May 18, 1976

To the Graduate Council:

I am submitting herewith a dissertation written by Clare Potter Counts, entitled "Student Ratings of Biology Instructors' Performance at the College Level." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Curriculum and Instruction.

A. Paul Wishart

A. Paul Wishart, Major Professor

We have read this dissertation
and recommend its acceptance:

C. Edward Roeske

Wilbur S. Lawson

Benjamin Hochman

Accepted for the Council:

Hilton A. Smith

Vice Chancellor
Graduate Studies and Research

STUDENT RATINGS OF BIOLOGY INSTRUCTORS' PERFORMANCE AT THE COLLEGE LEVEL

A Dissertation

Presented for the

Doctor of Education

Degree

The University of Tennessee, Knoxville

Clare Potter Counts

June 1976

TO

Roberta, William, Diane, Herbert,

Henry, and Eric

who, on the basis of random sampling,

are highly improbable

ACKNOWLEDGMENTS

Strictly speaking, an author's indebtedness knows no beginning and possibly no end, and so an arbitrary point of departure must be taken. I begin therefore with my teachers—too many to be listed—to whom I acknowledge my obligations and hope they will be pleased with what they find here.

My second thanks are to the teaching staff of the Biology Department and various members of the administrative staff at East Tennessee State University who trusted me with the information necessary to launch this project.

I am grateful for the counsel and autonomy given me by my friend and chairman, Dr. Paul Wishart—where there was never criticism without advice; where there was always that extraordinary feeling of good will which is constantly present in him and of which, on this occasion, he gave with particular abundance.

To the other members of my supervisory committee, especially Dr. Benjamin Hochman, I owe a great debt for their encouragement of my work.

My heaviest obligation, however, is to Mr. George Killough for helping with the statistical design and offering many incisive criticisms and imaginative suggestions for improving upon my proposed project. It was a privilege, in the full sense of the word, to share with him the excitement of my venture and to find his encouragement unfailing.

I wish to thank Mr. Ron Gross and Mr. John Boling, computer consultants at the University of Tennessee, for their invaluable assistance in writing the computer programs and interpreting the printouts and whose expertise can only be described as superior.

Thanks are due to Wayne Wilcox who expressed his understanding and affection for this research by doing a wonderful job coding and recording the data for computer analysis and to Debbie Davis for typing the manuscript.

I also express my thanks to my colleague and confidante, Roberta Bashor, who went over the final manuscript with a thoroughness that is so characteristic of everything she does and offered many constructive suggestions.

My children, Scott, Marc, and Kim enthusiastically supported this venture, and they have my thanks for that. But my true gratitude extends to their forbearance—a point which requires no elaboration.

Above all, I wish to thank my husband, Perry, who not only relinquished, without complaint, many hours of time with me which were rightfully his but who also did more than his fair share to create an atmosphere within which the work could proceed with a bare miminum of discord or distraction. To him I offer my deepest love and affection.

And last, to my mother and brother, Bob, who always let me know that they were there, I express my enduring love and appreciation.

ABSTRACT

The primary purpose of the study was to investigate the effects of three student characteristics on student ratings of instruction in biology courses at the college level. The variables were the students' theoretical orientation, attitude toward biology, background in high school chemistry courses and the course instructor. Other purposes of the study were: (1) to learn more about the extent to which students report accurately such information as their cumulative grade point average, sex, and class level; (2) to determine relationships between students' ratings of instruction and their reactions to class size, class work load, and difficulty of the course; (3) to examine the relationship between a students' perceived progress (expected grade) and his actual progress (assigned grade); (4) to attempt a validity study of the student instructional report (SIR) by analyzing the relationships between the student's ratings on 20 specific performance items and the students' global (general) rating of instructional performance, and to determine the relationship between the students' achievement and their rating of instruction; (5) to determine the influence of students' attitude toward biology on their achievement in a biology course and on their perception of the difficulty of the course; and (6) to determine the relationship of the students' theoretical orientation and the way students' perceived the course difficulty.

The basic population for the study was comprised of 374 students enrolled in the third quarter of Freshman Biology classes at East Tennessee State University in the Spring Quarter 1975.

The results of the study indicated that: (1) students with a more positive attitude toward biology gave significantly higher ratings than those with a less positive (i.e., negative) attitude toward biology; (2) students with a high theoretical orientation gave significantly higher ratings than those with a low theoretical orientation; (3) students who included chemistry in their high school program gave similar ratings to those who did not include chemistry in their high school program; (4) students differentiated among their biology instructors' performances; (5) students reported with a high degree of accuracy such information as their GPA, sex, and class level; (6) expected grades of students had virtually no influence on ratings of instructional performance; (7) the grades students expected corresponded closely to the grades assigned by the instructor; (8) there was a weak relationship between students' achievement in a biology course and the ratings of their biology instructors' performance on 20 specific performance items and the students' general ratings of their biology instructors' performance; (10) the reactions of students to class size, work load in the class, and course difficulty influenced the ratings of their biology instructors' performance. Students who thought class size was too large gave the highest ratings; those who said class size "didn't matter" gave the lowest ratings. Students who perceived the work load as much heavier than in other courses gave much lower ratings than did students who perceived the work load as being "about right." Students who perceived the work load as much lighter than in other courses gave the

highest ratings. Students who thought the difficulty of the course was very elementary or very difficult gave significantly lower ratings than those who perceived the course difficulty as "about right." Students who had a positive attitude toward biology achieved on a higher level than those who had a negative attitude toward biology. Also, students with a positive attitude perceived the course as less difficult than did students who had a negative attitude. Students with a high theoretical orientation thought the course less difficult than students who had a low theoretical orientation.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The recent interest in formal student ratings of college teaching might lead one to believe that it is a new phenomenon. It is not. Students rated teachers at some American universities 50 years ago, and 500 years ago student committees at the University of Bologna helped enforce strict rules on how professors conducted their classes (Centra 1972). Today, the sheer volume of published research and opinions on evaluation of instruction attests to the considerable attention this topic has received.

Interest in student ratings has been greatly accelerated since the middle of the 1960's and its momentum has come from a variety of sources. Curriculum developers seek student opinions to produce information for instructors—that is, to generate feedback that an instructor can use to improve his course or his teaching methods. Student groups on many campuses gather information on course and instructor "evaluations" to help students select courses to take. Some institutions allow or in some instances require instructors to submit ratings by students as evidence for tenure and promotion consideration (Centra 1972; Hoyt 1974; Cameron 1975).

Calls for increased use of student input, for a variety of purposes, may be found in the recent AAUP Statement of Teaching Evaluation (1974). In fact, a group of chemists at the University of Wisconsin agreed that student evaluation is apparently the easiest method of

evaluation to use relative to the quantity of data obtained. There seems to be considerable agreement among chemists in education that teaching can be evaluated with the expenditure of reasonable time, money and effort and that student evaluations are presently being employed for this purpose more frequently than any other method (Siebring and Schaff 1974). Regardless of the purpose for which student evaluations are conducted there is little debate that students' perceptions of their learning experiences are of general value (Silberman 1974).

The use of formal student ratings over the years has triggered an avalanche of research on their reliability, validity, and utility. While there are still some unanswered questions, it seems clear that student ratings have been found generally to be reliable (Remmers 1960; Centra 1972; Frey 1974). The validity issue is more difficult to analyze than the reliability issue. Validity studies which have been conducted over the past several decades have followed designs in which control procedures have been poor and the dependent variable inappropriate (Frey 1974). Attempts to answer the validity question are hampered by the absence of acceptable criterion measures of the impact of a teacher on student learning (Crittenden and Norr 1975). Although more recent studies involving greater technical sophistication can be interpreted with greater confidence (Frey 1974) no clear pattern has emerged from the many studies relating student ratings to student characteristics. Comments on the utility of student ratings have appeared with greater frequency in literature. Scott (1975)

alluded to the growing realization that students are in the most advantageous position to observe actual in-class activities and that their collective input should, therefore, be used as a legitimate source of information when comprehensively assessing the teaching function. Because student opinions may be obtained with greater ease and a slight expenditure of time and effort, student evaluations of teaching are increasingly being proposed as a solution to the current lack of systematic information. Trent and Cohen (1973) observed that the use of student ratings of teachers has steadily increased over the past several years despite strong skepticism about their validity and serious questioning about the use to which they are put.

Zelby (1974) admits it would be preposterous to suggest--much less advocate--that student-faculty evaluations be postponed until all factors that may affect their results can be included in such evaluation; however, Tetenbaum (1975) makes the recommendation to halt the use of student ratings of faculty until these ratings are better understood. Crittenden (1975), however, contends that enlightened use of these ratings can come about only through extensive research that takes into account the inevitable complexity of the teaching situation; and that such research can be conducted as a by-product of the use of these ratings in real classrooms.

I. PURPOSES OF THE STUDY

Major purposes of this study were:

1. To investigate the effects of three student characteristics

on student ratings of teaching in biology courses at the college level. Variables were the students' theoretical orientation, attitude toward biology, background in chemistry courses in high school, and the course instructor.

2. To learn the extent to which students report accurately such information as sex, grade point average (GPA), and class level.
3. To determine the correspondence of students' expected grades to the assigned grade and to investigate the effects of students' perceived progress on their ratings of instruction.
4. To determine relationships between students' ratings of instruction and student reaction to class size, course work load, and difficulty of the course.
5. To examine the relationships between students' attitudes toward biology and their achievement in a biology course; the students' theoretical orientations and their perception of the difficulty of the course; and students' attitudes toward biology and their perception of the course.
6. To collect data which would permit a validity study of the rating instrument (SIR) used in this study.

II. HYPOTHESES

The following eighteen hypotheses were tested:

1. A student's rating of his biology instructor's performance is not influenced by the student's attitude toward biology.
2. A student's rating of his biology instructor's performance is not influenced by his level of cognitive development, i.e., his theoretical orientation.
3. A student's rating of his biology instructor's performance is not influenced by the presence or absence of chemistry courses in his high school program.
4. There is no difference in the way students rate different instructors in a multisection biology course.
5. The correlation coefficient obtained between a student's reported GPA and actual GPA is not significant.
6. The correlation coeficient obtained between a student's reported sex and actual sex is not significant.
7. The correlation coefficient obtained between a student's reported class level and actual class level is not significant.
8. The correlation coefficient obtained between a student's expected grade and the grade assigned is not significant.
9. The correlation coefficient obtained between a student's expected grade and the student's rating of his instructor's performance is not significant.

10. The correlation coefficient obtained between a student's achievement in a biology course and his rating of instructor performance is not significant.
11. The correlation coefficient obtained between a student's instructor score and the instructor's global score assigned by the student is not significant.
12. The correlation coefficient obtained between a student's attitude toward biology and his achievement in a biology class is not significant.
13. There is no relationship between a student's reaction to class size and the student's rating of his biology instructor's performance.
14. There is no relationship between a student's perception of the course work load and the student's rating of his biology instructor's performance.
15. There is no relationship between a student's perception of the difficulty of the course and the student's rating of his biology instructor's performance.
16. There is no difference in the achievement of students with a positive attitude toward biology and students with a negative attitude toward biology.
17. A positive or negative attitude toward biology has no effect on the way a student perceives the difficulty of a course.
18. A high or low theoretical orientation has no effect on the way a student perceived the difficulty of a course.

