*	-.06	.27*	.26*	.29*	.26*	.20*	.19*	.19*	.22*	.18*	.27*	.22*	.21*	.23*	.22*	.24*	.31*	1,759
$X_{4,4}$	.34*	.00	.11*	.28*	-.05	.26*	.32*	.34*	.31*	.26*	.19*	.23*	.24*	.20*	.23*	.29*	.23*	.29*	.27*	.27*	.35*	1,786
$X_{4,5}$	.33*	-.01	.11*	.27*	-.06*	.27*	.30*	.32*	.31*	.24*	.21*	.20*	.23*	.19*	.24*	.26*	.25*	.26*	.27*	.27*	.34*	1,761
$X_{4,6}$	.33*	-.02	.14*	.29*	-.03	.30*	.31*	.33*	.29*	.24*	.20*	.23*	.25*	.21*	.25*	.27*	.23*	.27*	.26*	.29*	.35*	1,761
$X_{4,7}$	.31*	-.02	.16*	.29*	-.03	.28*	.31*	.31*	.29*	.25*	.17*	.23*	.23*	.18*	.20*	.26*	.21*	.26*	.25*	.33*	.34*	1,721
S_4	.38*	-.04	.17*	.32*	-.04	.33*	.35*	.37*	.33*	.26*	.22*	.25*	.28*	.22*	.26*	.30*	.26*	.31*	.30*	.32*	.39*	1,892

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{4,j}$ (j=1-7) refers to individual attitude scales on PART III of rating form.

S_4 refers to the composite attitude for attitude variable 4 on PART III of the rating form.

the coefficient of determination, was found to be .15.

Student Attitude Toward Montana Tech

Results of hypotheses tested. Table 15 shows the correlation coefficients computed in testing hypotheses involving teacher rating variables and the various attitude scales used in measuring student attitude toward Montana Tech. These correlation coefficients resulted in the rejection (at the $\alpha = .01$ level of significance) of the following null hypotheses:

$H_0: \rho(X_{5,1}; Y_k) = 0$ for all k except $k = 2$ and 5

$H_0: \rho(X_{5,1}; T) = 0$

$H_0: \rho(X_{5,2}; Y_k) = 0$ for all k except $k = 2$ and 5

$H_0: \rho(X_{5,2}; T) = 0$

$H_0: \rho(X_{5,3}; Y_k) = 0$ for all k except $k = 2,3$ and 5

$H_0: \rho(X_{5,3}; T) = 0$

$H_0: \rho(X_{5,4}; Y_k) = 0$ for all k except $k = 2,3$ and 5

$H_0: \rho(X_{5,4}; T) = 0$

$H_0: \rho(X_{5,5}; Y_k) = 0$ for all k except $k = 2$ and 5

$H_0: \rho(X_{5,5}; T) = 0$

$H_0: \rho(X_{5,6}; Y_k) = 0$ for all k except $k = 2$ and 5

$H_0: \rho(X_{5,6}; T) = 0$

$H_0: \rho(X_{5,7}; Y_k) = 0$ for all k except $k = 2$ and 5

$H_0: \rho(X_{5,7}; T) = 0$

Table 15

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Montana Tech

Attitude Variable	Rating Variable																				n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{5,1}$	.22*	.01	.09*	.17*	-.03	.19*	.21*	.20*	.22*	.18*	.17*	.17*	.17*	.18*	.22*	.23*	.16*	.21*	.23*	.24*	.26*	1,863
$X_{5,2}$	.24*	.03	.08*	.20*	-.01	.20*	.21*	.23*	.24*	.20*	.19*	.21*	.18*	.16*	.26*	.23*	.19*	.22*	.23*	.25*	.29*	1,748
$X_{5,3}$	.21*	.01	.04	.17*	-.04	.15*	.19*	.17*	.22*	.18*	.19*	.15*	.15*	.18*	.21*	.21*	.16*	.21*	.19*	.20*	.24*	1,759
$X_{5,4}$	.23*	.00	.03	.19*	-.05	.16*	.22*	.22*	.22*	.20*	.18*	.20*	.18*	.19*	.20*	.26*	.17*	.21*	.22*	.22*	.26*	1,783
$X_{5,5}$	.22*	.03	.06*	.17*	-.01	.21*	.21*	.22*	.19*	.21*	.20*	.19*	.20*	.18*	.20*	.23*	.19*	.19*	.22*	.22*	.27*	1,777
$X_{5,6}$	.26*	-.01	.06*	.21*	-.02	.22*	.24*	.24*	.24*	.22*	.20*	.20*	.19*	.18*	.25*	.27*	.19*	.23*	.24*	.25*	.29*	1,744
$X_{5,7}$	.20*	.02	.07*	.15*	-.05	.19*	.21*	.20*	.20*	.20*	.16*	.15*	.15*	.13*	.21*	.21*	.17*	.18*	.20*	.25*	.24*	1,706
S_5	.26*	.01	.08*	.20*	-.03	.21*	.23*	.23*	.25*	.23*	.21*	.20*	.20*	.20*	.24*	.27*	.20*	.24*	.25*	.25*	.30*	1,888

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{5,j}$ (j=1-7) refers to individual attitude scales on PART III of rating form.

S_5 refers to the composite attitude for attitude variable 5 on PART III of the rating form.

$H_0: \rho(S_5; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(S_5; T) = 0$

Interpretation of results. Student attitude toward Montana Tech, as indicated by the null hypotheses rejected, was found to be significantly related to all teacher rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). The rating variable Y_3 (SUBJECT CENTERED--STUDENT CENTERED), although significantly related to student attitude toward Montana Tech, was again found to have a correlation coefficient only one fourth to half the magnitude of the other significant variables.

The signs on the significant coefficients indicate that students having a more favorable attitude toward Montana Tech tended to rate the teacher higher than did the students with a less favorable attitude toward the school. The strength of the relationship, as measured by the coefficient of determination, was found to be .09.

Student Attitude Toward Topics of the Course

Results of hypotheses tested. Table 16 shows correlation coefficients computed in testing hypotheses involving teacher rating variables and the various attitude scales used in measuring student attitude toward topics of course. These coefficients resulted in the rejection (at the $\alpha = .01$ level of significance) of the following

null hypotheses:

$H_0: \rho(X_{6,1}; Y_k) = 0$ for all k except k = 5

$H_0: \rho(X_{6,1}; T) = 0$

$H_0: \rho(X_{6,2}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,2}; T) = 0$

$H_0: \rho(X_{6,3}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,3}; T) = 0$

$H_0: \rho(X_{6,4}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,4}; T) = 0$

$H_0: \rho(X_{6,5}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,5}; T) = 0$

$H_0: \rho(X_{6,6}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,6}; T) = 0$

$H_0: \rho(X_{6,7}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{6,7}; T) = 0$

$H_0: \rho(S_6; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(S_6; T) = 0$

Interpretation of results. Student attitude toward topics of the course, as the rejected null hypotheses show, was found to be strongly and positively related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). These coefficients indicate that students who had a more

Table 16

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Topics of the Course

Attitude Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{6,1}$	.56*	-.07*	.28*	.40*	.05	.43*	.50*	.45*	.52*	.34*	.30*	.42*	.41*	.34*	.36*	.45*	.33*	.46*	.58*	.55*	.59*	1,881
$X_{6,2}$	.47*	-.02	.27*	.38*	.03	.40*	.46*	.43*	.45*	.30*	.27*	.39*	.40*	.32*	.35*	.41*	.29*	.41*	.47*	.48*	.53*	1,725
$X_{6,3}$	.42*	-.01	.24*	.36*	.02	.40*	.43*	.41*	.40*	.32*	.26*	.36*	.37*	.32*	.39*	.39*	.29*	.37*	.43*	.45*	.57*	1,715
$X_{6,4}$	.48*	.01	.24*	.36*	.03	.38*	.47*	.43*	.45*	.35*	.30*	.40*	.39*	.35*	.34*	.43*	.31*	.43*	.50*	.49*	.55*	1,756
$X_{6,5}$	.42*	.01	.22*	.33*	.04	.37*	.41*	.40*	.42*	.30*	.26*	.36*	.37*	.34*	.32*	.39*	.29*	.38*	.44*	.45*	.50*	1,733
$X_{6,6}$	.46*	-.01	.25*	.38*	.03	.41*	.46*	.44*	.45*	.35*	.28*	.41*	.40*	.35*	.36*	.42*	.30*	.41*	.48*	.50*	.55*	1,727
$X_{6,7}$	.45*	-.01	.25*	.36*	.06	.37*	.45*	.40*	.45*	.36*	.26*	.41*	.37*	.31*	.32*	.42*	.29*	.41*	.48*	.55*	.53*	1,696
S_6	.55*	-.03	.30*	.43*	.05	.45*	.52*	.49*	.52*	.38*	.31*	.45*	.45*	.39*	.39*	.48*	.36*	.47*	.56*	.56*	.62*	1,891

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{6,j}$ (j=1-7) refers to the individual attitude scales on PART III of rating form.

S_6 refers to the composite attitude for attitude variable 6 on PART III of the rating form.

favorable attitude toward the topics of the course definitely tended to give the teacher a higher rating than did students having a less favorable attitude toward the course.

Of all variables investigated in this research, student attitude toward topics of course was found to be the most strongly related to the composite teacher rating T. The coefficient of determination between S_6 (composite attitude) and T was found to be .38, indicating that as much as .38 of the rating variance was attributable to student attitude toward topics of the course.

Since the coefficient of determination for the relationship between composite teacher rating and whether or not the course was required or elective was found to be only .02, it appears that favorable and unfavorable attitudes toward topics of the course were distributed in both required and elective courses.

