



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

© Copyright by Dennis Richard Haley (1974)

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.

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. King
Chairman, Examining Committee

Robert J. Thifault
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

	Page
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER	
I. INTRODUCTION	1
SETTING	1
FORMAT	5
STATEMENT OF PROBLEM	6
NEED OR PURPOSE OF THE STUDY	8
GENERAL QUESTIONS TO BE ANSWERED	11
GENERAL PROCEDURE	13
LIMITATIONS	15
DEFINITION OF TERMS	16
SUMMARY	17
II. REVIEW OF RELATED RESEARCH	19
INTRODUCTION	19
RESEARCH FINDINGS	26
Student's Grade in Course	26
Student's Grade Point Average	36
Class Size	37
Class Level and Student's Age	39

CHAPTER	Page
Student's Sex	42
Required vs. Elective	44
Student's Attitudes	45
Multiple Relationships	47
Other Variables	48
ANALYSIS OF RESEARCH	49
Representativeness of Samples	51
Rating Instruments Used	52
Time of Study	53
SUMMARY	54
III. PROCEDURES	56
INTRODUCTION	56
POPULATION DESCRIPTION AND SAMPLING PROCEDURE	57
AREA OF INVESTIGATION	59
DATA COLLECTION AND VALIDATION	60
Data Collection Instruments and Procedure	60
Reliability	67
Validity	72
ORGANIZATION OF DATA	80
ANALYSIS OF DATA	81
PRECAUTIONS TAKEN FOR ACCURACY	86

CHAPTER	Page
SUMMARY	87
IV. FINDINGS	89
INTRODUCTION	89
FINDINGS	91
Contextual Variables	91
Non-Attitudinal Student Variables	95
Student Attitude Toward the Classroom	101
Student Attitude Toward Occupation of Teaching	104
Student Attitude Toward the Hour the Class is Held	106
Student Attitude Toward Teacher Evaluations	109
Student Attitude Toward Montana Tech	112
Student Attitude Toward Topics of the Course	114
Student Attitude Toward Grade Expected in Course	117
Student Attitude Toward the Way He Feels	120
Multiple Regression Findings	123
SUMMARY	126
V. A MODEL	129
INTRODUCTION	129
VALIDATION OF REGRESSION FORMULA	132
Theoretical Background	132

CHAPTER	Page
Validation Results	136
REGRESSION TRANSFORMATION MODEL	140
Theoretical Background	140
Application of Transformation to Data of Present Study	144
SUMMARY	148
VI. CONCLUSIONS	150
INTRODUCTION	150
SUMMARY OF STUDY	150
The Problem	150
The Procedure	152
Literature Reviewed	155
Findings	156
A Model	159
CONCLUSIONS	160
Specific Conclusions	160
General Conclusions	168
RECOMMENDATIONS	169
Further Research	169
Use of Student Ratings of Teachers	173
APPENDIX A	176
APPENDIX B	180

	Page
APPENDIX C	184
APPENDIX D	185
BIBLIOGRAPHY	186

LIST OF TABLES

Table	Page
1. Research Related to Student Ratings of Teachers . . .	22
2. Reliability Coefficients for Items on PART I of Rating Instrument	68
3. Reliability Coefficients for Items on PART II of Rating Instrument	69
4. Reliability Coefficients for Variables on PART III of Rating Instrument	70
5. Reliability Coefficients for Teacher Rating Variables on PART IV of Rating Instrument	73
6. Validity Coefficients for Items on PART I and PART II of Rating Instrument	75
7. Variable Names for Various Items on Rating Form and Rating Packet	83
8. Hypotheses Tested	84
9. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Certain Contextual Variables	92
10. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Certain Student Variables	96
11. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Classroom	103
12. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Occupation of Teaching	105
13. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Hour the Class was Held	108

Table	Page
14. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Teacher Evaluations	111
15. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Montana Tech	113
16. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Topics of the Course	116
17. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward Grade Expected in the Course	119
18. Correlation Coefficients for Hypotheses Involving Teacher Rating Variables and Student Attitude Toward the Way He Feels	122
19. Teacher Ranks for Unadjusted and Adjusted Ratings .	147

LIST OF FIGURES

Figure	Page
1. Regression of Y on X	141
2. Components of Y	142

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 its 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 categories: 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; Discrepancy 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 Livery 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 undoubtedly 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 idiosyncrasy 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