III. RATIONALE FOR THE STUDY

Part of the basis of this investigation was drawn from recent studies by Griffiths (1973) and McKinnon (1971) in attempts to apply Piagetian theory to college students.*

It is generally agreed that freshman science courses require thinking on a formal level. Since Piaget placed an upper limit for this level at fifteen years of age, there would appear to be no problem.

Recent studies by McKinnon (1971), Schwebel (1972), and Griffiths (1973) have found that the majority of beginning college students are below the formal level in mental development. The implication of these studies is that there exists a mismatch between the mental level of college students and the content and teaching methods of college science courses which are geared toward formal level students. Could the inability of a student to grasp material on the formal level influence his perception of the course or the performance of the instructor? This investigation sought to determine whether the theoretical orientation of

---

*Piaget believes that the process of intellectual development results from interaction between an organism and its environment. Using mathematical and logical models, Piaget has outlined four stages that evolve as a result of the development of the mind. Each stage is characterized by the ability to perform certain mental functions and the inability to perform others. We are only interested here in the last two stages. The concrete operational student, that is, one in the third mental stage, is able to make logical decisions, but can only apply his logic to problems involving concrete (real or observable) situations. The main difference is that the concrete operational student cannot deal with hypothetical problems or those involving the future; he can see what "is," not what "could be." The student at the formal level can deal with all kinds of problems, present, past, future, or hypothetical, (Kolidiy 1975).

a student, i.e., his level of cognitive development, influenced the rating of his instructor in a college biology class.

A second basis for this study came from studies which considered the effects of a student's attitude toward biology on his level of performance in the course.* Since an important consequence of instruction is the student's attitude toward the subject, it is very important for teachers to be aware of the attitudinal characteristics of their students (Russell and Hollander 1975). As Mager (1968) observed:

> The likelihood of the student putting his knowledge to use is influenced by his attitude for or against the subject. Things disliked have a way of being forgotten. . . . One objective toward which to strive is that of having the student leave your influence with as favorable an attitude toward your subject as possible. In this way, you will help to maximize the possibility that he will remember what he has been taught and will willingly learn more about what he has been taught.

Using a highly reliable biology attitude scale, Russell and Hollander (1975) conducted a study which yielded information about students' attitudes toward biology in college courses. Three of the courses included in the study were: an introductory course for biology majors, a course for life sciences majors, and a course for nonbiology

---

*In the past few years, a number of writers have lamented the problem of a growing antiscience attitude by students and the public. Some writers have even speculated that a youth revolution against science was evidenced in the mid-1960's uprisings on university campuses. An American Institute of Physics survey of attitudes of high school students indicated no significant antiscience attitude in its sample (American Institute of Physics, 1972). Balvin and R. R. Boedeker (1973) in a study of nonscience student attitudes toward science revealed that there is no antiscience attitude among these students and expressed surprise that the physical science group indicated by a three-to-one majority that science courses should be required of all students.

majors. It is of interest that the students enrolled in the less rigorous courses left the course with a more positive attitude toward biology than students in the more rigorous courses for biology majors and life sciences majors.

It would appear that many college students have not developed the theoretical orientation necessary to produce a more positive attitude toward specific science courses which deal mostly with scientific concepts requiring formal thinking. Could the inability of a student to operate on the level of thinking required in most science courses lead to a negative attitude about the course and the instructor? This research effort sought to determine whether a student's attitude toward biology constitutes a biasing factor in the rating of his instructor's performance.

A third basis for this investigation involved studies which relate a student's high school chemistry background with instructional performance in beginning college biology classes. A discussion of the rationale for inclusion of this variable appears below.

During the early 1960's advances in molecular biology quickly changed the focus of the life sciences from the more traditional taxonomic approach to one which emphasized the chemistry of living systems. This new molecular approach became the focal point of most college biology courses and many high school biology programs. Thus a knowledge of chemistry became a necessary prerequisite or corequisite for mastering biological concepts.

In a new study by Bennett (1975) the effects of previous high school programs on achievement in college biology were examined. Bennett found that the influence of a high school chemistry background seemed to be a contributing factor to achievement in college biology. In fact, the relationship between high school chemistry background and achievement in college biology was highly significant at the 1 percent level. Bennett also stated that high school rank is generally thought to be the best predictor of overall success in college. However, the results of his study revealed that other indicators such as mathematics and English college entrance test scores better predicted success in college biology courses. For this reason ACT math and English scores were selected as covariants in this study.

IV. DEFINITIONS OF TERMS USED

1. Achievement. Performance gains measured by total examination points residualized for ability.

2. Global rating. Overall (general) rating of the instructor's performance assigned by students on the Student Instructional Report (SIR).

3. Teacher effectiveness. Measurement of the incremental knowledge gained by the students as a consequence of the contact with a particular instructor.

4. Teacher performance. Behaviors of the instructor measured in some way (e.g., by observations made by administrators, peers, or students).

5. Theoretical orientation. A student's level of cognitive development as measured by the theoretical orientation scale of the Omnibus Personality Inventory (OPI).

6. Instructor's score. Student rating obtained from responses to 20 specific items on the Student Instructional Report (SIR).

V. PROCEDURE AND INSTRUMENTATION

The basic population for the study was comprised of students enrolled in the third quarter of freshman biology classes in the Spring Quarter 1975, at East Tennessee State University. During the final week of classes, the Student Instructional Report (SIR), the Omnibus Personality Inventory (OPI), and a Biology Attitude Scale were administered to the subjects. ACT scores for the subjects and each student's record of high school chemistry courses taken were available from the records office. The student's report of their sex, cumulative grade point average, and expected grade in the course were obtained from answers to questions on the SIR. Total number of exam points each student accumulated for the course and the grade for the course for each student were obtained from the coordinator of the freshman program, with permission from the instructors. A total of 374 students were available for the study.

The subjects were divided into two groups based on their scores on a biology attitude scale. Subjects who scored at or above the 84th percentile were labeled high scorers and described as having a positive attitude. Subjects scoring below the 84th percentile were designated

low scorers and described as having a less positive, i.e., a negative attitude toward biology. All subjects whose standard score was at or above 50 were labeled high scorers on the theoretical orientation scale of the Omnibus Personality Inventory. Subjects whose standard score was below 50 were named low scorers on the scales.* The students were divided into two groups on the basis of chemistry courses taken in high school. One group consisted of students who had included chemistry in their high school program, and the other had not. Finally, the group was divided into three groups identified by the instructor assigned to their section. This completed the $2 \times 2 \times 2 \times 3$ factorial designed for the analysis of variance.

Relationships between a student's reported GPA and actual GPA, reported sex and actual sex, reported class level and actual class level, expected grade and actual grade, and reported grade point average and actual grade point average were determined by correlational analyses and chi square analyses. Relationships between expected grade and instructor score, total exam points (achievement) and instructor score, and global rating and instructor score were determined by correlational analyses.

Relationships between attitude toward biology and achievement, student reported difficulty of the course and student theoretical

---

*The point at which any score may be defined as a high score is relative. On most scales standard scores of 60 (84th percentile) or above are interpreted as sufficiently high for the essence of the respective definition to apply (OPI Manual, 1968).

orientation, and student reported difficulty of course and attitude toward biology were determined by correlational analyses and t-tests.

Correlation coefficients and one-way analyses of variance were calculated to determine relationships between the instructor score and student reaction to class size, instructor score and student reaction to course work load, and instructor score and student perception of difficulty of the course.

Selection of the Rating Instrument

A student rating form developed by Centra (1972) focuses on specific performance traits covering only those aspects of teaching that students are competent to judge. It is based on the proposition developed by Menne (1971) that is necessary, and at least for the present, it is sufficient to measure a teacher's performance. This proposition is analogous to the situation in which aspirin "A" cannot be proved to be more effective than aspirin "B." In such a case, other characteristics such as taste, price, availability, etc., are the determining choice factors—and these factors should be considered. Menne supports his reasoning as follows:

> Many will agree that the main intent of a teacher evaluation procedure is to measure the teacher's effectiveness, even though the usual or more common procedures involves the measurement of performance. Thus, the following Proposition A has been implicity accepted by many as true:
>
> Proposition A: If the teacher is performing well, effectiveness follows:
>
> This proposition is not demonstrably true. Research has been carried out, but the amount of research and the consistency of the findings do not lead to the conclusion that the proposition is always true. In fact, at present,

it may be questioned whether it is possible to prove that effectiveness either does or does not follow performance. Let us state this as Proposition B:

Proposition B: In many, if not most situations differences in effectiveness of two or more teachers cannot be measured reliably (accurately).

Analysis of final grades for students in a course shows that much of the variance in class performance is explained by student differences before they entered the class. This is similar to the situation found when only the OUTPUT of a college, school, or special program is evaluated. It has been found repeatedly that if you want the OUTPUT of an institution or program to look better , selection of better students for INPUT will change the OUTPUT much more than changes in program. So the teacher who is being evaluated for effectiveness by an OUTPUT measure will find that selecting better students and/or eliminating poor students will make him look more "effective."

But if you are concerned with teacher effectiveness in a strict sense, then you are concerned with the difference, "OUTPUT minus INPUT," and "effectiveness" in this sense is both generally quite small and difficult to measure. The reason for this is that most of the output is explained by input. Consider that if final grades from the proceding course in a sequence of courses or from a similar course are correlated with present "OUTPUT" or course grades, the correlation will typically be at least 0.70 and very often in the 0.80 to 0.90 range. (When it gets up to about 0.90 it can be suspected that the teacher may have been more influenced by the performance in previous courses rather than by performance in the current course!)

A correlation of 0.70 indicated that about 50 per cent of the variance in the final scores can be explained by the situations that occurred before the particular teacher had a chance to influence the student. Thus, for example, there may be a 0.70 correlation between the previous cummulative grade record and the final grade in a current course. Or there may be a correlation of 0.80 between first and second semester English grades. In this latter case 64 per cent of the second semester grade variance is explained or accounted for by the first English course grade and so there is only 36 per cent of the variance left to be explained by current course grade unreliability, teacher performance, student effort, student-teacher interaction, and possibly other factors. The reliability of the scores (final grades) is generally about 0.90 or less. This means that 10 per cent or more of the variance of the scores being analyzed is measurement error.

> Thus, there is at most only about 40 per cent and often only 25 per cent or 30 per cent of the variance which remains to be explained by factors such as:
> (a) teacher characteristics, effort, performance;
> (b) student motivation, effort, etc. unique to this class
> (c) teacher x student interaction, etc.
> So far it is not known how to partial these last three sources of variance empirically, but is seems evident that the proportion of the variance due to teacher influence is a fairly small proportion--perhaps 20 per cent, 10 per cent or less. This small proportion of variance leads to the practical impossibility of measuring the difference in effectiveness (OUTPUT-INPUT) between teachers so that it would be fair and accurate to say that one teacher is better or worse than another.