Student Attitude Toward Grade Expected in Course

Results of hypotheses tested. Table 17 shows the correlation coefficients found in testing hypotheses involving teacher rating variables and the various attitude scales used in measuring the student's attitude toward the grade he expected to receive in the course (at the time of rating). As a result of these coefficients the following null hypotheses were rejected at $\alpha = .01$ level of significance:

$H_0: \rho(X_{7,1}; Y_k) = 0$ for all k except k = 2

$H_0: \rho(X_{7,1}; T) = 0$

$H_0: \rho(X_{7,2}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{7,2}; T) = 0$

$H_0: \rho(X_{7,3}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{7,3}; T) = 0$

$H_0: \rho(X_{7,4}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{7,4}; T) = 0$

$H_0: \rho(X_{7,5}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{7,5}; T) = 0$

$H_0: \rho(X_{7,6}; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(X_{7,6}; T) = 0$

$H_0: \rho(X_{7,7}; Y_k) = 0$ for all k except 2 and 5

$H_0: \rho(X_{7,7}; T) = 0$

$H_0: \rho(S_7; Y_k) = 0$ for all k except k = 2 and 5

$H_0: \rho(S_7; T) = 0$

Interpretation of results. The student's attitude toward the grade he expected to receive in the course, as indicated by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). The significant coefficients are, again, all positive indicating that students having

Table 17

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Grade Expected in the Course

Attitude Variable	Rating Variable																				n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{7,1}$	.43*	.01	.26*	.37*	.06*	.38*	.41*	.39*	.37*	.26*	.19*	.35*	.36*	.27*	.26*	.38*	.23*	.40*	.47*	.46*	.48*	1,869
$X_{7,2}$	.41*	-.01	.24*	.37*	.03	.37*	.41*	.39*	.34*	.26*	.19*	.35*	.36*	.29*	.28*	.38*	.23*	.37*	.43*	.45*	.47*	1,697
$X_{7,3}$	.37*	.02	.23*	.34*	.03	.37*	.39*	.37*	.35*	.27*	.19*	.32*	.34*	.27*	.28*	.36*	.25*	.35*	.40*	.43*	.46*	1,677
$X_{7,4}$	.39*	.02	.21*	.34*	.04	.37*	.40*	.39*	.35*	.24*	.17*	.36*	.34*	.28*	.26*	.37*	.21*	.37*	.43*	.44*	.46*	1,688
$X_{7,5}$	.35*	.04	.20*	.30*	.02	.32*	.36*	.34*	.33*	.24*	.19*	.31*	.32*	.28*	.23*	.34*	.23*	.33*	.40*	.40*	.43*	1,689
$X_{7,6}$	.40*	-.01	.23*	.35*	.03	.38*	.41*	.40*	.35*	.28*	.22*	.35*	.36*	.29*	.30*	.40*	.25*	.38*	.42*	.46*	.48*	1,682
$X_{7,7}$	.39*	-.01	.21*	.34*	.05	.37*	.39*	.37*	.34*	.27*	.20*	.32*	.33*	.27*	.26*	.36*	.27*	.37*	.45*	.48*	.46*	1,679
S_7	.44*	.01	.26*	.40*	.06	.41*	.44*	.43*	.38*	.29*	.21*	.38*	.38*	.31*	.29*	.41*	.26*	.40*	.47*	.49*	.51*	1,882

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{7,j}$ (j=1-7) refers to the individual attitude scales on PART III of rating form.

S_7 refers to the composite attitude for attitude variable 7 on PART III of the rating form.

a more favorable attitude toward the grade they were expecting in the course tended to rate the teacher higher than students who were less favorably disposed toward the grade they expected.

The coefficient of determination for the student's composite attitude toward the expected grade (with respect to the composite rating) was found to be .26. This value, when compared to the coefficient of determination (.09) between the student's expected grade and the composite rating, shows that of the two, the attitude variable is the more strongly related to the composite rating.

The Student's Attitude Toward the Way He Feels

Results of hypotheses tested. Table 18 shows the correlation coefficients computed in testing hypotheses involving rating scales and the various attitude scales used in measuring the student's attitude toward the way he felt at the time of the rating. As a result of these correlation coefficients the following null hypotheses were rejected at the $\alpha = .01$ level of significance:

$H_0: \rho(X_{8,1}; Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0: \rho(X_{8,1}; T) = 0$

$H_0: \rho(X_{8,2}; Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0: \rho(X_{8,2}; T) = 0$

$H_0: \rho(X_{8,3}; Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0: \rho(X_{8,3}; T) = 0$

$H_0:\rho(X_{8,4};Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0:\rho(X_{8,4};T) = 0$

$H_0:\rho(X_{8,5};Y_k) = 0$ for all k except k = 2 and 5

$H_0:\rho(X_{8,5};T) = 0$

$H_0:\rho(X_{8,6};Y_k) = 0$ for all k except k = 2, 3, and 5

$H_0:\rho(X_{8,6};T) = 0$

$H_0:\rho(X_{8,7};Y_k) =$ for all k except k = 2 and 5

$H_0:\rho(X_{8,7};T) = 0$

$H_0:\rho(S_8;Y_k) = 0$ for all k except k = 2 and 5

$H_0:\rho(S_8;T) = 0$

Interpretation of results. The student's attitude toward the way he felt at the time of rating, as indicated by the null hypotheses rejected, was found to be significantly related to all rating variables except Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS). The significant coefficients of Table 18 further indicate that students with a more favorable attitude toward the way they felt at the time of rating tended to rate the teacher higher than did the students with a less favorable attitude.

The coefficient of determination for the composite attitude (with respect to the composite rating) was found to be .09, indicating that as much as .09 of the composite rating variance was accounted

Table 18

Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Way He Feels

Attitude Variable	Rating Variable																					n
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	T	
$X_{8,1}$	.20*	.01	.10*	.16*	-.01	.20*	.20*	.16*	.19*	.16*	.14*	.15*	.18*	.17*	.19*	.22*	.14*	.18*	.20*	.27*	.25*	1,866
$X_{8,2}$	.20*	.03	.11*	.16*	-.01	.22*	.22*	.18*	.21*	.15*	.13*	.17*	.19*	.18*	.21*	.20*	.15*	.18*	.20*	.26*	.26*	1,748
$X_{8,3}$	.17*	.03	.06	.15*	.01	.20*	.20*	.17*	.19*	.18*	.16*	.16*	.17*	.20*	.24*	.19*	.16*	.18*	.1.	.24*	.25*	1,747
$X_{8,4}$	.21*	.01	.06	.19*	-.04	.20*	.23*	.20*	.25*	.19*	.17*	.19*	.22*	.19*	.21*	.23*	.16*	.22*	.23*	.28*	.28*	1,729
$X_{8,5}$	.20*	.02	.10*	.18*	-.02	.21*	.23*	.19*	.23*	.19*	.17*	.18*	.21*	.18*	.25*	.22*	.18*	.21*	.22*	.29*	.28*	1,728
$X_{8,6}$	.16*	.00	.06	.14*	-.01	.17*	.20*	.17*	.18*	.16*	.15*	.14*	.18*	.17*	.21*	.17*	.14*	.17*	.17*	.24*	.23*	1,763
$X_{8,7}$	.16*	.01	.07*	.13*	.00	.06*	.18*	.14*	.17*	.16*	.13*	.14*	.14*	.17*	.19*	.19*	.12*	.17*	.18*	.25*	.22*	1,712
S_8	.22*	.01	.10*	.18*	-.02	.23*	.23*	.20*	.22*	.20*	.17*	.19*	.22*	.20*	.22*	.23*	.16*	.21*	.22*	.31*	.28*	1,884

*Indicates coefficient was found to be significant at α=.01 level.

n designates the number of paired observations involving the composite rating T.

Y_k (k=1-20) refers to the individual teacher rating scale of PART IV of the rating form.

$X_{8,j}$ (j=1-7) refers to the individual attitude scales on PART III of rating form.

S_8 refers to the composite attitude for attitude variable 8 on PART III of the rating form.

for by this attitude.

Multiple Regression Findings

In order to estimate the relationship between the composite teacher rating and the combination of contextual, student, and student attitude variables just discussed a multiple linear regression analysis was made. The dependent or criterion variable for this analysis was T, the composite teacher rating. Independent variables for this multiple regression analysis, as stated in CHAPTER III, were the twenty-three contextual, student, and student attitude variables just reported. For convenience in interpreting the multiple regression findings these twenty-three independent variables are here named as follows:

X_1: Date on which rating took place

X_2: Period class met

X_3: Level of the course

X_4: Sex of the teacher

X_5: Number of students in the class

X_6: Difference between sex of teacher and student

X_7: Student grade point average

X_8: Number of credits student taking

X_9: Number of years student knew teacher

X_{10}: Number of exams student had on day of rating

X_{11}: Age of the student

X_{12}: Course required or elective

X_{13}: Sex of the student

X_{14}: Grade student expected in course

X_{15}: Difference between major of student and teacher

X_{16}: Student attitude toward classroom

X_{17}: Student attitude toward occupation of teaching

X_{18}: Student attitude toward the hour class was held

X_{19}: Student attitude toward teacher evaluations

X_{20}: Student attitude toward Montana Tech

X_{21}: Student attitude toward topics of the course

X_{22}: Student attitude toward grade expected in course

X_{23}: Student attitude toward the way he felt

Results and interpretation of regression analysis. In developing this regression model a sample of 1,396 complete sets of observations were used; and as can be seen in Table 9 through 18 this was a somewhat smaller sample than was used in computing the correlation coefficients. The reason for this was that some of the students failed to respond to all of the items on the instrument and in so doing eliminated one whole set of observations in the multiple regression analysis. In other words, only those data which included

responses for every variable were usable in the regression model. Thus, with respect to the dependent variable T and the independent variables just named the following multiple regression equation was constructed (the following are b coefficients and not β's):

$$T' = a + \sum_{i=1}^{23} b_i X_i \quad \text{where,}$$

$a = 2.537$

$b_1 = -.005$

$b_2 = -.001$

$b_3 = -.032$

$b_4 = -.110$

$b_5 = -.005$

$b_6 = .049$

$b_7 = -.139$

$b_8 = .001$

$b_9 = -.018$

$b_{10} = -.013$

$b_{11} = .004$

$b_{12} = .187$

$b_{13} = .049$

$b_{14} = -.098$

$b_{15} = -.032$

$b_{16} = .016$

$b_{17} = .146$

$b_{18} = -.002$

$b_{19} = .074$

$b_{20} = -.003$

$b_{21} = .332$

$b_{22} = .079$

$b_{23} = .033$

Using the regression coefficients from this equation, the multiple R correlation coefficient was computed, and the resulting value of .69 was found to be significant at the $\alpha = .01$ level of significance. R^2, the coefficient of multiple determination, was then found to be .48, indicating that as much as .48 of the variance of T was attributable to the combined dependent variables.