The Centra rating form selected for this study considers the student as an information source rather than an evaluator. The rating items are worded in terms of observables and statements of opinion and attitudes are avoided when possible. In addition, the selected rating form contains a section which elicits certain student characteristics and responses to the course and instructor necessary to this investigation.

It should be noted that there are many factors or aspects of a teacher's performance. If performance is rated as a global construct, it is to be expected that some raters will think of factors such as clarity and stimulation value of material presentation, others will think of the teacher's personality, mode of interaction with students or competence in the content area. Thus, it is necessary to be concerned about and delimit the rather specific aspects of performance being evaluated in order to measure something when using raters (Menn 1972).

Research has shown that certain types of items used to evaluate teaching performance seldom lead to agreement in the rating group. In

very broad terms, items that relate teacher performance to educational philosophy or concern behaviors that seem to indicate personality variables do not measure in a way that meets the condition of agreement in rating group. Even though educators may rightfully feel that such factors are important, such items can be misleading and usually cannot be performance measures. Items that concern the preparedness, effort, interest, fairness, concern, respect, and organization demonstrated by the teacher generally—but not always—do measure in such a way as to meet the condition of agreement in the rating group (Menne, 1972).

The design for the validity study was suggested by Frey (1974) who argued that validity studies of student rating instruments can be accomplished only if "Teaching" is defined in a relatively narrow way, such as preparing students to solve calculus problems on a written examination. Frey contended that only when we use such an operational definition is it feasible to determine if some teachers do a more adequate job than others. According to Frey, the ideal design is one in which students are randomly assigned to different instructors who each cover the same subject matter. Each class uses the same textbook and course syllabus and common exams are used to determine how much each student has learned. The final exam score for each instructor's students provides the external criterion indicating which teachers are most effective. Student ratings of the different instructors are then validated against this criterion. The research in this study deviates from the ideal design outlined by Frey in that students chose their sections in the normal fashion during course

registration. Several studies which have appeared in the literature recently have deviated from the last requirement of this design. In the present investigation, mathematics ACT scores were used to adjust (linear regression) for ability differences among sections possibly produced by nonrandom section assignments.

VI. LIMITATIONS

1. The study was limited to the students enrolled in third quarter freshman biology classes at East Tennessee State University who attended class the days the data was collected. The total number of subjects available was 374. Because of the above facts the study was limited both as to subjects and geographic area.

2. The subjects did not meet the rigorous requirments for a random sample, in that students were not randomly assigned to class but chose their sections in the normal fashion during registration.

3. The operational definition of student achievement was arbitrary and of necessity narrow.

4. The ACT scores used were obtained from tests administered at different times and places prior to the subjects' entrance to college.

VII. ORGANIZATION OF THE STUDY

Chapter I is an introduction, statement of the problem, overview of the procedure and instrumentation needed for the study, and limitations of the study. A review of related and pertinent literature is presented in Chapter II. The procedure used, sources of data, a

description of measuring instruments, and a discussion of the scoring and statistical analysis are presented in Chapter III. Chapter IV presents the results and Chapter V the summary, conclusions, and recommendations for further study.

CHAPTER II

REVIEW OF THE LITERATURE

The extensive body of literature dealing with construction, validation, application, and interpretation of student ratings of college teaching has been recently compiled in several reviews* (Aleamoni 1974; Costin, Greenough, and Menges 1971; McKeachie 1969; and Menges 1973). Much of the present review is based on those expositions; other more recent studies pertinent to this study are also included.

The portion of the literature dealing with student ratings of instruction which is relevant to the present study is that comprised of the several attempts to validate student ratings of instruction by correlating them with student characteristics, e.g., demographic data, student achievement, personality traits, and other miscellaneous variables. It is within the limited context of these studies that the present investigation may be most clearly understood.

The preponderance of studies attempting to validate student ratings of instruction have treated student ratings of instruction as chronologically subsequent reflections of teacher or student characteristics. Of this group of studies, about half have used teacher characteristics as predictors and the other half have used student characteristics as predictors. A few studies have used the interaction

---

*The Education Index lists well over 150 studies published in the last decade alone on ratings of instruction and professors; an ERIC search found over 200 unpublished papers on the topic of student evaluation of teachers.

of student-observed teacher behavior and student characteristics as predictors of student ratings.

Because there are no studies which directly anticipate most of the hypotheses in the present paper, the literature reviewed was selected on the basis of its relevance to the broad question of how student characteristics have been found to be related to student ratings of instruction. It should be noted, however, that since the literature related to this broad topic went well beyond the scope of the present hypotheses, this review was limited to a summation of this extensive material. Rather than offering a detailed discussion of the available literature, this review was intended as an overview which would convey to the reader the flavor of the reported findings.

I. RELATED LITERATURE

Reliability of Student Ratings*

Reliability is seen as a precondition for validity, and considerable research has been directed to establishing the stability and internal consistency of student ratings and their dimensionality (Centra 1966, 1973; Cohen and Humphreys 1960; Costin 1968; Drucker and Remmers 1950, 1951; French 1957; Frey 1973; Guthrie 1954; Harvey and Barker 1970; Hildebrand and Wilson 1970; Isaacson et al. 1964;

---

*Antagonists (i.e., faculty members) of student ratings claim that student ratings are unreliable (no matter who prepares them), that the ratings will favor the entertainer over the instructor who gets his material across effectively, that ratings are highly correlated with expected grades, and that students are not competent judges of instruction since long-term benefits of a course may not be clear at the time it is rated (Costin et al., 1971, p. 511).

Lovell and Haner 1955; Maslow and Zimmerman 1956; Norr and Crittenden 1975; Spencer and Aleanomi 1970). This consistency provided evidence of students' ability to distinguish realiably among different instructors and different dimensions of teaching performance.* It is of interest to note that the Guthrie study found that student judgments were more stable than faculty judgments of teaching quality; and the Remmers study found that when 25 or more ratings were averaged, the means were as reliable as means of the better mental tests available. Frey (1974), examining ratings of students for 13 different instructors, obtained results which suggested that students can clearly discriminate among the teaching performances of different instructors. The outcome of the study by Frey (1973) which analyzed ratings from 72 students who had enrolled in both biology and organic chemistry strongly supports the contention that students can function as reliable observers.

Student and alumni ratings. The relationship between student and alumni ratings of teachers has been demonstrated in two studies, both of which have not substantiated the temporal quality of student ratings.** In a study conducted in 1951, Drucker and Remmers found that the average ratings given to 17 instructors by students and alumni were correlated

---

*Although there are several acceptable procedures for assessing the reliability of a measuring instrument, the most meaningful index of reliability for student rating data is the extent to which students agree about the strengths and weaknesses of different teachers (Frey, 1974).

**One of the complaints frequently voiced about student ratings of instruction is that the ratings, which are generally obtained at the end of a course, do not adequately reflect the long-term effects of instruction. Student immaturity or lack of perspective are reasons often cited for this shortcoming, and it is assumed that later ratings—say when the students are alumni—could be more valid measures of teacher effectiveness (Centra 1973).

positively. The correlations ranged from .40 to .68 on 10 teacher traits such as presentation of subject matter and interest in the subject.

The results of the Drucker and Remmers (1950, 1951) investigations refute the charge that student are immature and lack long-term perspective. The results of a study by Centra (1973) on the collinearity of student and alumni ratings indicated the judgment of teachers by their students at the end of a course are fairly permanent and mature. The agreement between current student and alumni regarding who have been effective or ineffective teachers is substantial, particularly in identifying teachers at the extremes (Centra, 1973).

Validity of Student Ratings*

Demographic variables. One approach to the problem of validation is through analyses of the correlations of ratings with other variables. A valid measure would be one which correlates with factors theoretically related to teaching effectiveness and which can be distinguished from so-called "biasing factors."

Very recent investigators have addressed the independence of rating from such biasing factors as student sex, major, years in school, age, and values; instructor sex and rank; and class size and level (Costin, Greenhouse and Menges 1972; Crittenden, Norr and LeBailey 1975; Norr and Crittenden 1975; Wood, Linsky and Straus 1974).

---

*Validity information speaks to the meaning of ratings both in the sense of internal structure and relationships to external criteria, e.g., exam performances residualized for ability or other premeasures of ability.

Sex. Classifying student ratings on the basis of the respondents' sexes, numerous investigators have found no significant differences between males and females in several types of ratings (Bendig 1953; Caffrey 1969; Dowie 1952; Elliot 1950; Heilman and Armentrout 1936; Lowell and Haner 1955; McKeachie, Lin and Mann 1971; Remmers 1939). Both Walker (1969) and Bendig (1953) have found that female students are inclined to differentiate between male and female instructors with female instructors as the beneficiaries of their preferences. McKeachie, Lin and Mann (1971) found that females thought teachers whom they rated high on skill and structure more effective.

Class level. Numerous studies have reported on the relationships between student ratings of instruction and class level, i.e., number of terms in college. Frey, Leonard and Beatty (1974) found that more senior students rated instructors more favorably than their less experienced classmates. This effect is consistent with the previous weak tendency reported by Rosenshine, Cohen, and Forst (1973) and Granzin and Painter (1973) for more experienced students to give higher ratings. One explanation for this relationship suggested by Frey et al. (1974) is the more positive attitude of upperclassmen; students with less positive attitudes would be less likely to become juniors and seniors.

Congruent with the Rosenshine et al. (1974) outcomes are other findings which report that students in upper division (i.e., juniors and seniors) tend to give higher ratings than do their lower division peers (i.e., freshman and sophomores) (Clark and Keller 1954; Dowie

1952; Gage 1961; Lowell and Haner 1955). In a comparison of graduate versus undergraduate students, Guthrie (1954), Heilman and Armentrout (1936), Stewart and Malpass (1966), and Rayder (1968) have found no real differences in the ratings given by students of graduate and undergraduate classifications. In contrast, Remmers and Elliot (1949) have found that graduate students assign higher ratings than do undergraduates.

Required courses versus elective courses. There is conflicting evidence as to whether there are differences between ratings given by students taking required courses on a required basis and those taking required courses on an elective basis. Gage (1961) and Lowell and Haner (1955) reported that teachers of required courses received significantly lower ratings than teachers of elective courses. Granzin and Painter (1973) reported a relatively weak but interesting correlation which showed required courses tend to receive lower ratings while there was virtually no relationship to the fact that the course was in one's major field.

Major versus nonmajor. Another demographic variable of interest, the student's major subject area, seems uncorrelated with student ratings. Cohen and Humphreys (1960), for example, have reported that psychology majors tended to rate instruction in a psychology course about the same as nonpsychology majors. Granzin and Painter (1973) found virtually no relationship between students' ratings of students who were majors and those who were nonmajors in 17 courses offered in 11 different departments.