A validation of this regression model, using a technique developed by Mosier (1951), is further discussed in CHAPTER V. Also included in CHAPTER V is an extended model that attempts to eliminate this multiple relationship.

SUMMARY

The main research questions of this study as stated in CHAPTER I and as restated in the form of null hypotheses in CHAPTER III, were answered in CHAPTER IV. These questions all dealt with

the presence or absence of relationships between teacher ratings at Montana Tech and various contextual, student, and student attitude variables. In this regard, student ratings of teachers at Montana Tech were found to be significantly related to all of the following:

1. The number of students in the class,
2. The difference between the sex of the teacher and the sex of the student,
3. The number of credits the student was taking,
4. The age of the student,
5. Whether the course was required or elective,
6. The sex of the student,
7. The grade the student expected in the course (at the time of rating),
8. Whether the student's major was the same or different than the teacher's,
9. The student's attitude toward the classroom,
10. The student's attitude toward the occupation of teaching,
11. The student's attitude toward the hour the class was held,
12. The student's attitude toward teacher evaluations,
13. The student's attitude toward Montana Tech,
14. The student's attitude toward topics of the course,
15. The student's attitude toward the grade he was expecting

to receive,

16. The student's attitude toward the way he felt on the day of rating.

As indicated by the various coefficients of determination the strongest relationships were found between ratings and student attitude variables. The strongest relationship, (r^2 = .38) was found between composite rating and student attitude toward topics of the course. The second strongest relationship was found between composite rating and the student's attitude toward the grade expected in the course (r^2 = .26). For the non-attitude variables, the strongest relationship was found to be between composite rating and the grade the student expected to receive (r^2 = .08).

The multiple regression analysis, involving all of the contextual, student, and student attitude variables as independent variables resulted in an R^2 value of .48. Thus, comparison of R^2 (.48) to the largest pairwise r^2 (.38 between rating and student attitude toward topics of the course) shows the combined variables to be the better predictor of teacher ratings.

Teacher rating variables Y_2 (STRESSES FACTS--STRESSES CONCEPTS) and Y_5 (CONTROLS STUDENTS--FREES STUDENTS) were found not to be significantly related to any of the student attitude variables. In fact Y_2 was found to be significantly related to only one variable; namely, W_1 (the student's cumulative grade point average). In this

instance the students with the higher grade point averages tended to rate the teachers as stressing concepts more than did the students with the lower grade point averages. An interpretation of this lack of relationship for these two variables cannot overlook the possibility that students may not have been sure which was the better teacher, the one who stressed facts or the one who stressed concepts, and likewise, the one who controlled students or the one who freed students.

CHAPTER V

A MODEL

INTRODUCTION

As reported in CHAPTER IV the student ratings of teachers at Montana Tech were shown to be strongly associated with a variety of contextual, student, and student attitude variables. Since these variables were, for the most part, not related to teacher behaviors, and also, since these variables were not directly controllable by the teacher, the fairness of such ratings appears questionable, especially in such consideration as teacher salary, promotion or tenure.

Whether fair or unfair it is quite probable, however, that student ratings as well as other kinds of ratings will be used for such purposes in the future; the trend toward accountability and the lack of other adequate measures of teaching effectiveness support such usage. The approach, then, appears to be to find a method of controlling these variables in such a way that the above mentioned relationship is eliminated. This approach, because of the number of variables to be controlled, would require a much more involved and time consuming procedure for the administration of the rating forms than is presently used. This difficulty alone seems to rule out such an approach. An alternate approach involves the

alteration or transformation of the teacher ratings in such a way as to statistically eliminate the above mentioned undesirable relationship. This latter approach has in the past been used in educational and psychological research to eliminate the effect of a non-controllable variable in an experimental situation (Ferguson, 1971:387). This transformation uses a regression equation to predict the portion of an observation that is attributed to the uncontrolled variable or variables. This predicted portion of the observation is then subtracted leaving only the residual portion. The remainder of CHAPTER V reports the development and application of such a regression transformation. This transformation was designed to remove the effect of various contextual, student and student attitude variables from composite teacher ratings. Results of the subsequent application of this transformation to the fall semester, 1973, teacher ratings at Montana Tech are also reported in this chapter.

Chapter V, then, is divided according to the development of this model. The main divisions of the chapter are: INTRODUCTION, VALIDATION OF REGRESSION FORMULA, REGRESSION TRANSFORMATION MODEL, and SUMMARY. VALIDATION OF REGRESSION FORMULA is further divided into subdivisions entitled: Theoretical Background and Application of Transformation to Data of Present Study.

VALIDATION OF REGRESSION FORMULA

Theoretical Background

The principal purpose of the regression model in CHAPTER IV was not to predict teacher rating from the combination of contextual, student, and student attitude variables used as independent variables in that equation. Although prediction is the usual use of a regression equation, there was in this research, no practical reason for predicting teacher rating from these independent variables. In fact, the greater the predictive effectiveness of these contextual, student, and student attitude variables in predicting teacher ratings, the less useful and fair the teacher ratings become.

The principal purpose, then, of the regression equation of CHAPTER IV was not prediction, but instead, was to determine the strength of the relationship between teacher ratings and the various independent variables used. In CHAPTER IV an estimate of the strength of the relationship was made by computing the multiple R coefficient of correlation between the independent variables and the criterion variable. That multiple R was computed using the regression coefficients from the complete set of data and was found to be .69. From the R value of .69 the multiple coefficient of determination R^2 was found to be .48 and was interpreted to mean

that as much as 48% of the variance of the teacher ratings was attributable to the independent variables. Since these "independent" variables are variables over which the teacher has limited control, it follows that this relationship is undesirable if teacher ratings are supposed to be measures of teacher performance. This being the case, the remainder of CHAPTER V is directed toward the development of a statistical model that may be used to eliminate this undesirable relationship.

This model, constructed from the regression equation of CHAPTER IV, was designed for future use in future teacher ratings at Montana Tech. For this reason it was essential that some measure of the predictive validity of the model be established. Furthermore, since this model is dependent upon the regression equation of CHAPTER IV it was also necessary to determine some measure of the predictive validity of the regression equation. In fact, as will be shown later in this chapter, the predictive validity of the model designed to eliminate the undesirable relationship is identical to the predictive validity of the regression equation from which it was developed. Thus, even though prediction was not the principal intent of the regression equation reported in CHAPTER IV, the use of that equation in the model necessitated an estimate of its predictive validity.

An apparent estimate of the predictive effectiveness of the regression equation of CHAPTER IV lies in the multiple R and subsequent R^2 obtained in that chapter. These values of R and R^2 were, however, obtained by means of the same data used in establishing the regression equation itself. According to Mosier (1951:5) this is not the proper procedure for determining the predictive validity of a regression equation: "If the combining weights of a set of predictors have been determined from the statistics of one sample, the effectiveness of the predictor-composite must be determined on a separate, independent sample." This technique, which Mosier (1951) termed cross-validation, implies that when all the sample data is used to determine the regression coefficients for the equation, it then becomes impossible (because all the data has been used up) to determine an unbiased estimate of the predictive effectiveness of the regression equation. This difficulty was well described by Mosier (1951:9):

> In prediction problems, exemplified by cross-validation, we are faced with two distinct (and, as we shall see later, incompatable) goals. The first of these is the determination of those weights which will best predict the criterion from the predictor information; the second is the most accurate determination of how effective our prediction will be--usually in terms of a simple or multiple-correlation coefficient. As we analyse these objectives further, we shall see that they are incompatible--if we use all of our data to make the best determination of weights, we cannot arrive at an unbiased estimate of the

> effectiveness of prediction. If, on the other hand, we use some of our data to get a more stable and unbiased estimate of multiple R, we necessarily settle for less than the most stable weights which could be obtained from the data at hand.

Faced with this inherent difficulty the cross-validation design proceeds as follows: Data for the regression model is divided into two samples, sample 1 and sample 2. In the basic design the data from sample 1 is used to determine regression coefficients ${}_1b_i$ and constant ${}_1a$ for the equation:

$$Y' = {}_1a + \Sigma {}_1b_i X_i$$

This regression equation is then applied to the data of sample 2 and predicted values Y' are obtained. Since criterion values Y are present in sample 2 it is now possible to obtain an estimate of the population R. This is done by computing the zero-order product-moment correlation coefficient between the predicted Y' and the corresponding criterion values Y of sample 2.

A refinement of this basic design, termed double cross-validation by Mosier (1951), provides for a better estimate of R. In double cross-validation the data is again divided into two samples, sample 1 and sample 2. At this point, however, two regression equations are developed, one using data from sample 1 and the other using data from sample 2. Let the equation developed from sample 1 be denoted by:

$$_1Y' = {}_1a + \Sigma_1 b_i X_i \qquad (1)$$

In similar fashion let the equation developed from sample 2 be denoted by:

$$_2Y' = {}_2a + \Sigma_2 b_i X_i \qquad (2)$$

Equation (1) is then applied to the data of sample 2 to obtain a set of predicted values $_1Y'$ and equation (2) is applied to data of sample 1 to obtain predicted values $_2Y'$. If $_1Y$ and $_2Y$ are used to represent actual criterion values for sample 1 and sample 2 respectively, then multiple R is estimated by computing the zero-order product-moment correlation coefficient between the combined set of $(_2Y', {}_1Y)$ and $(_1Y', {}_2Y)$ observations.

If the primary goal of the model is to attain accurate prediction while having an accurate estimate of R is only a secondary goal, then regression equations (1) and (2) are used only to estimate R and neither of the two is used for prediction purposes. For prediction purposes a third regression equation is developed using the complete set of available data.

Validation Results

The validation technique used in this study was the double cross-validation design just discussed. In this validation the complete set of 1,396 observations used in the development of the regression model in CHAPTER IV was randomly divided into two

half-samples each of size 698. The data from half-sample one was then used to develop the regression equation:

$$_1T' = {}_1a + \sum_{i=1}^{23} {}_1b_i X_i$$

where T' denotes the predicted composite teacher rating and the X_i's denote the various contextual, student, and student attitude variables used as predictors. The coefficients were found as follows:

$$_1a = 2.420$$

$$_1b_1 = -.017$$

$$_1b_2 = -.003$$

$$_1b_3 = -.040$$

$$_1b_4 = -.197$$

$$_1b_5 = -.003$$

$$_1b_6 = .072$$

$$_1b_7 = -.172$$

$$_1b_8 = .011$$

$$_1b_9 = .010$$

$$_1b_{10} = -.014$$

$$_1b_{11} = .009$$

$$_1b_{12} = .247$$

$$_1b_{13} = .082$$

$$_1b_{14} = -.123$$

$_1b_{15} = -.019$

$_1b_{16} = .016$

$_1b_{17} = .209$

$_1b_{18} = -.033$

$_1b_{19} = .053$

$_1b_{20} = .006$

$_1b_{21} = .338$

$_1b_{22} = .093$

$_1b_{23} = -.058$

The data from half-sample 2 was used to develop the regression equation:

$$_2T' = {}_2a + \sum_{i=1}^{23} {}_2b_i X_i$$

where T' denotes the predicted composite teacher rating and the X_i's denote the various contextual, student, and student attitude variables. The coefficients were found as follows:

$_2a = 2.76$

$_2b_1 = .003$

$_2b_2 = .001$

$_2b_3 = -.025$

$_2b_4 = -.010$

$_2b_5 = .006$

$_2b_6 = .029$

$_2b_7 = -.105$

$_2b_8 = -.009$

$_2b_9 = .020$

$_2b_{10} = -.010$

$_2b_{11} = .003$

$_2b_{12} = .119$

$_2b_{13} = .020$

$_2b_{14} = .082$

$_2b_{15} = -.069$

$_2b_{16} = .015$

$_2b_{17} = .075$

$_2b_{18} = .020$

$_2b_{19} = .091$

$_2b_{20} = -.009$

$_2b_{21} = .342$

$_2b_{22} = .058$

$_2b_{23} = -.009$

Following the double cross-validation design the regression equation from half-sample 1 was then applied to the data of half-sample 2 and the regression equation from the half-sample 2 was applied to the data of the half-sample 1. As described in the previous section, these predicted values for both half-samples were then

correlated with the actual criterion values using the zero-order product-moment correlation coefficient. From these combined samples a correlation of r = .68 was obtained. In addition the r values computed for each half sample were found to be .66 and .68 respectively.

In CHAPTER IV the same null hypothesis (H_0:R = 0) was tested using the multiple R value computed from the entire set of data. In that test the computed R was found to be .69 and the hypothesis was rejected. Comparison of the multiple R computed in CHAPTER IV (.69) with the value of r (.68) computed in this chapter using the double cross-validation technique shows the two values to be comparable. Both values indicate (since the multiple coefficients of determination for both were greater than (.46) that approximately 46% of the composite rating variance was attributable to the combined independent variables of the regression equation.

REGRESSION TRANSFORMATION MODEL

Theoretical Background

Because the relationship between ratings and the various contextual, student, and student attitude variables involves variables over which the teacher has limited control and since the intent of such ratings is to measure teacher behavior, the relationship is undesirable. Thus, the remainder of this chapter reports

the design and application of a statistical model which alters the teacher ratings in such a way that the above stated relationship is eliminated.

Consider, first, the linear regression of dependent variable Y on a single independent variable X. Figure 1 shows

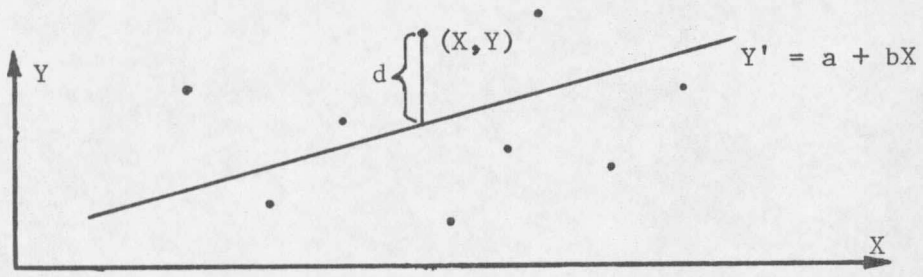

Figure 1

Regression of Y on X

the regression line $Y' = a + bX$ as determined from the set of observations $\{(X,Y)\}$. This line is constructed by choosing the constants a and b in such a way that Σd^2, the sum of the squares of distances from the points (X,Y) to the line $Y' = a + bX$ (along lines parallel to the Y axis), is minimized. The resulting line ($Y' = a + bX$) provides an estimate (Y') of Y for a given value of variable X. A common use, then, of such a regression line is to obtain predicted values of Y (Y') for given values of X. The efficiency or goodness of such predictions is often measured by

the standard error of the estimate:

$$s_{x.y} = \sqrt{\frac{\Sigma(Y - Y')^2}{n - 1}}$$

As indicated by this formula smaller values $s_{x.y}$ indicate a better fit of the regression line to the observed values and, hence, tend to result in better predictions of Y from X.

Figure 2 shows a regression line $Y' = a + bX$ for a set of observation (X,Y). In this figure a given observation (X,Y) is shown to be comprised of two components,

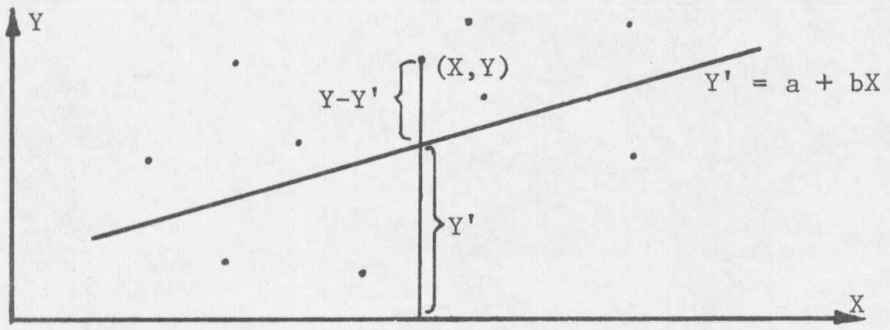

Figure 2

Components of Y

(Y - Y') which is called the residual or error of the estimate and Y' which is called the predicted value of Y. Thus, the predictive component Y' may be considered as an estimate the portion of Y accounted for by the relationship between X and Y while the residual component (Y - Y') may be attributed to other factors.

If the relationship between X and Y is an undesirable relationship, as is the case in this study, then a transformation of the observed Y variable is suggested. The conversion of each of the observations (X,Y) to a new observation (X,Y_c) by means of transformation $Y_c = Y - Y'$ results in the elimination of the predictive component of Y and retains the residual component. Using the correlation coefficient, r, or the coefficient of determination, r^2, as an estimate of the strength of the relationship between variable X and Y, it is not difficult to show that the resulting r for the transformed set of observations $\{(X,Y_c)\}$ is equal to zero (Draper and Smith, 1968). The transformation $Y_c = Y - Y'$ is called a regression transformation, and according to Ferguson (1971:387) is often used as a means of removing the effect of a non-controllable variable in an experimental situation.

This same transformation may also be applied to the data associated with a multiple linear regression equation. In this generalized case a set of observations $\{(X_1,X_2,\ldots,X_n,Y)\}$, where $X_1, X_2, \ldots, X_n$ are considered as independent variables and Y as the dependent variable, is used to establish the regression equation:

$$Y' = a + \sum_{i=1}^{n} b_i X_i$$

As in the case of one independent variable, the constants

b_1, b_2, ..., b_n and a are chosen in such a way that the sum of the squares of the quantities (Y - Y') is minimized. Thus, given the set of observation $\{(X_1,X_2,X_3,\ldots,X_n,Y)\}$ and the associated multiple regression equation:

$$Y' = a + \sum_{i=1}^{n} b_i X_i$$

the transformation $Y_c = Y - Y'$ results in a new set of observations $\{(X_1,X_2,X_3,\ldots,X_n,Y_c)\}$. If the multiple coefficient of correlation (R) or the multiple coefficient of determination (R^2) is used to estimate the strength of the relationship between a variable Y and the weighted predictors X_1, X_2, X_3, ..., X_n of Y, then it can be shown that such an R obtained from the transformed set of observations $\{(X_1,X_2,X_3,\ldots,X_n,Y_c)\}$ is equal to zero (Draper and Smith, 1968).