Student-teacher interactions (size of class). Several investigations reported in the literature were designed to detect relationships in ratings of instruction and the size of the class.* Although this is supported to some extent by the results of several studies (Heilman and Armentrout 1936; Lowell and Haner 1955; McDaniel and Feldhasen 1970), other investigations do not support such a belief. For example, Goodhartz (1948) and Guthrie (1954) found that students were not reluctant to assign high ratings of overall teaching ability to instructors of large classes. Gage (1961) reported an interesting curvilinear relationship: teachers in courses with 30-39 students received lower ratings than teachers in courses with either more or fewer students. These studies relate directly to the present investigation.

Student hostility. A misgiving about student ratings (sometimes expressed by teachers) that students may assign poor ratings simply as a function of hostility, independent of the quality of instruction, was not supported by an investigation by Crannell (1948), who found no evidence that students tried to damage a faculty member through low ratings—possibly because they did not expect the rating forms to have any effect. Support for the Crannell study comes from Centra (personal communication) who found that students who believed their ratings would be used for tenure and promotion purposes gave virtually identical ratings as students who thought their ratings would be used for course improvement.

---

*Faculty members frequently suggest that teachers of larger classes may receive lower ratings because students generally prefer small classes which permit more student-teacher interaction (Costin et at. 1971).

Students' criteria for effective teaching. Several factor analyses of students' ratings of instruction have been performed, and depending on the number and nature of items included in each instrument the factors differ somewhat. Yet certain factors were common to several instruments: (a) organization or structure, (b) teaching skill or communication, or verbal fluency, and (c) student rapport or empathy (Coffman 1954; Gebb 1955; Hodgson 1958; Isaacson et al., 1964; Centra 1964; Lantrop and Richmond 1967; McKeachie, Lin and Mann 1969). It is significant to note that at least two studies indentified a factor as significant that halo effect,* if present, has a relatively minor influence on student ratings (Hodgeson 1958). More recently, Crawford and Bradshaw (1968) asked students to describe the most effective teachers they had ever had. The four most frequently mentioned characteristics were: (a) thorough knowledge of subject matter, (b) well-planned and organized lectures, (c) enthusiastic, energetic, lively interest in teaching, and (d) student-oriented, friendly, willing to help students. Although phrased somewhat differently, these characteristics resemble the ones in a list by French (1957) and a student-generated list in a study by Guthrie (1954).

---

*More than thirty years ago, Thorndike pointed out on the basis of experimental evidence that a rater has a constant tendency to rate an individual either high or low in many traits because the rater knows (or thinks) the individual to be high or low in some specific or particular trait. Thorndike called this tendency the "halo" effect (Thorndike, 1920). Applied to the learning situation, Thorndike's statement means that if the student regards an instructor as very satisfactory in terms of a trait such as general personality or the ability to "entertain," he is likely to rate the instructor high also in such traits as teaching skill, i.e., communication or verbal fluency, organization or structure, student empathy, ingenuity and perhaps many other traits.

In an attempt to determine the basis on which students make their judgments, McKeachie et al. (1969b) measured teaching effectiveness as rated by students who were tested on various aspects of "critical thinking." They found that students who performed particularly well on the test rated their teachers as "more effective" than did students who performed poorly on the test. Similar findings were reported by Elliot (1950) and Mann (1969). Other supportive studies have been conducted by Lantrop and Richmond (1967), McKeachie and Solomon (1958), and Pohlmann (1975). These studies relate directly to the present investigation.

It would appear, then, that students are at least partially capable of distinguishing certain qualities of instruction which increase their knowledge or motivation.

Costin et al. (1971) reported that two-thirds of the students randomly selected in a study on student opinion of student ratings of instructor performance expressed interest in rating even to the extent of spending time outside of class. Costin concluded that students did not think ratings are a waste of time. Moreover, the Costin study emphasized that the students used independence in their ratings and did not rely merely on the interchange of "student gossip."

Student personality traits. The predictive power of student personality characteristics has been investigated by several researchers. Yonge and Sassenrath (1968) have found that self-reported student personality characteristics and student ratings of instruction are correlated within classes but that the observed correlations do not hold

from instructor to instructor. They have concluded that ratings of instruction have different meanings from instructor to instructor because of the differences among instructors in teaching methods and because of differences in the reactions of various student personalities to these differing methods.

The results of an investigation by Grush and Costin (1975) showed that teacher skill was not related to students' self-perceived personality traits or their attraction to teachers "as persons," but were correlated positively with certain teacher personality traits and attraction to instructors "as teachers." It was concluded that students are objective consumers of the teaching process and their judgments can be used to identify variables important for teacher effectiveness. A unique study by Rezler, 1965, found that student needs (as measured by the Edwards Personal Preference Scale) influenced some items on the Purdue Rating Scale.

Teacher personality measures and student ratings. When teacher personality measures and student ratings have been correlated, the relationship has been generally negligible (Borg, 1957; Bendig 1955). However, psychology instructors rated high in "general cultural attainment" by their peers tended to receive good ratings from students (Isaacson, McKeachie and Milholland 1963). This latter finding has been substantiated by other research (Cosgrove 1959), although in subject areas other than psychology students may rate other faculty qualities as more important (Riley, Ryan, and Lifshitz 1950).

Faculty research and instructional outcomes. Research data on faculty research involvement and instructional outcomes are inconsistent. Positive relationships have been reported by Bresler (1968) and Singhal and Stallings (1969). Hayes (1971) found a positive relationship when both were judged by department heads (halo?) but no relationship when students judged teaching effectiveness and research productivity was either rated by department heads or determined by publication count. No relationship was found between effectiveness (as rated by students) and research productivity in studies reported by Hoyt (1974), Harry and Goldner (1972), McDaniel and Feldhusen (1970), Voeka (1962), and Guthrie (1949).

Teacher experience and instructional performance. Teachers with less than five years experience tend to be rated lower than teachers with more than eight years experience (Remmers 1929). Similarly, graduate teaching assistants as a group were consistently rated lower than full-time faculty (Centra 1966).

External or criterion-oriented validity. An additional group of studies having a direct relationship to the present study addresses the external or criterion-oriented validity of ratings using student learning as the criterion. With learning defined as examination performance residualized by some premeasure of ability, correlations between overall ratings and learning have been significant and positive (Wherry 1952; Spencer 1968; Walker 1969; Weaver 1960; Treffinger and Feldhusen 1970), near zero (Wherry 1952), and significantly negative (Rodin and Rodin

1972).* Correlations involving more specific ratings have generally been positive and significant or positive but not significant; no specific negative correlations seem to have been found for these more specific ratings (Doyle and Whitely 1974; Elliot 1950; Frey 1972; McKeachie, Lin and Mann 1971; Remmers, Martin and Elliot 1949). More recent studies reported by Frey (1974) of his other work and that of Authur Sullivan and Peter Gessner, indicated that findings from older studies—low positive correlations (.32-.39)—can be replicated when acceptable methodological procedures are employed.

A recent investigation by Scott (1975) provides additional support for the contention that at least some information from student ratings is positively related to achievement. These studies relate directly to one hypothesis in the present study.

Expected grades and ratings. Another group of studies having a direct relationship to the present studies are those attempting to correlate student expectations for grades and grades actually received with student ratings of instruction. Numerous studies have found no relationship between expected grades and/or received grades and student ratings of instruction; however, studies using received grades are included here because a strong relationship has been observed between

---

*The Rodin and Rodin study has been contested on methodological grounds by Frey (1974) and by Rodin, Frey and Gessner (1975). Subsequent replications of the basic design using regular faculty rather than teaching assistants and using more sophisticated rating forms have reported a positive relationship rather than a negative relationship (Frey 1973; Gessner 1973; Sullivan and Skanes 1974).

expected and received grades (Bendig 1953; Blum 1936; Cohen and Humphreys 1960; Eckert 1950; Garverick and Carter 1962; Guthrie 1949, 1954; Heilman and Armentrout 1936; Hudelson 1951; Voeks and French 1960; and Walsh 1967). Also, students with higher overall grade-point averages do not rate instructors differently than lower-achieving students (Guthrie 1954).

Student ratings and the halo effect. There is no direct evidence to support the viewpoint that students may judge instruction on the basis of its "entertainment" value rather than on its information contribution to learning or long-term usefulness. Royce (1956) reported that higher ratings were a function of "superficial popularity." To the contrary, Weaver (1960) found that ratings were not a product of a popularity "halo." Supportive of the Weaver (1960) outcome were research by Remmers (1934) and Remmers and Stalnker (1928) who found that raings of multiple attributes of teaching were relatively low on "halo effect." In agreement was an investigation by Guthrie (1954) which indicated that teachers who obtained the highest ratings were also "substance" teachers and not merely entertainment oriented. Findings in a study by Costin (1971) seem to support a conclusion that students use independence in their ratings and do not rely merely on the interchange of "student gossip."

Anonymity of ratings. Researchers disagree on the context in which student ratings of teachers are obtained. Costin et al., (1971) believes that a guarantee of anonymity or immunity would seem to be mandatory,

especially when ratings are administered (as is typical) prior to final grading. In contrast, Siebring and Schaff (1974) reported the following response of a chemistry professor to an open-ended question concerning student ratings of instructional performance:

> . . . one can obtain useful information about student response to a given course or instructor—but only by detailed analysis of the results. To do this properly student anonymity must be abandoned—to date no one has had the courage to insist on this. (p. 154)*

Brooks and Levenson (1974) categorized the anonymity issue as one of ethics, not one of validity, and concluded from a well-designed statistical study that self-identification does not cause the ability of students to discriminate among lecturers to be dramatically altered. In fact, Brooks and Levenson are convinced that identified forms definitely have their place in finding out what is going on in a course.

Self-reported student grades. In many cases the data reported in these studies concerning grades depended on the accuracy of students' self-reports. It is of interest to note, therefore, a recent investigation by Walsh (1967) which found that students do report accurately such data as course grades and grade point average. The Welsh results were substantiated by results of a major study, conducted by two Massachusetts

---

*In the present study, a detailed analysis of results was possible by providing for identification of each respondent participating in the investigation. While it is difficult to judge trends in the literature, it appeared to the writer that some investigators are utilizing the identification technique more frequently in an attempt to obtain technical sophistication in experimental designs commensurate with designs in the natural sciences.

researchers, published in The College Board Review, January 1976. The two researchers, Armstrong and Jenkins, reported results showing that students generally report their high school grades accurately when they fill out the Student Descriptive Questionnaire (SDQ) accompanying the registration forms for the "college boards."

The analysis of these self-reported grades showed that an average of 97 percent of the grades were reported accurately within one grade category and an average of 72 percent of the grades were reported with exact accuracy. The two research efforts mentioned above directly anticipate three of the hypotheses in the present study.