Application of Transformation to Data of Present Study

As discussed earlier in this paper twenty-three different contextual, student, and student attitude variables were used as dependent variables or weighted predictors of the composite teacher rating. The resulting multiple regression equation (See page 125.) with composite teacher rating, (T), as the independent variable and the various contextual, student, and student attitude variables (X_i) as dependent variables was determined from a set of 1,396 usable

observations. By using this regression equation to estimate T' (the component of the composite rating due to the weighted predictors) each of the 1,396 observations, $(X_1, X_2, X_3, \ldots, X_{23}, T)$, was adjusted or transformed by means of the transformation $(T_c = T - T')$. As shown in the previous section the multiple coefficient of correlation R between the adjusted teacher rating (T_c) and weighted predictors X_1, X_2, ..., X_{23} is necessarily equal to zero. Thus, the adjusted teacher ratings (T_c) may be considered as being independent of the combined contextual, student, and student attitude variables; and since these contextual, student, and student attitude variables are not, for the most part, controllable by the teacher, the adjusted ratings are an improvement over the unadjusted ratings.

At the end of each semester at Montana Tech the faculty members who participated in the student rating process are ranked according to the average rating they received. In order to estimate the effect of the various contextual, student, and student attitude variables on these teacher rankings two different rankings were made for the fall semester, 1973. One set of rankings was made using the averages of the unadjusted composite teacher ratings for each teacher; the second set of rankings used the averages of the adjusted or transformed ratings for each teacher (See Table 19.) Since the unadjusted ratings were found to be related to the contextual, student,

and student attitude variables while the adjusted ratings were not, an estimate of the effect of these variables on the teacher rankings was made by comparison of the two sets of ranks. More specifically, this comparison was made by means of the Spearman coefficient of rank correlation:

$$\rho = 1 - \frac{6 \Sigma d^2}{N (N^2 - 1)}$$

where N is the number of paired ranks and d is the difference between paired ranks. Using the 46 pairs of ranks (unadjusted and adjusted) shown in Table 19, the value of $\rho = .72$ was computed. Thus, although the transformation appears to have altered the teacher rankings, the alteration was not large enough to disassociate the two sets of ranks (i.e. The hypothesis that $\rho = 0$ was rejected at the $\alpha = .01$ level of significance.).

Table 19 does show that for some teachers the adjusted ratings resulted in a considerable change in rank. Teacher number 37 dropped 32 places in rank; teacher number 5 dropped 22 places in rank; teacher number 41 increased 19 places in rank; teacher number 32 increased 16 places in rank; teacher number 24 decreased 16 places in rank; teacher number 46 increased 15 places in rank; teacher number 40 increased 15 places in rank; and teacher number 34 increased 14 places in rank. In fact 16 different teachers (more than one third of the faculty) changed more than ten places

Table 19

Teacher Ranks for Unadjusted and Adjusted Ratings

Teacher Code Number	Unadjusted Rating Rank	Adjusted Rating Rank	Difference in Rank	Number of Ratings
1	43	43	0	19
2	2	4	2	65
3	14	18	4	55
4	8	7	- 1	47
5	4	26	22	3
6	30	32	2	37
7	36	41	5	46
8	25	27	2	30
9	19	29	10	12
10	6	5	- 1	20
11	15	11	- 4	13
12	17	30	13	17
13	10	10	0	66
14	24	13	-11	18
15	7	17	10	20
16	32	35	3	26
17	20	23	3	44
18	37	34	- 3	19
19	16	14	- 2	29
20	3	2	- 1	15
21	27	15	-12	30
22	45	44	- 1	42
23	11	22	11	43
24	9	25	16	17
25	42	37	- 5	66
26	44	39	- 5	49
27	23	12	-10	32
28	41	36	- 5	33
29	33	38	5	33
30	1	1	0	7
31	39	42	3	32
32	22	6	-16	30
33	26	33	7	4
34	38	24	-14	47
35	40	46	6	9
36	31	31	0	36
37	13	45	32	6
38	5	16	11	11
39	33	28	- 5	7
40	18	3	-15	11
41	28	9	-19	43
42	29	20	- 9	28
43	21	21	0	39
44	46	40	- 6	49
45	12	8	- 4	67
46	35	19	-15	24

in rank as a result of the adjusted ratings.

It is necessary to note that the two teachers (number 37 and number 5) receiving the greatest changes in rank were both rated by a small number of students. Teacher number 37 was rated by only six students and number 5 by only three students. Thus, the adjusted ratings for these two teachers were based respectively on the transformation of six and three composite ratings. Since this transformation is dependent upon the measurement of student attitudes, and since the scales used to measure these attitudes have been shown to be best used for measuring group attitudes (Shaw and Wright, 1967) there is a question of the validity of the transformation for these two teachers. By a similar argument, however, there is some question as to the validity of the unadjusted ratings of teachers receiving as few as three or six ratings.

SUMMARY

Because of the number of contextual, student, and student attitude variables found to be significantly related to teacher ratings at Montana Tech, and because many of these variables are not directly controllable by the teacher, a technique was developed for the statistical elimination of this relationship. This technique is called a regression transformation because it utilizes a regression equation to predict what portion of the teacher's

rating is attributable to the non-controllable variables. The ratings are then transformed or adjusted by subtracting out this predicted portion.

Since the regression transformation is dependent upon the predictive validity of the multiple regression equation developed in CHAPTER IV it was necessary to obtain an estimate of the validity of that equation. This was done by means of cross-validation and resulted in a "validity" coefficient comparable to the multiple R found in CHAPTER IV.

This regression equation was then used to obtain a predicted teacher rating from the data of each of 1,396 usable teacher rating forms for the fall semester, 1973, at Montana Tech. The actual teacher rating T computed for each of these forms was then transformed or adjusted by subtracting the predicted rating. In order to determine the effect of this transformation on teacher rankings two sets of ranks were compared. In one set the teachers were ranked according to the average value of the unadjusted ratings while in the other set the teachers were ranked according to the average values of all their adjusted ratings. Correlation of these pairs of ranks yielded a $\rho = .72$. Although this coefficient was not found to be small enough to conclude that the two sets of ranks were dissassociated, there were, nevertheless, substantial changes in rank for as many as one third of the faculty involved.

CHAPTER VI

CONCLUSIONS

INTRODUCTION

CHAPTER VI contains a summary of the study, conclusions of the study, and recommendations resulting from the study. The chapter is divided into the following four divisions: INTRODUCTION, SUMMARY OF STUDY, CONCLUSIONS, and RECOMMENDATIONS. SUMMARY OF STUDY is further divided into five subdivisions: The Problem, The Procedure, Literature Reviewed, Findings, and A Model. CONCLUSIONS is divided into two sections: Specific Conclusions and General Conclusions, Specific Conclusions is further divided (according to variables studied) into: Contextual variables, Student variables, Student attitude variables, and Combined variables. General Conclusions is further divided into Bias, Bias elimination, and Effect of bias on teacher rankings. RECOMMENDATIONS is also divided into two sections: Further Research and Use of Student Ratings of Teachers.

SUMMARY OF STUDY

The Problem

The research of this study was directed, primarily, toward student ratings of teachers at Montana College of Mineral Science and Technology. More specifically the study was designed and conducted

in order to determine whether or not student ratings of teachers at that college are related to certain contextual, student, and student attitude variables. Because the contextual, student, and student attitude variables included in the study are, for the most part, variables over which teachers have little or no control, it was important to know whether or not these variables are related to the teacher ratings, especially since these ratings have been used in salary and promotion considerations.

Variables included in the study were divided into three groups: contextual variables, student variables, and student attitude variables. Contextual variables included are:

1. Day of the week on which ratings were conducted,
2. The period the class met,
3. The level of the course,
4. The sex of the teacher,
5. The number of students in the class,
6. Difference between sex of teacher and sex of student.

Student variables included are:

1. The student's cumulative grade point average at the time of rating,
2. The number of credits the student was taking,
3. The number of years the student had known the teacher,
4. The number of exams the student had taken on the day

of rating (prior to rating),

5. The student's age,

6. Whether the course was required or elective for the student,

7. The sex of the student,

8. The grade the student expected to receive in the course,

9. Whether the student's major was the same or different than the teacher's major.

Student attitude variables included are:

1. Student's attitude toward the classroom,

2. Student's attitude toward the occupation of teaching.

3. Student's attitude toward the hour the class was held,

4. Student's attitude toward teacher evaluations,

5. Student's attitude toward Montana Tech,

6. Student's attitude toward topics of the course,

7. Student's attitude toward grade expected in the course,

8. Student's attitude toward the way he felt at the time of rating.

The Procedure

The procedure followed in determining the relationship between teacher ratings and each of the contextual, student, and

student attitude variables just named was a correlational one. Each of these contextual, student, and student attitude variables was correlated (Pearson product-moment) with the teacher rating. Those correlation coefficients found to be significantly different from zero at the $\alpha = .01$ level of significance resulted in the rejection of a null hypothesis $\rho(X,Y) = 0$ (ρ designates a population correlation coefficient; X designates the particular contextual, student, or student attitude variable; and Y designates the particular teacher rating variable). In addition to these pairwise correlations, a multiple regression analysis was also made using the teacher rating as the dependent variable and the twenty-three contextual, student, and student attitude variables as the independent variables. From this analysis a multiple R, coefficient of correlation, was also computed.

Data for the study were collected during the fall semester, 1973, when each student at Montana Tech was given the opportunity to rate each teacher of each class he was taking. Instruments used in the collection of these data are shown in the appendixes. PART I and PART II of the teacher rating form (See APPENDIX A.) were used to collect the student data; PART III was used to collect the student attitude data; and PART IV was used to collect the teacher rating data. Contextual variable data were collected using the instrument shown in APPENDIX C. Reliability of these

instruments was estimated by means of a test-retest application of the instruments to a cross-section of classes at Montana Tech. This was done early in the fall semester, 1973, using a period of approximately two weeks between test and retest. The resulting reliability coefficients of .90 or greater for most items of PART I and PART II, .64 or greater for most composite attitude measurements of PART III, and .77 for the composite teacher rating of PART IV were interpreted as showing that the instruments have acceptable reliability. Concurrent validity of items of PART I and PART II was estimated by comparing student responses to these items with the registrar's records. The resulting coefficients showed these items to have acceptable concurrent validity. Validity of PART III of the form was verified by reference to the works of Osgood (1957) and Shaw and Wright (1967). Validity of PART IV was verified by reference to the works of Ryans (1960) and Combs (1969).