Skewness of student ratings. Data collected from student ratings of instruction using the SIR are typically skewed in a positive direction. On a five point (excellent to poor) scale, for example, a mean score of 3.6 is numerically above average, but in comparison to other teachers it may be only average or even slightly below average. This is a fact which is entirely necessary for a clear interpretation of results of student ratings, particularly for individual instructors who receive summary reports of their own students' evaluations. The above information alludes to the interesting fact that the student population in general is "soft" on teachers when they rate them.

II. DISCUSSION OF THE LITERATURE

Taken as a whole, the outstanding characteristics of the studies under discussion is that student ratings of instruction provide reliable and valid information on the quality of courses and instruction and the

data which have been reviewed strongly suggest that the formal use of student ratings provides a reasonable way of measuring student reaction.

There are many possible causes of the conflicting results in the data which have been reviewed; such as, dissimilar samples, dissimilar instruments, dissimilar experimental procedures, dissimilar student populations, different setting events, etc., and there can be little doubt that these causes have played an important role in the production of these conflicting results. It was, however, the position of this investigation that most of these studies would have produced conflicting results even without these problems because of (1) poor experimental designs lacking proper technical sophistication for rigorous testing of hypotheses, (2) failure to identify students (because of the anonymity issue) in investigations of this type which prevented detailed analysis of the results, and (3) failure to address the question of how various student characteristics, e.g., attitudes toward the subject, level of cognitive development, and high school background, might vary with different modes of teacher behavior, and the level of abstraction needed by the students to understand the material discussed in the course.

Actually researchers have not been completely unaware of this possibility. In fact, this type of interaction has been suggested by Centra (1973b) who examined several types of subgroups of students in a typical class, each of which reacted a little differently to the teacher and the course.

Central (1973b) emphasized the importance of the context (i.e., the particular course) in understanding and interpreting the meaning of

student ratings. Too, this type of interaction has been explicitly suggested by Yonge and Sassenrath (1968) who found that various student personality characteristics were correlated with student ratings in some classes but not in others. Finally, Zelby (1974) found that student ratings in two sections of a course called "Technology in Society" were dichotomized—students in the section in which more structured techniques were used gave significantly higher ratings than students in the section employing unstructured techniques. This finding corresponds to the results of the investigations by Griffiths (1973) and McKinnon (1971) discussed in Chapter I, page 7. Griffiths and McKinnon found that over half of the students enrolled in an introductory physics class in college were not able to employ abstract thinking in dealing with scientific activities. It is reasonable to assume that students in this category might exhibit dislike for unstructured classroom methods since their abilities to work in a scientific paradigm, i.e., deal with novel situations and ideas, complexity (as opposed to simplicity), and ambiguities and uncertainties, would be quite restricted.

There is at least one other explanation for the conflicting results which have been reported. It may be that many of the variables which have been and which are being studied have not correlated with student ratings of instruction simply because they are theoretically and empirically irrelevant. But it was the position of this study that the predictive relevance of these variables will remain indeterminate until studies using some form of experimental and/or statistical control for these variables (student, teacher, and course characteristics) and interactions among them are implemented.

This study identified the existence of several subpopulations of students in introductory biology classes and emphasized the need for detailed analyses of student characteristic variables which might lead to different points of view of instruction and courses which may affect ratings. The present investigation underscores the point made by Yonge and Sassenrath (1968) and Centra (1973) that the context of the ratings (i.e., the particular course and instructor) are often important in understanding the meaning of student ratings.

CHAPTER III

PROCEDURE

The procedure used in this study involved four major steps. These steps were the selection of the subjects, a description of the method, a description of the instruments used, and a scoring and computation of statistics.

I. SUBJECTS

The population from which the observed sample was drawn was comprised of students attending classes in the Biology Department of East Tennessee State University during the Spring Quarter of 1975. The students were primarily first-year students enrolled in a multisection course. The course consisted of 13 sections of introductory biology taught by four full-time staff members of the Biology Department, and one full-time staff member from the Department of General Science and Science Education.

Despite the sensitive nature of student ratings of college teaching, all of the instructors involved in the study participated. Each staff member was assured confidentiality of results; also, each staff member was told that he alone would be provided with a copy of his individual rating summary reports. This assurance undoubtedly contributed to the excellent cooperation of the Biology Department teaching staff; in fact, 100 percent of those contacted participated fully in all phases of the study. Memoranda to the Biology Department Chairman and to the staff members involved in the study are included in Appendix A.

Students chose their sections in the normal fashion during course registration. The instructor for each section had primary responsibility for the organization and conduct of classroom activities. All sections used a common course syllabus, common textbook, and common exams.

II. METHOD

Each of the 374 students in the sample completed three questionnaires administered during the last week of the quarter during regularly scheduled class periods. The first instrument gathered information on student reactions to instruction and requested information about each student, such as cumulative grade-point average (GPA), and anticipated grade in the course. The second instrument asked about the respondent's attitude toward biology. The third instrument collected information pertaining to the student's theoretical orientation. In addition, ACT test scores, cumulative grade-point averages, information concerning high school chemistry courses, and final grades assigned for the course were obtained from the records office. The total number of points accumulated on three common exams, from which the final grade was computed, was obtained from the coordinator of the freshman program, with permission of the instructors.

Questionnaires were presented as collecting information for a curriculum study. Members of the University's student personnel group (Dean of Students, Director of Financial Aids, Director of Religious Affairs), clearly not connected with the class in which students were enrolled, administered the questionnaires. Each proctor was provided

identical and explicit instructions for administering each instrument. A copy of these instructions is shown in Appendix B. An orientation session just prior to the collection of the data was conducted in an attempt to standardize procedures.

Students were not asked to provide their names; instructors were provided with summary reports only. Questionnaires were marked to facilitate matching of the three instruments for each student and to allow identification of each student's reported cumulative GPA, expected grade, actual grade, and number of total exam points accumulated. This information was needed to test several of the hypotheses in the study.

III. INSTRUMENTATION

The three questionnaires administered to all subjects were the Student Instructional Report (SIR), a Biology Attitude Scale, and the Omnibus Personality Inventory (OPI). Appendix F contains a copy of the SIR, the Biology Attitude Scale, and sample questions from the OPI.

IV. THE STUDENT INSTRUCTIONAL REPORT (SIR)*

The SIR, an instrument that students can use to report on the instruction in their classes, was developed by John A. Centra, senior research psychologist at Educational Testing Service, Princeton, New Jersey. The project was carried out with the cooperation of students and faculty members of Northwestern University.

---

*Sources of rating forms and related issues in Appendix F, page 152.

The Student Instructional Report is a relatively new instrument—in use since May 1972. In that time over 250 institutions have used SIR in specific departments or throughout the institution. Thirty institutions, with a total of 3,302 classes have been used to prepare Comparative Data Reports. Universities, liberal arts colleges, and two-year colleges have been included in collecting these comparative data for the SIR (SIR Report Booklet 1973).

The Comparative Data Reports are summarized and categorized for combined Reports and for Percentile Distributions in a variety of breakdowns (Comparative Data Report Booklet 1973). These are shown in Appendix C. The combined report for the biological sciences and health fields is shown in Appendix C. Appendix C also contains the Percentile Distribution of SIR items means for biological sciences and health fields. It is appropriate to compare the standard deviations from an individual instructor's class report with these figures.

Either local college comparative data and/or the data in the Report Booklet are important because student ratings typically are skewed in a positive direction. On a five point scale (excellent to poor), for example, a mean score of 3.6 is numerically above average, but in comparison to other teachers, it may be only average or even slightly below average (Comparative Data Report Booklet 1973).

The SIR is printed on an IBM sheet ready for computer scoring, but it can also be easily hand scored. Instructions for administering the SIR and for interpreting the results are clear. The SIR usually can be completed in 10 to 15 minutes (Student Instruction Report Booklet 1973).

The SIR contains several items designed to generate helpful information concerning student characteristics. The majority of the items are descriptive and thereby intended to reflect the instructor's teaching style. There are 39 items with responses of agree, disagree, multiple choice and five point ratings.

Appendix C, Table C-1, shows percentile distribution of item means, standard deviations of the means and average standard deviations for SIR items 1-24 and 32-39. For the percentile distributions, all the item means were distributed at 10 percent intervals. The shaded column contains the 50 percent or median mean—that is for each item half of the means are higher and half are lower than the one in the shaded column. In the 80 percent column, 20 percent of the means for that item are higher and 80 percent are lower. For items 6 and 16 a higher mean is negative and a lower mean is positive or "better."

Means for Combined Reports and also those from individual class reports, may be compared to the means shown in Appendix C. The Mean Column contains the mean of class reports means. Means from Combined Reports and also those from individual class reports may be compared to these means. The Standard Deviation of the Means Column contains the standard deviations from Combined Reports with these figures. The Average Standard Deviation Column contains the average student deviation of the class report means.

Analysis of various reliability estimates for the SIR items were generally about .70 for 20 or more students in a class and slightly less for 15 or more students. The data are presented in Appendix C,

Table C-2. The lowest reliabilities may be noted on items dealing with such aspects as the level of difficulty of the course, for which students within a class might more likely be expected to differ.

Test-retest correlations, as indicated in Appendix C, Table C-2, were moderately high with the majority of items near or above .70. The SIR has as high a reliability as the better mental tests (SIR Booklet, Report 1, 1972).

Item reliabilities were estimated by comparing the variance of responses within classes to the variance between classes. Using this approach an estimated reliability coefficient was obtained (Winer, 1962). Through use of the Spearman-Brown formula, estimated reliabilities for classes of various sizes were computed.

V. THE BIOLOGY ATTITUDE SCALES

The Biology Attitude Scales was developed by James Russell, Purdue University and Steven Hollander, Standard Oil Company. The entire scale is short—only 22 questions—and easily scored. The scales contain 14 Likert-type items and 8 items in a standard bipolar adjectival scale (semantic differential scale). It can be used in detecting and in measuring change in attitude or absolute attitude. The items were selected if they were either clearly favorable or clearly unfavorable to the concept of attitude toward biology.

Concurrent validity of the Biology Attitude Scales was determined by correlating the total scores on the two instruments included in the scales—the semantic differential and the Likert scale. The mean

correlation for the two was 80 percent. The test-retest reliability of the semantic differential averaged about 80 percent over seven weeks, while the reliability for the Likert Scale was never under 90 percent.

The Biology Attitude Scale was published in the May, 1972 issue of the American Biology Teacher.

VI. THE OMNIBUS PERSONALITY INVENTORY

The Omnibus Personality Iventory (OPI) was developed by Paul Heist and George Yonge at the Psychological Corporation and was published by the Psychological Corporation.

The OPI-form F is an instrument containing 385 statements designed to measure differences among college students with regard to their attitudies, opinions, and feelings on a variety of subjects. Each subject is chiefly relevant in areas of normal ego-functioning and intellectual activities. Each item belongs to one or more of the 14 scales which constitute the OPI. Almost all dimensions in the inventory were chosen for their general importance in understanding and differentiating among students in an educational context.