Data collection resulted in a sample which included 46 out of 49 members of the teaching faculty and 144 out of 153 of the classes taught during the fall semester, 1973. Since the number of student responses varied from item to item on the data collection instruments, it was not possible to determine an exact student sample size. The number of student responses did tend, however, to be greater than 1,400 and less than 2,000 for any given item.

Literature Reviewed

Review of the literature related to student ratings of teachers revealed that these ratings and research on these ratings have been in evidence for over forty years. During this forty year period a substantial amount of this research has been directed toward the relationship between student ratings of teachers and many of the contextual and student variables included in this present study.

The relationship that has received the greatest amount of research is the relationship betweeen the student's grade (final course grade, expected grade, and grades in prior courses) and the student's rating of the teacher. Findings in this area seem to be divided chronologically; studies of the 1930's and 1940's by H. H. Remmers and others indicated that student grades are not related to student ratings of teachers while more recent studies have indicated that for certain classes and teachers, student ratings are related to student grades. The majority of these studies investigated the relationship between final grade received and the teacher's rating. However, recent research studies by Weaver (1960), Stewart and Malpass (1966), and Bausell and Magoon (1972) dealt with the relationship between the grade the student expected (at the time of rating) and the student's rating of the teacher. All three of these studies resulted in the conclusion

that expected grades are related to the ratings.

Other variables that have been investigated in relation to student ratings of teachers are: student grade point average the size of the class, class level, student age, student sex, required vs. elective courses, length of time student knew the teacher, social class of the student, and the hour the class was held. These variables, unlike the student grade variable, have received limited research. As was the case for the grade variable, many of the findings are contradictory.

Recent research on the relationship between student ratings of teachers and class level has indicated that this relationship is, perhaps, idiosyncratic to the particular setting in which the ratings are performed. Similar findings have appeared for the relationship between student ratings and student sex.

The divergence of the findings of the research on student ratings of teachers supports the contention that these rating relationships may vary according to the setting in which they are conducted. Implicit in this finding is the need for further research on student ratings, especially by and for those institutions currently using these ratings as assessments of teacher effectiveness.

Findings

The findings of this study, as reported in CHAPTER IV,

appear in the form of sample correlation coefficients, coefficients of determination, and multiple R coefficients of correlation. Each of the correlation coefficients resulted in the acceptance or rejection of a null hypothesis stating that a population correlation coefficient was equal to zero. The coefficients of determination were interpreted as estimates of the strength of the relationship between variables.

Thus, according to the null hypotheses rejected, student ratings of teachers at Montana Tech were found to be significantly correlated with each of the following contextual, student and student attitude variables:

1. The number of students in the class,
2. The difference between the sex of the teacher and the sex of the student,
3. The number of credits the student was taking,
4. The age of the student,
5. Whether the course was required or elective,
6. The sex of the student,
7. The grade the student expected at the time of rating,
8. Whether the student's major was the same or different than the teacher's,
9. The student's attitude toward the classroom,

10. The student's attitude toward the occupation of teaching,

11. The student's attitude toward the hour the class was held,

12. The student's attitude toward teacher evaluations,

13. The student's attitude toward Montana Tech,

14. The student's attitude toward topics of the course,

15. The student's attitude toward the grade he was expecting to receive,

16. The student's attitude toward the way he felt on the day of rating.

The strongest of these relationships, as measured by coefficients of determination (r^2), were found for the relationships involving attitude variables. The strongest relationship ($r^2 = .38$) was found to be between student ratings and student attitudes toward topics of the course.

Computation of the multiple R coefficient of correlation from a regression equation in which the twenty-three contextual, student, and student attitude variables were used as independent variables and the composite teacher rating as the dependent variable resulted in a value of $R = .69$. The multiple coefficient of determination ($R^2 = .48$) was interpreted as indicating that as much as .48 of the variance of the composite teacher rating may be

attributable to these twenty-three variables.

A Model

Because of the strength of the relationship between teacher ratings and the twenty-three contextual, student, and student attitude variables, and because these twenty-three variables are, for the most part, not controllable by the teachers, a statistical technique was developed to eliminate the relationship. This technique, known as a regression transformation, has in the past been used as a means of removing an uncontrollable variable from an experimental situation. To do this a linear regression equation is used to predict what portion of an observation is attributable to the uncontrolled variable; this predicted portion is then subtracted from the observation. In the present study the uncontrollable variables were considered to be the twenty-three contextual, student, and student attitude variables; the regression equation was the multiple regression equation used in CHAPTER IV to estimate the relationship between these variables and the teacher ratings. The portion of each teacher rating predicted from this equation was then subtracted from the rating.

In order to determine the effect of this transformation on teacher rankings, two sets of ranks were compared. In one set all the teachers were ranked according to the average value of their untransformed ratings while in the other set the teachers

were ranked according to the average of their transformed ratings. The two sets of ranks were then compared using a rank order correlation. The resulting $\rho = .72$ indicated that, although the two sets of ranks were different, they were not sufficiently different to be interpreted as being disassociated. The transformed ratings, however, did result in changes in rank of over ten places for as many as one-third of the faculty.

CONCLUSIONS

Specific Conclusions

Contextual variables. The list of research questions in CHAPTER I includes questions about the relationship between student ratings of teachers and certain contextual variables. With respect to teachers and students at Montana Tech this study provided the following answers to these questions:

1. The day of the week on which ratings are held is not related to student ratings of the teacher.

2. The period of the day the class is held is not related to student ratings of the teacher.

3. The level of the course is not related to the student ratings of the teacher.

4. The sex of the teacher is not related to the ratings

the teacher receives.

5. The number of students in the class is related to student ratings of the teacher; teachers of larger classes tend to receive lower ratings than teachers of smaller classes.

6. The difference between the sex of the teacher and the sex of the student is related to the rating the teacher receives; teachers tend to be rated higher by students of the opposite sex.

Although two of these contextual variables were found to be significantly related to student ratings of teachers, none of these contextual variables can be considered as being strongly related to the ratings. In fact the maximum amount of rating variance that may be attributed to any one of these contextual variables is 3% (to the number of students in the class).

Comparison of these conclusions to the conclusions of other researchers shows both agreement and disagreement. The conclusion that course level is not related to ratings agrees with the conclusions of Starrak (1934), Heilman and Armentrout (1936), Remmers (1939), Brookover (1940), Riley et al. (1950), and Drucker and Remmers (1951); the conclusion disagrees with the findings of Davenport (1944), Remmers and Elliot (1949), Bendig (1952), Stewart and Malpass (1966), and Bledsoe, Brown and Strickland (1971). The conclusions that class size is related to student ratings of teachers, although not strongly related, is supported by the findings

of Starrak (1934), Riley et al. (1950), and Cornwell (1974). The conclusion is not supported by Heilman and Armentrout (1936) who found class size not to be related to student ratings. The conclusion that the hour the class is held is not related to student ratings is supported by the findings of Overturf and Price (1966) and by Cornwell (1974).

Due to the limited number of supportive studies and due also to the divergent findings for the class level and class size variables, generalization is not possible for relationships between student ratings and these contextual variables. The seemingly contradictory findings are resolved, however, by the conclusion that relationships between student ratings and contextual variables are idiosyncratic to the particular institutions, teachers, and classes in which the ratings are performed.

Student variables. Included in CHAPTER I are research questions involving a variety of student variables, or more specifically, relationships between student variables and student ratings of teachers. With respect to student ratings of teachers at Montana Tech this study provided the following answers to these questions:

1. The cumulative grade point average of the student is not related to the student's rating of the teacher.

2. The number of credits taken by a student is related to the student's rating of the teacher. Students taking the larger number of credits tend to give lower ratings than students taking fewer credits.

3. The number of years the student has known the teacher is not related to the rating the student gives the teacher.

4. The number of exams the student takes on the day of rating (prior to rating) is not related to the rating the student gives the teacher.

5. The age of the student is related to the rating the student gives the teacher. Older students tend to give higher rating than do younger students.

6. Whether a course is required or elective is related to the rating the teacher receives. Teachers of elective courses tend to receive higher ratings than teachers of required courses.

7. The sex of the student is related to the rating the student gives the teacher. Female students tend to give higher ratings than do male students.

8. The grade the student expects to receive in the course is related to the rating the student gives the teacher. Students expecting higher grades tend to give higher ratings than do students expecting lower grades.

9. Whether the student's major is the same or different

than the major of the teacher is related to the rating the student gives the teacher. Students having the same major as the teacher tend to give the teacher higher ratings than do students having a different major.

Of all the contextual and student variables included in this study, the student variables--grade expected in the course--appears to be the most strongly related to student ratings. The coefficient of determination of .08 for this relationship supports the conclusion that as much as 8% of the rating variance may be attributed to this student variable. Less than 2% of the rating variance may be attributed to any of the other student variables.

The conclusions that student grade point average is not related to student ratings of teachers is supported by the findings of each of the studies by Overturf and Price (1966), Rayder (1968), and Bausell and Magoon (1972). The conclusion that the number of years the student has known the teacher is not related to the rating the student gives the teacher is supported by Davenport (1944) but is not supported by Brookover (1945). Of all studies reviewed only the one by Davenport (1944) dealt specifically with student age as a variable. The findings of Davenport (1944) are in agreement with the conclusion that student age is related to student ratings of teachers. The conclusion that teachers of elective courses tend to receive higher ratings than teachers of required

courses is in disagreement with the findings of both Heilman and Armentrout (1936) and Riley et al. (1950); neither of these studies found that teachers of required courses received different ratings than teachers of elective courses. The conclusion that female students tend to give teachers higher ratings than male students is supported by Bryan (1937) and Davenport (1944). This conclusion does, however, disagree with Brookover (1949), Remmers and Elliot (1940), and Bendig (1952); Brookover (1949) and Remmers and Elliot (1940) found no relation between student sex and ratings, while Bendig (1952) found that female students gave lower ratings than did male students. This conclusion also disagrees with Riley et al. (1950), and Bledsoe, Brown and Strickland (1971); neither of these two studies found ratings related to student sex. The conclusion that students expecting higher grades in a course tend to rate the teacher higher than do students expecting lower grades is also supported by other research. Weaver (1960), Stewart and Malpass (1966), and Bausell and Magoon (1972) all found student ratings to be related to expected grade. No other research studies were found that dealt specifically with the variable--expected grade.

As was the case for the contextual variables, generalization about relationships between student ratings of teachers and the student variables included in this study is not possible. Only two of these relationships--the relationship between student ratings

and student grade point average, and the relationship between student rating and expected grade--were supported unanimously by other research. In each case, however, there were only three supportive studies. For all the other relationships the research findings were either non-existent or contradictory. These contradictory findings, as was the case for the relationships between the contextual variables and ratings, again indicate that relationships in student ratings of teachers are idiosyncratic to the particular school, department, or class in which the ratings are held.

Student attitude variables. Also included in the research questions of CHAPTER I are questions involving student attitudes. With respect to student ratings of teachers at Montana Tech, this study provided the following answers to these questions:

1. Student attitude toward the classroom is related to the student rating of the teacher.

2. Student attitude toward the occupation of teaching is related to the student rating of the teacher.

3. Student attitude toward the hour the class is held is related to the student rating of the teacher.

4. Student attitude toward teacher evaluations is related to the student rating of the teacher.

5. Student attitude toward Montana Tech is related to the

student rating of the teacher.

6. Student attitude toward topics of the course is related to the student rating of the teacher.

7. Student attitude toward grade expected in the course is related to the student rating of the teacher.

8. Student attitude toward the way he feels at the time of rating is related to the student rating of the teacher.

In addition it may be concluded that students having the more favorable attitudes tend to rate the teacher higher than do the students with the less favorable attitudes. Furthermore, the relationships between student attitudes and student ratings of teachers are the strongest of all the relationships investigated. Student attitude toward topics in the course may account for as much as 38% of the teacher rating variance, student attitude toward the occupation of teaching may account for as much as 29% of the teacher rating variance, and student attitude toward the grade expected in the course may account for as much as 26% of the rating variance. Even though these attitude variables are not independent, they are, when combined, significantly and strongly related to student ratings of teachers.

Comparison of these conclusions to other research is impossible since little or no research has been done on the relationship between student attitudes and student ratings of teachers.

Combined variables. The combination of all twenty-three contextual, student, and student attitude variables is related to student ratings of teachers at Montana Tech. Furthermore, when used as weighted predictors these twenty-three variables account for up to 48% of the teacher rating variance.

General Conclusions

Bias. Student ratings of teachers at Montana College of Mineral Science and Technology are related to the combination of the twenty-three contextual, student, and student attitude variables included in the study. With the possible exception of the variable--the grade the student expects to receive in the course--these variables are not directly controllable by the teachers, and thus, the above relationship constitutes a bias in the rating process.

Bias elimination. This bias, since it is defined in terms of the relationship between student ratings of teachers and contextual, student, and student attitude variables, can be eliminated by elimination of the relationship.

The regression transformation model developed in this study does, within its limitations, provide for the statistical elimination of the bias relationship. However, since this model is dependent on randomization of error through large numbers of ratings and

since it is also dependent on measurement of student attitude, its use is questionable when applied to teachers receiving only a small number of student ratings.

Effect of bias on teacher rankings. If bias is defined operationally as the relationship between student ratings of teachers and variables which the teachers cannot control, then the effect of this bias on teacher ratings can be estimated by comparing the ranks of the teachers determined by transformed ratings with the ranks determined by untransformed ratings. Comparisons of these teacher rankings at Montana Tech showed the effect of the bias to be considerable for as many as one-third of the teaching faculty.

RECOMMENDATIONS

Further Research

Because of the expanded use of student ratings of teachers and because of the contradictory findings of much of the research associated with these ratings, it is apparent that further research is needed. The strength of the biases in student ratings of the teachers at Montana Tech gives clear indication that these ratings can be dangerous and unfair, especially when used in salary, promotion, and tenure considerations. Although this does not imply that the

same biases will exist in student ratings of teachers at other institutions or departments, it does say that institutions and departments using these ratings should include, as part of the rating process, research designed to detect and eliminate such biases. From this research will come added evidence as to which of the biases tend to be general and inherent in the rating process and which tend to be local or idiosyncratic to the particular setting in which the ratings are conducted.

Since many of the student attitudes are not directly controllable by the teacher and since certain of these attitudes were found in this study to be strongly related to ratings, they should be the central focus of much of the future research.

Student attitudes, unlike many of the student and contextual variables, are constructs and hence, are much more difficult to measure. Because the present study involved the measurement of so many different attitudes and because of the length limitations of the rating instrument, all of the attitudes were measured using the same set of semantic differential scales. Even though this practice has been used by other researchers, Shaw and Wright (1967) have contended that different semantic differential scales are necessary to measure different attitudes. Thus, it would be of interest to compare the findings of this study with the findings of studies that use different semantic differential scales to measure student

attitudes. Studies using Thurstone scales would provide yet another, and perhaps, better comparison.

As indicated by the various coefficients of determination, the eight attitude variables included in this study are not independent of each other, otherwise they would account for more than 100% of the rating variance. Thus, further research is necessary to determine which of these attitudes or groups of attitudes are independent of each other. This suggests for the use of factor analysis on these and additional student attitude variables.

Unanswered questions relative to the elimination of bias in student ratings also suggest future research. The bias elimination model developed in the present study utilizes a multiple regression equation to predict what portion of each teacher's rating is attributable to the biasing variables. Development of this equation was dependent upon measurement of student attitudes, and therefore, emphasizes the need for additional research on the measurement of these attitudes.

Also since the bias elimination model is used to adjust each rating that each teacher receives there is also a question as to how many ratings a teacher must receive before adjustments are made. Certainly an adjusted teacher rating based on the average ratings of only one or two students is subject to a greater possibility of a random error than is an average based on the

adjusted ratings of twenty or thirty students. Thus, it is necessary to determine how many student ratings a teacher must receive in order to apply the model. In this regard it is necessary also to determine whether or not teachers receiving a small number of ratings should even be included in the rating process. What is the minimum number of ratings needed?

Since the bias elimination model developed in this study was developed specifically for the ratings at Montana Tech, it should not, without proper modification, be used at other institutions. To do so would be to extend the validity of the model beyond its limits. Any institution could, however, use the same approach to measure the dependent variables and to develop a regression equation suitable to that particular institution. This does not say that the specific model developed in this study could not be used anywhere else. It would, in fact, be of interest to apply this model to ratings conducted in other institutions. Thus, by correlating the actual ratings at these institutions with the ratings predicted by the model, it would be possible to determine to what degree the validity of the model may be extended beyond its use at Montana Tech.

Yet another area of needed research relative to the model developed in this study is the area of validation of the model. In the present study, the model was validated by estimating the

predictive validity of the multiple regression equation used to determine the component of the ratings attributable to the dependent or biasing variables. However, since the objective of this model is to eliminate the relationship between the biasing variables and the teacher rating, another validation approach might consist of the following: the model would first be used to adjust the ratings of each teacher, then for each teacher these adjusted ratings would be correlated with the biasing variables. The model would be considered satisfactory or valid for a given teacher if the resulting correlation coefficient is not found to be significantly different from zero.

Use of Student Ratings of Teachers

Montana Tech. Because of the number of biases and because of the strength of the biases in student ratings of teachers at Montana Tech, it is recommended that these ratings not be used in future determination of teacher salary, tenure, and promotion at that school. If the ratings must be used in such considerations, then the adjusted or transformed ratings should be used rather than the unadjusted ratings. Even the adjusted ratings should be used with reservation, since it is possible that they too are biased by variables that were overlooked in the development of the model. Furthermore, the use of the model necessitates a long and involved

rating form which the students do not like to fill out. In addition both the students and the faculty do not understand such a rating form, especially that part of the form which measures student attitudes.

This is not a recommendation that student ratings of teachers be suspended at Montana Tech; it is a recommendation only that the ratings not be used in salary, promotion, and tenure decisions. If for no other reason the ratings should be conducted as a means of providing teachers with feedback on their teaching. Furthermore, the need for additional research on the rating process is, itself, a recommendation that the ratings be continued as part of the institutional research of the school.

Other schools. Although the biases found in the rating process at Montana Tech cannot be extended to other schools, the approach to student ratings can be. The number of biases and the strength of the biases in the rating process at Montana Tech give an indication that such ratings can be quite unfair when used to determine teacher,salary, promotion and tenure. Thus, a final recommendation is that any institution using student ratings in such a manner should include, as institutional research, some means of determining what biases are present and to what extent these biases influence the ratings conducted in that institution. If biases

are found, then either the ratings should not be used in salary, promotion, and tenure decisions, or the biases must be eliminated.

APPENDIX A

MONTANA COLLEGE OF MINERAL SCIENCE AND TECHNOLOGY

TEACHER RATING FORM

The information being requested of you on this rating form is being used in a research study designed to measure the fairness of student ratings of teachers. Please fill the form out carefully paying special attention to the instructions for each part. Thank you.