The student is asked to read each statement in the booklet and mark his responses on a separate answer sheet. He is instructed to mark true if the statement is true or mostly true for him, and false or not usually true as applied to him. The measured characteristic is generally defined in terms of a description of high scorers, the logical opposite of this description would, in most cases, characterize low scorers.

The point at which any score may be defined as a high score relative. On most scales, standard scores of 60 (84th percentile) or above are interpreted as sufficiently high for the essence of the respective score to apply. Twelve or more colleges or universities selected all of their enrollees from the top 10 to 15 percent of secondary school graduates. Consequently, the standard scores are based on means and standard deviations computed for a group that may be above average of college freshman. Since the participating University in this study has open enrollment, a high scorer for this study is defined as one having a standard score of 50 or above.

A more complete description of a high scorer on the theoretical orientation scale (TO) used in this study appears below (OPI Manual 1968).

> High scorers indicate a preference for dealing with theoretical concerns and problems and for using the scientific method in thinking; many are also exhibiting an interest in science and in scientific activities. High scorers are generally logical, analytical, and critical to their approach to problems and situations. (p. 4)

For the OPI, validation data have been drawn from many sources and utilize several types of information. Correlations with other measures provide the core of these measurements. Several of the intercorrelations of the 14 scales contribute to the construct validity of a given scale. The results of a factor analysis support the construct validity. A number of investigators to date have included some or all of the OPI scales in predictive studies (OPI Manual 1968).

Validation Data OPI

The validation data available for the OPI is presented scale by scale. The theoretical scale consists of 33 items. Evidence in

support of the TO scale, as a measure reflecting an interest in problem-solving, logical or critical thinking, and science is also found in quite a variety of correlations with established measures. The substantial correlation with the Study of Values theoretical scale (.62) is entirely expected (Appendix D, Table D-1) when considering the development of this dimension (Center for the Study of Higher Education 1962). Referring now to (Appendix D, Table D-2) the TO scale correlates with an experimental measure of problem solving ability (.53) as well as with the mathematical test (.46) from the CEEB Scholastic Aptitude Test (SAT) (College Entrance Examination Board 1959). It should also be stated that TO correlates substantially with SAT verbal (.52) (OPI Manual 1968). The TO correlation with problem solving remains significant when SAT verbal and mathematical abilities are partialled out. Thus, to some extent, this scale reflects problem-solving ability, as well as interests along these lines.

The theoretical-scientific interest components of TO is manifested in correlations with the Strong Vocational Interest Blank for men. Attention is called to coefficients for TO and the following scales: Psychologist (.51), Physician (.42), Mathematician (.39), Physicist (.39), Chemist (.44), Banker (-.15), Mortician (-.46) (OPI Manual, 1968).

The correlation of .51 with the thoughtfulness measure on the Guilford-Zimmerman Temperament Survey (Appendix D, Table D-3) supports the contention that TO reflects, to some extent, interest in logical, critical thinking. With respect to the Activities Index (Stern 1958), TO correlates highest with a need for Scientism (.47) and understanding (.44) (Appendix D, Table D-4).

The TO scale correlates with faculty ratings (Appendix C, Table D-5) for "self-reliance and originality" (.35) and for overall effectiveness as a graduate student and future scholar (.33).

Reliability Data for the OPI

Three different approaches to the estimation of the reliability of the OPI exists. Since it is incorrect to refer to "the" reliability of a scale or inventory, the results of three different approaches to the estimation of the reliability of the OPI are presented in the OPI Manual, 1968. These estimates are based on different samples. Coefficients derived by the Kuder-Richardson formula 21 (KR21) (Appendix D, Table D-6) and the corrected split-half method are estimates of the internal consistency of the OPI measure. The KR 21 figures range from .67 to .89. The great majority of the test-retest reliability coefficients (Appendix D, Table D-6) are above .85, with approximately half falling at .89 or above. The intercorrelations of the 14 OPI scales are presented in Appendix D, Table D-7.

Validity, chiefly as determined by external tests and criteria is dealt with in the previous section, but it is appropriate to point out here that several of the intercorrelations in Table D-7 of Appendix D contribute to the construct validity of a given scale.

The norms for the OPI are based on a large sampling of entering students at diverse institutions of higher learning. For convenient interpretation, the norms of the OPI are presented as standard scores. Raw scores on all 14 scales are easily converted to a standard score or percentile score.

VII. SCORING AND COMPUTATION OF STATISTICS

An instructor score was derived from each respondent of the SIR by adding the responses of items 1-20. Four points were possible for each of the 20 items making a total maximum score of 80. Items which were marked by students as not applicable or items which were marked omitted by the respondent were deleted from the teacher score before it was calculated. The students responded to each item by marking the statement as not applicable, 0 points; strongly agree, 4 points; agree, 3 points; disagree, 2 points; or strongly disagree, 1 point; with the exception of items 6 and 16 which reversed the point system outlined above.

After the instructor score was computed for each respondent the answer sheets were sent to Educational Testing Service, Princeton, New Jersey, for further compilation and analysis. A summary report was produced for each class in which the SIR was administered. A copy of a class report is shown in Appendix C. Each instructor participating in the study received a copy of the summary reports for his sections. The number of students who responded to the questionnaire appears in a box in the shaded bar at the top of the report.

The two shaded vertical columns on the right contain the mean and standard deviation for each item where it is appropriate to compute them. The mean for an item is obtained by multiplying the number giving each response by the number at the top of the appropriate column (the numerical value of each response on the rating scale) and then averaging the responses. The number, not the percentage of all students giving a

particular response, is used to calcuate all statistics. With the exception of items 6 and 16, a high mean is "better" or more positive.

The standard deviation is best described as the spread among the responses to any item. The smaller the standard deviation, the greater the consensus among students in the class.

The Biology Attitude Scale was hand-scored. Response categories for the Likert Scale range from 1 to 5. Scores are assigned such that 5 reflects a strongly favorable attitude. An example of a negative item and positive item is shown in Appendix E.

The score for the Likert Scale was computed by totaling the individual scores for each item.

For the semantic differential scale the bipolar adjectives are assigned values from 1 to 5 with the positive adjective 5 and the negative adjective 1. An example of an array of adjectives from the scale appears in Appendix E. The semantic differential score is computed by totaling the individual scores for each item. A single score is derived by adding the scores of the two scales.

The OPI was hand-scored by the use of a hand-scoring key available from the Psychological Corporation.

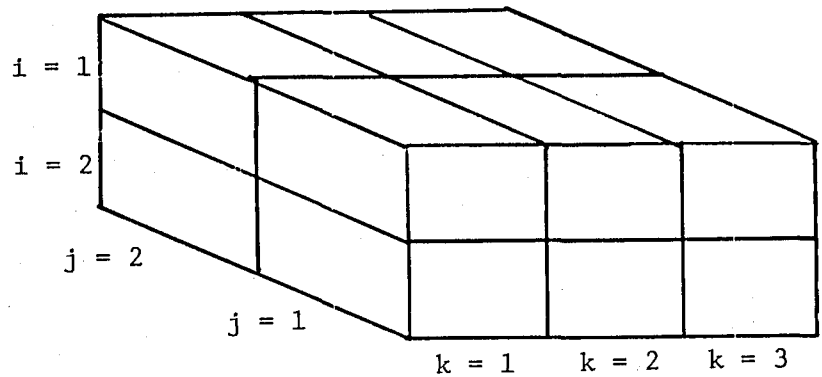
A 2 × 2 × 2 × 3 factorial design was used to test for the effects of three students characteristics—attitude toward biology, completion of chemistry courses in high school, and theoretical orientation—on their ratings of college instructors in freshman biology classes. The factorial design is limited to those students enrolled in sections taught by professors from the Biology Department whose entire teaching assignment

was for the main campus; it is also limited to the number of students who attended classes the days the data were collected. The students were divided into two groups (high or low) based on their scores on a Biology Attitude Scale. The students were divided into two groups on the basis of whether or not chemistry was included in their high school program. The students were divided into two groups on the basis of their scores on a personality test (high or low). Three instructors composed the fourth factor. English and mathematics ACT test scores were used as covariants.

The factorial design described above resulted in an unequal number of subjects in the cells. The procedure of unweighted means analysis was used to handle the problem of unequal $\underline{n}$.

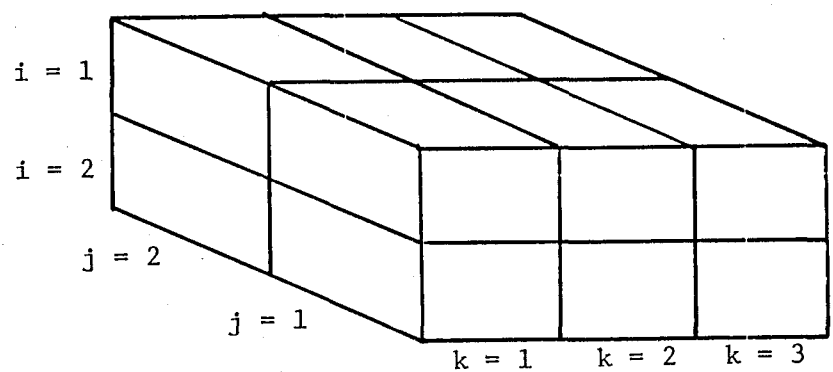
The partitioning of the data for the analysis of variance can be visualized conveniently in the form of Figure 1. The definition of two attitude groups, two theoretical orientation groups and three instructors may be thought of as creating $2 \times 2 \times 3 = 12$ cells represented by individual blocks in both visual representation 1 and visual representation 2 in the figure; the total number of cells equals 24. With each cell are associated three letters, i, j, and k ($i = 1$ or 2); ($j = 1$ or 2); ($k = 1$, 2, or 3); i identifies the two attitude groups (positive or negative), j the theoretical orientation groups (high or low), and k the instructor (A, B, or C). Visual representation 1 denotes one of the two levels of the fourth variable, chemistry in the high school program, and visual presentation 2 connotes the other level of variable four, no chemistry in the high school program.

No High School Chemistry



Visual Presentation 1

High School Chemistry



Visual Presentation 2

i = attitude	j = theoretical orientation	k = instructor	Visual Presentations
1 = positive	1 high	1 = instructor A	1. No chemistry in high school
2 = negative	2 low	2 = instructor B	2. Chemistry in high school
		3 = instructor C	

Figure 1. Visual presentation of the partitioning of the analysis of variance with the identification scheme.

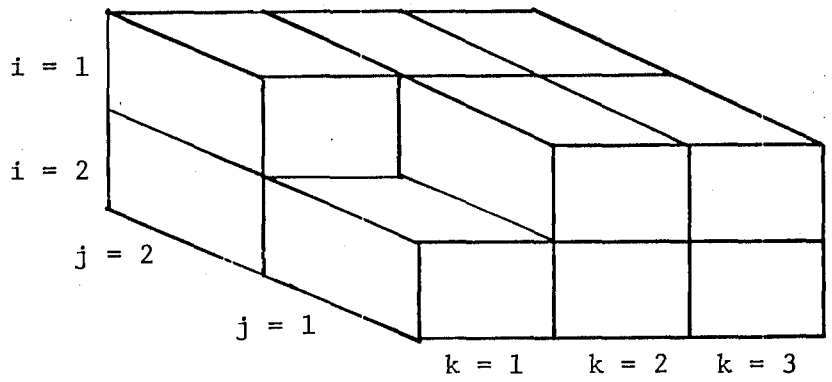
Every observation in the i, j, k cell will belong to each of the individual groups, and conversely every observation in the sample which belongs to these respective groups is in cell i, j, k. For example, in visual representation 2 of Figure 1, cell 1,1,1, consists of all the students with a positive attitude in the high theoretical orientation group who included chemistry in their high school program and who rated instructor A. In the figure, this cell is in the upper layer and is at the corner nearest the viewer. Figure 2 displays this information more completely. The number of subjects in each cell and the mean rating for each cell is displayed in Table 1.

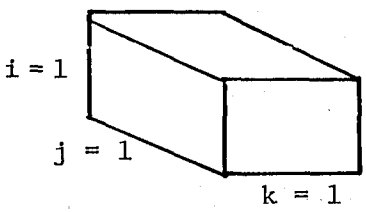
Our first concern is for the appropriateness of the interaction terms in the model because the presence of these obscures the main effects and complicates the interpretation of results. Therefore, it is reasonable to ask whether the data supports their inclusion. We thus formulated the null hypothesis that the interaction variability in the data is attributable to chance and adopted the model for further analysis.

Hypotheses associated with the four main effects were tested by means of F-tests. The covariants, GPA, MACT and English ACT scores, were used to adjust for errors due to ability.

Hypotheses 9, 10, 11 and 13 were tested by use of correlation coefficients.

Hypotheses 14, 15 and 16 were tested with one-way analysis of variance tests. Barlett's test for homogeneity of variances checked for skewness in the distribution which might lead to a





Block 1, 1, 1
i j k

The block removed consists of all subjects in the study who fell into the following categories:

1. High school chemistry
2. Positive attitude
3. High theoretical orientation
4. Instructor A

Figure 2. A visual display of block 1, 1, 1 listed from Visual Presentation 1, Figure 1, with contents of the block listed.

Table 1

Means for Each Block in the Factorial Design

Attitude	Chemistry	Theor. Orient.	Instructor	No.	Rating
0	0	0	0	9	62.00
0	0	0	1	18	54.33
0	0	0	2	19	60.47
0	0	1	0	13	57.54
0	0	1	1	23	54.09
0	0	1	2	33	59.03
0	1	0	0	4	58.75
0	1	0	1	7	58.43
0	1	0	2	5	70.40
0	1	1	0	3	54.00
0	1	1	1	10	53.50
0	1	1	2	10	60.70
1	0	0	0	3	55.00
1	0	0	1	8	56.00
1	0	0	2	13	55.62
1	0	1	0	16	54.81
1	0	1	1	28	47.54
1	0	1	2	26	57.00
1	1	0	0	4	61.75
1	1	0	1	9	53.56
1	1	0	2	14	58.71
1	1	1	0	10	50.30
1	1	1	1	20	49.40
1	1	1	2	25	55.60

Attitude	0 = Positive 1 = Negative
Chemistry	0 = High School Chemistry 1 = No High School Chemistry
Theoretical Orientation	0 = High 1 = Low
Instructor	0 = A 1 = B 2 = C

misinterpretation of results. An Eta was calculated for each analysis to determine whether curvilinearity was present to eliminate the possibility of a underestimation of r.

Hypothesis 12, 17, and 18 were tested with correlation coefficients and t-tests.*

*For a clear but brief review of interpretation of F tests, t-test, and correlation coefficients see A Simplified Guide to Statistics by G. Milton Smith (fourth edition) 1970. For a discussion of the less well-proven and/or more complicated procedures (i.e., Barlett's tests, Scheffe Procedure, and calculation and interpretation of ETA) see Quinn McNemar's Psychological Statistics (fourth edition), and B. F. Winer's Statistical Principles in Experimental Design, 1962.

CHAPTER IV

RESULTS

Hypotheses may be couched in four broad statements: the literary null, operational null, literary alternative, and operational alternative. The hypotheses presented in this study were stated in the literary null or operational null.

I. ANALYSIS OF RESULTS

The statistical analysis of the data was accomplished primarily through the use of The University of Tennessee IBM 360/65 computer. The programs used were selected from those described in the Statistical Analysis System (SAS) designed and implemented by Antony James Barr and James Howard Goodnight at North Carolina in 1972;* and from the Statistical Package for the Social Sciences (SPSS), published by McGraw Hill Publishing Company in 1975.

Tests of Hypotheses

The first three hypotheses were designed to detect biases in student ratings of instructorial or professorial performance in introductory biology courses at the college level due to the effects of the students' theoretical orientation, attitude toward biology, and

---

*A user's guide to the Statistical Analysis System is available through students supply stores, North Carolina State University, Raleigh, North Carolina 27606.

chemistry background in high school. The probability level for rejection of all hypotheses was set at the .05 level.* This decision rule, by necessity, was set prior to collection of the data.

Prior to the analysis of the first four hypotheses it was necessary to test for possible interaction in the model since interaction obscures the main effects and complicates the interpretation of results. In other words, interaction implies that the main effects reinforce or neutralize each other in some way at some point. Table 2 includes a summary of the analysis of interaction and shows that none of the eleven computed F's were significant at the .05 level. These results failed to reject the null hypotheses at the .05 level and the conclusion was that interaction variability in the data was attributable to chance.

Statement of the first hypothesis was that a student's rating of his biology instructor's performance is not influenced by the student's attitude toward biology. The probability of obtaining a computed F of 10.87 was .0011; therefore the analysis rejected the null hypothesis.

The second hypothesis was that a student's rating of his biology instructor's performance is not influenced by his theoretical orientation, i.e., his level of cognitive development. There was a significant difference in the ratings of high and low scorers; therefore the analysis rejected the null hypothesis. The computed F was 13.60 which was significant to the .0003 level.

---

*Behavioral scientists ordinarily are willing to reject the hypothesis of no effect at a probability level of .05 or .01. In other words, it is conventional to reject the hypothesis of randomness where the chances are either nineteen to one against it or 99 to one against it. The question of whether 5 percent or 1 percent significance level will be employed seems to be a matter of personal preference. This convention has been widely agreed upon (From: Introduction to the Statistical Method by Kenneth R. Hammond and James E. Householder, Random House, 1962, p. 349).

Table 2

Analysis of Variance Table, Regression Coefficients, and Statistics of Fit for Dependent Variable Rating*

Source	df	ss	ms	F Value	Prob F
Regression	23	6109.45	265.63	3.66	0.0001
Error	306	22236.24	72.67		
Corrected Total	329	28345.70			
		Partial ss			
Attitude	1	790.15		10.87	0.0011
Chemistry	1	40.81		0.56	0.4542
Theoretical Orient.	1	988.30		13.60	0.0003
Instructor	2	2017.99		13.88	0.0001
Att - Chem	1	5.62		0.076	0.7812
Att - TO	1	0.10		0.002	0.9696
Att - Instr	2	125.35		0.863	0.5738
Chem - TO	1	236.31		3.252	0.0723
Chem - Ist	2	189.41		1.303	0.2722
TO - Ist	2	42.90		0.295	0.7490
Att - Chem - TO	1	2.76		0.037	0.8456
Att - Chem - Ist	2	186.63		1.284	0.2776
Att - TO - Ist	2	227.35		1.564	0.2091
Chem - TO - Ist	2	138.25		0.951	0.6106
Att - Chem - TO - Ist	2	165.19		1.137	0.3224

N = 330

R - square = 0.22

C.V. = 15.31%

S.D. = 8.52

Rating Mean = 55.70

Att = Attitude toward biology

TO = Theoretical Orientation

Instr = Instructor

Chem = Chemistry variable

*Two levels of attitude toward biology; two levels of theoretical orientation; two levels of high school chemistry; and three levels of instructors.

Statement of the third hypothesis was that a student's rating of his biology instructor's performance is not influenced by the presence or absence of chemistry courses in his high school program. The probability of obtaining a computed F of 0.56 was .4542. The results failed to reject the null hypothesis.

The fourth hypothesis was that there is no difference in the way students rate different instructors in a multisection biology course. The computed F was 13.89; the significance level was .0001. The null hypothesis was rejected. The results of the F tests are gathered and conveniently displayed in Table 2 as a single 4-way analysis of variance table.

Statement of the fifth hypothesis was that the correlation coefficient obtained between a student's reported GPA and actual GPA is not significant. The computed coefficient of 0.704 was significant at the 0.001 level and the hypothesis was rejected. A chi-square of 443.02 with 25 degrees of freedom was significant beyond the 0.001 level. A computer-generated contingency table of the reported GPA and actual GPA is displayed in Table 3.

The sixth hypothesis was that the correlation coefficient obtained between a student's reported sex and the actual sex of the student is not significant. The computed coefficient of 0.788 was highly significant at the .001 level and the hypothesis was rejected. A chi-square of 202.47 with 1 degree of freedom was significant beyond the .001 level. A computer-generated contingency table of reported sex and actual sex is displayed in Table 4.

Table 3

A Computer-Generated Contingency Table of Students' Reported GPA and Actual GPA

RGPA	COUNT / ROW PCT / COL PCT / TOT PCT	AGPA 2.	3.	4.	5.	6.	7.	ROW TOTAL
2.		0	0	1	0	0	0	1
		0.0	0.0	100.0	0.0	0.0	0.0	0.3
		0.0	0.0	1.2	0.0	0.0	0.0	
		0.0	0.0	0.3	0.0	0.0	0.0	
3.		1	9	2	0	1	0	13
		7.7	69.2	15.4	0.0	7.7	0.0	3.8
		50.0	27.3	2.4	0.0	1.2	0.0	
		0.3	2.6	0.6	0.0	0.3	0.0	
4.		0	16	36	5	6	1	64
		0.0	25.0	56.3	7.8	9.4	1.6	18.7
		0.0	48.5	43.9	5.7	7.1	1.9	
		0.0	4.7	10.5	1.5	1.7	0.3	
5.		1	2	33	61	2	5	104
		1.0	1.9	31.7	58.7	1.9	4.8	30.3
		50.0	6.1	40.2	69.3	2.4	9.3	
		0.3	0.6	9.6	17.8	0.6	1.5	
6.		0	5	7	17	64	5	98
		0.0	5.1	7.1	17.3	65.3	5.1	28.6
		0.0	15.2	8.5	19.3	76.2	9.3	
		0.0	1.5	2.0	5.0	18.7	1.5	
7.		0	1	3	5	11	43	63
		0.0	1.6	4.8	7.9	17.5	68.3	18.4
		0.0	3.0	3.7	5.7	13.1	79.6	
		0.0	0.3	0.9	1.5	3.2	12.5	
COLUMN TOTAL		2	33	82	88	84	54	343
		0.6	9.6	23.9	25.7	24.5	15.7	100.0

Chi Square = 443.02 with 25 df

Significance > 0.001 level

N = 343

(Number of Missing observations = 31)

Pearson r = 0.71

Significance = 0.001

Table 4

A Computer-Generated Contingency Table of Students' Reported Sex and Actual Sex

	ASEX	0	1	
RSEX				
	0	155	18	173
		89.6	10.4	52.3
		90.1	11.3	
		46.8	5.4	
	1	17	141	158
		10.8	89.2	47.7
		9.9	88.7	
		5.1	42.6	
	Column Total	172	159	331
		52.0	48.0	100.0

Chi Square = 202.47 with 1 df

Significance > 0.001 level

N = 331

(Number of Missing observations = 43)

Pearson r = 0.79

Significance = 0.001

Statement of the seventh hypothesis was that the correlation coefficient between a student's reported class level and actual class level is not significant. The computed coefficient of 0.612 was significant at the 0.001 level and the hypothesis was rejected. A chi-square of 340.69 with 12 degrees of freedom was significant beyond the 0.001 level. A computer-generated contingency table of the reported class level and actual class level can be seen in Table 5.