PART I

For each of the following five items please fill in the blank with the number that best completes the statement. If you do not know the number leave the space blank.

1. My cumulative grade point average at the beginning of this semester was: _____
2. The number of credits I am taking this semester is: _____
3. The number of years I have known this teacher is: _____
4. The number of exams I have taken today is: _____
5. My age is: _____

PART II

For each of the following five items place an X in the appropriate blank.

1. For me this course is: _____
   - Required _____
   - Elective _____
2. I am: _____
   - Male _____
   - Female _____
3. My major is: _____
   - Chemistry _____
   - Engineering _____
   - English _____
   - Geology _____
   - History _____
   - Mathematics _____
   - Other _____
4. The grade I expect in this course is: _____
   - A _____
   - B _____
   - C _____
   - D _____
   - F _____
5. My major is the same as the major of this teacher: _____
   - Yes _____
   - No _____

PART III

For each of the following eight items there are seven word or symbol opposites. These opposites are separated by seven blanks. For each set of word opposites place an X in the blank that best describes the item for you.

EXAMPLE: SCHOOL

1. GOOD ____:_X_:____:____:____:____:____ BAD

The X in this sample indicates a feeling that school is good, but not completely good.

1. THIS CLASSROOM

1. GOOD ____:____:____:____:____:____:____ BAD
2. CRUEL ____:____:____:____:____:____:____ KIND
3. CLEAN ____:____:____:____:____:____:____ DIRTY
4. REGRESSIVE ____:____:____:____:____:____:____ PROGRESSIVE
5. REPUTABLE ____:____:____:____:____:____:____ DISREPUTABLE
6. SICK ____:____:____:____:____:____:____ HEALTHY
7.  ____:____:____:____:____:____:____ 

2. OCCUPATION OF TEACHING

1. GOOD ____:____:____:____:____:____:____ BAD
2. CRUEL ____:____:____:____:____:____:____ KIND
3. CLEAN ____:____:____:____:____:____:____ DIRTY
4. REGRESSIVE ____:____:____:____:____:____:____ PROGRESSIVE
5. REPUTABLE ____:____:____:____:____:____:____ DISREPUTABLE
6. SICK ____:____:____:____:____:____:____ HEALTHY
7.  ____:____:____:____:____:____:____ 

3. THE HOUR THIS CLASS IS HELD

1. GOOD ____:____:____:____:____:____:____ BAD
2. CRUEL ____:____:____:____:____:____:____ KIND
3. CLEAN ____:____:____:____:____:____:____ DIRTY
4. REGRESSIVE ____:____:____:____:____:____:____ PROGRESSIVE
5. REPUTABLE ____:____:____:____:____:____:____ DISREPUTABLE
6. SICK ____:____:____:____:____:____:____ HEALTHY
7.  ____:____:____:____:____:____:____ 

4. TEACHER EVALUATIONS

1. GOOD ____:____:____:____:____:____:____ BAD
2. CRUEL ____:____:____:____:____:____:____ KIND
3. CLEAN ____:____:____:____:____:____:____ DIRTY
4. REGRESSIVE ____:____:____:____:____:____:____ PROGRESSIVE
5. REPUTABLE ____:____:____:____:____:____:____ DISREPUTABLE
6. SICK ____:____:____:____:____:____:____ HEALTHY
7.  ____:____:____:____:____:____:____ 

5. MONTANA TECH

1.	GOOD	____:____:____:____:____:____:____	BAD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	CLEAN	____:____:____:____:____:____:____	DIRTY
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	REPUTABLE	____:____:____:____:____:____:____	DISREPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

6. TOPICS OF THIS COURSE

1.	GOOD	____:____:____:____:____:____:____	BAD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	CLEAN	____:____:____:____:____:____:____	DIRTY
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	REPUTABLE	____:____:____:____:____:____:____	DISREPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

7. GRADE I EXPECT IN THIS COURSE

1.	GOOD	____:____:____:____:____:____:____	BAD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	CLEAN	____:____:____:____:____:____:____	DIRTY
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	REPUTABLE	____:____:____:____:____:____:____	DISREPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

8. TODAY I FEEL

1.	GOOD	____:____:____:____:____:____:____	BAD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	CLEAN	____:____:____:____:____:____:____	DIRTY
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	REPUTABLE	____:____:____:____:____:____:____	DISREPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

PART IV

Rate your teacher on each of the following item scales by placing an X in the blank that best describes the item for you.

1.	EXCELLENT TEACHER	____:____:____:____:____:____:____	POOR TEACHER
2.	STRESSES FACTS	____:____:____:____:____:____:____	STRESSES CONCEPTS
3.	STUDENT CENTERED	____:____:____:____:____:____:____	SUBJECT CENTERED
4.	UNINVOLVED	____:____:____:____:____:____:____	INVOLVED
5.	CONTROLS STUDENTS	____:____:____:____:____:____:____	FREES STUDENTS
6.	FRIENDLY	____:____:____:____:____:____:____	UNFRIENDLY
7.	HINDERING	____:____:____:____:____:____:____	HELPFUL
8.	ALOOF	____:____:____:____:____:____:____	RESPONSIVE
9.	STIMULATING	____:____:____:____:____:____:____	DULL
10.	UNCERTAIN	____:____:____:____:____:____:____	CONFIDENT
11.	SYSTEMATIC	____:____:____:____:____:____:____	DISORGANIZED
12.	INFLEXIBLE	____:____:____:____:____:____:____	ADAPTABLE
13.	BROAD	____:____:____:____:____:____:____	NARROW
14.	PARTIAL GRADER	____:____:____:____:____:____:____	FAIR GRADER
15.	NEAT	____:____:____:____:____:____:____	SLOPPY
16.	CONFUSED OBJECTIVES	____:____:____:____:____:____:____	CLEAR OBJECTIVES
17.	PREPARED	____:____:____:____:____:____:____	UNPREPARED
18.	BAD COMMUNICATOR	____:____:____:____:____:____:____	GOOD COMMUNICATOR
19.	I LEARNED MUCH	____:____:____:____:____:____:____	I LEARNED LITTLE
20.		____:____:____:____:____:____:____	

APPENDIX B

REORIENTED SCALES

PART III

1. THIS CLASSROOM

1.	BAD	_____:_____:_____:_____:_____:_____:_____	GOOD
2.	CRUEL	_____:_____:_____:_____:_____:_____:_____	KIND
3.	DIRTY	_____:_____:_____:_____:_____:_____:_____	CLEAN
4.	REGRESSIVE	_____:_____:_____:_____:_____:_____:_____	PROGRESSIVE
5.	DISREPUTABLE	_____:_____:_____:_____:_____:_____:_____	REPUTABLE
6.	SICK	_____:_____:_____:_____:_____:_____:_____	HEALTHY
7.		_____:_____:_____:_____:_____:_____:_____	

2. OCCUPATION OF TEACHING

1.	BAD	_____:_____:_____:_____:_____:_____:_____	GOOD
2.	CRUEL	_____:_____:_____:_____:_____:_____:_____	KIND
3.	DIRTY	_____:_____:_____:_____:_____:_____:_____	CLEAN
4.	REGRESSIVE	_____:_____:_____:_____:_____:_____:_____	PROGRESSIVE
5.	DISREPUTABLE	_____:_____:_____:_____:_____:_____:_____	REPUTABLE
6.	SICK	_____:_____:_____:_____:_____:_____:_____	HEALTHY
7.		_____:_____:_____:_____:_____:_____:_____	

3. THE HOUR THIS CLASS IS HELD

1.	BAD	_____:_____:_____:_____:_____:_____:_____	GOOD
2.	CRUEL	_____:_____:_____:_____:_____:_____:_____	KIND
3.	DIRTY	_____:_____:_____:_____:_____:_____:_____	CLEAN
4.	REGRESSIVE	_____:_____:_____:_____:_____:_____:_____	PROGRESSIVE
5.	DISREPUTABLE	_____:_____:_____:_____:_____:_____:_____	REPUTABLE
6.	SICK	_____:_____:_____:_____:_____:_____:_____	HEALTHY
7.		_____:_____:_____:_____:_____:_____:_____	

4. TEACHER EVALUATIONS

1.	BAD	____:____:____:____:____:____:____	GOOD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	DIRTY	____:____:____:____:____:____:____	CLEAN
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	DISREPUTABLE	____:____:____:____:____:____:____	REPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

5. MONTANA TECH

1.	BAD	____:____:____:____:____:____:____	GOOD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	DIRTY	____:____:____:____:____:____:____	CLEAN
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	DISREPUTABLE	____:____:____:____:____:____:____	REPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

6. TOPICS OF THIS COURSE

1.	BAD	____:____:____:____:____:____:____	GOOD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	DIRTY	____:____:____:____:____:____:____	CLEAN
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	DISREPUTABLE	____:____:____:____:____:____:____	REPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
7.		____:____:____:____:____:____:____	

7. GRADE I EXPECT IN THIS COURSE

1.	BAD	____:____:____:____:____:____:____	GOOD
2.	CRUEL	____:____:____:____:____:____:____	KIND
3.	DIRTY	____:____:____:____:____:____:____	CLEAN
4.	REGRESSIVE	____:____:____:____:____:____:____	PROGRESSIVE
5.	DISREPUTABLE	____:____:____:____:____:____:____	REPUTABLE
6.	SICK	____:____:____:____:____:____:____	HEALTHY
			

8. TODAY I FEEL

1.	BAD	____	____	____	____	____	____	GOOD
2.	CRUEL	____	____	____	____	____	____	KIND
3.	DIRTY	____	____	____	____	____	____	CLEAN
4.	REGRESSIVE	____	____	____	____	____	____	PROGRESSIVE
5.	DISREPUTABLE	____	____	____	____	____	____	REPUTABLE
6.	SICK	____	____	____	____	____	____	HEALTHY
7.		____	____	____	____	____	____