The eighth hypothesis was that the correlation coefficient between a student's expected grade and grade assigned is not significant. The computed coefficient of 0.61 was significant at the 0.001 level and the hypothesis was rejected. A chi-square of 229.44 with 12 degrees of freedom was significant at the 0.001 level. A computer-generated contingency table of the expected grade and actual grade is displayed in Table 6.

Statement of the ninth hypothesis was that the correlation coefficient obtained between a student's expected grade and the student's rating of his instructor's performance is not significant. The computed coefficient of -0.182 was significant at the 0.001 level; the hypothesis was rejected.

The tenth hypothesis was that the correlation coefficient obtained between a student's achievement in a biology class and the student's rating of his instructor's performance is not significant. The computed coefficient of 0.136 was significant at the 0.007 level and the hypothesis was rejected.

Table 5

A Computer-Generated Contingency Table of Students' Reported Class Level and Actual Class Level

LEVEL	COUNT / ROW PCT / COL PCT / TOT PCT	ACLASS 1.	2.	3.	4.	5.	ROW TOTAL
1.		220	8	2	2	3	235
		93.6	3.4	0.9	0.9	1.3	68.9
		86.3	15.4	10.0	22.2	60.0	
		64.5	2.3	0.6	0.6	0.9	
2.		27	33	0	1	1	62
		43.5	53.2	0.0	1.6	1.6	18.2
		10.6	63.5	0.0	11.1	20.0	
		7.9	9.7	0.0	0.3	0.3	
3.		7	11	13	0	1	32
		21.9	34.4	40.6	0.0	3.1	9.4
		2.7	21.2	65.0	0.0	20.0	
		2.1	3.2	3.8	0.0	0.3	
4.		1	0	5	6	0	12
		8.3	0.0	41.7	50.0	0.0	3.5
		0.4	0.0	25.0	66.7	0.0	
		0.3	0.0	1.5	1.8	0.0	
COLUMN TOTAL		255	52	20	9	5	341
		74.8	15.2	5.9	2.6	1.5	100.0

Chi Square = 340.69 with 12 df

Significance > 0.001 level

N = 341

(Number of Missing observations = 33)

Pearson 4 = 0.61

Significance = 0.001

Table 6

A Computer-Generated Contingency Table of Students' Expected Grade and Grade Assigned

EGRADE	COUNT / ROW PCT / COL PCT / TOT PCT	AGRADE 0.	1.	2.	3.	4.	ROW TOTAL
0.		33	24	5	1	0	63
		52.4	38.1	7.9	1.6	0.0	17.9
		76.7	16.4	3.4	6.3	0.0	
		9.4	6.8	1.4	0.3	0.0	
1.		9	96	51	1	0	157
		5.7	61.1	32.5	0.6	0.0	44.7
		20.9	65.8	35.2	6.3	0.0	
		2.6	27.4	14.5	0.3	0.0	
2.		0	24	83	8	1	116
		0.0	20.7	71.6	6.9	0.9	33.0
		0.0	16.4	57.2	50.0	100.0	
		0.0	6.8	23.6	2.3	0.3	
3.		1	2	6	6	0	15
		6.7	13.3	40.0	40.0	0.0	4.3
		2.3	1.4	4.1	37.5	0.0	
		0.3	0.6	1.7	1.7	0.0	
COLUMN TOTAL		43	146	145	16	1	351
		12.3	41.6	41.3	4.6	0.3	100.0

Chi Square = 229.44 with 12 df

Significance > 0.001 level

N = 351

(Number of Missing observations = 23)

Pearson r = 0.61

Significance = 0.001

Statement of the eleventh hypothesis was that the correlation coefficient obtained between a student's rating of his biology instructor's performance on 20 specific items and the student's global rating of the instructor's performance is not significant. The computed correlation coefficient of 0.592 was high significant at the 0.001 level; the hypothesis was rejected.

The twelfth hypothesis was that the correlation coefficient obtained between a student's attitude toward biology and his achievement in a biology class is not significant. The computed correlation coefficient of -0.272 was significant at the 0.001 level and the hypothesis was rejected. All of the correlation coefficients are conveniently displayed in Table 7.

Statement of the thirteenth hypothesis was that there was no relationship between a student's reaction to class size and the student's rating of his biology instructor's performance. A one-way analysis of variance revealed a computed F of 3.294; the level of probability is .021. The null hypothesis was rejected. Table 8 gives a summary of the analysis of variance, subgroups and their means, results of tests for homogeneity of variances, and the results of Scheffe's procedure for detecting where differences occur.

The fourteenth hypothesis was that there was no relationship between a student's perception of the course work load and the student's rating of his biology instructor's performance. An analysis of variance showed a computed F of 1.94; the level of probability is .102. The null hypothesis was not rejected. Table 9 gives a summary of the analysis

Table 7

A Summary of the Correlation Coefficients Corresponding to Hypotheses 5-12

Hypothesis Concerning:	Hypothesis	Pearson r	Significance
Reported GPA and Actual GPA	5	0.704	0.001
Reported Sex and Actual Sex	6	0.788	0.001
Reported Class Level and Actual Class Level	7	0.612	0.001
Expect Grade and Actual Grade	8	0.607	0.001
Expected Grade and Instructor Rating	9	-0.185	0.001
Achievement and Instructor Rating	10	0.135	0.007
Instructor Rating and Global Rating	11	0.592	0.001
Attitude toward Biology and Achievement	12	-0.272	0.001

Table 8

Analysis of Variance Summary: Test for Relationship between Class Size and Student's Rating of the Biology Instructor's Performance; Tests for Homogeneity of Variances and Results of Scheffe's Procedure

Source	df	ss	ms	F	F Prob.
Between groups	3	827.56	275.86	3.29	0.021
Within groups	307	25707.06	83.74		
Total	310	26534.63			

N = 311

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

Group	Count	Mean	SD
1 = yes, most of the time	158	56.84	8.35
2 = no, class was too large	45	56.40	7.83
3 = no, class was too small	3	54.67	5.86
4 = it didn't make any difference one way or the other	105	53.31	10.74

Test for Homogeneity of Variances

Barlett-Box F = 3.486 P = 0.015

Maximum Variance/Minimum Variance = 3.362

Scheffe's Procedure

Ranges for the 0.050 level 3.97 3.97 3.97

Homogeneous Subsets*

Subset 1:	Group	4	3	2
	Mean	53.30	54.67	56.40
Subset 2:	Group	3	2	1
	Mean	54.67	56.40	56.84

*Subsets of groups—number of pairs of which have means that differ by more than the shortest significant range for a subset of that kind.

Table 9

Analysis of Variance Summary: Test for Relationship between a Student's Perception of the Course Work Load and the Student's Rating of His Biology Instructor's Performance

Source	df	ss	ms	F	F Prob.
Between groups	4	657.88	164.47	1.94	0.102
Within groups	324	27482.81	84.82		
Total	328	28140.69			

N = 329

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

Group (work load)	Count	Mean	SD
1 = much lighter	16	56.81	7.49
2 = lighter	14	55.17	9.38
3 = about right	171	56.23	9.11
4 = heavier	82	55.57	9.40
5 = much heavier	14	49.29	10.46

Test for Homogeneity of Variances

Barlett-Box F = 0.423 P = 0.794

Maximum Variance/Minimum Variance = 1.949

Scheffe's Procedure

Ranges for the 0.050 level 4.37 4.37 4.37

Homogeneous Subsets*

Subset 1:	Group	5	2	4	3	1
	Mean	49.29	55.17	55.57	56.23	56.81

*Subsets of groups—number of pairs of which have means that differ by more than the shortest significant range for a subset of that kind.

of variance, subgroups and their means, results of tests for homogeneity of variances and the results of Scheffe's procedure for detecting where differences occur.

The fifteenth hypothesis was that there was no relationship between a student's perception of the course difficulty and the student's rating of his biology instructor's performance. An analysis of variance revealed a computed F of 4.083. The probability level is .003. The null hypothesis was rejected. Table 10 gives a summary of the analysis of variance, subgroups and their means, results of tests for homogenity of variances, and the results of Scheffe's procedure for detecting where differences occur.

Statement of the sixteenth hypothesis was that there is no difference in the achievement of students with a positive attitude toward biology and students with a negative attitude toward biology. A summary of the results are displayed in Table 11. The computed t between means was significant beyond the .001 level; the hypothesis was rejected.

The seventeenth hypothesis was that a positive or negative attitude toward biology has no effect on the way a student perceives the difficulty of the course. The results are summarized and displayed in Table 12. The computed t was 1.645 at the 0.05 level; therefore, the hypothesis was rejected.

Statement of the eighteenth hypothesis was that a high or low theoretical orientation has no effect on the way a student perceives the difficulty of a course. A summary of the results is displayed in Table 13. The computed t between means was -2.06; the two-tailed probability is .043. The hypothesis was rejected.

Table 10

Analysis of Variance Summary: Test for Relationship between a Student's Perception of the Difficulty of the Course and the Student's Rating of His Biology Instructor's Performance

Source	df	ss	ms	F	F Prob.
Between groups	4	1350.56	337.64	4.083	0.003
Within groups	324	26790.13	82.69		
Total	328	28140.69			

N = 329

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

Group (level of difficulty)	Count	Mean	SD
1 = very elementary	5	51.20	15.25
2 = somewhat elementary	27	54.04	11.63
3 = about right	149	57.69	8.36
4 = somewhat difficult	127	54.48	9.16
5 = very difficult	21	51.38	8.47

Test for Homogeneity of Variances

Barlett-Box F = 2.202 P = 0.065

Maximum Variance/Minimum Variance = 3.331

Scheffe's Procedure

Ranges for the 0.050 level 4.37 4.37 4.37 4.37

Homogeneous Subsets*

Subset 1:	Group	1	5	2	4	3
	Mean	51.20	51.38	54.04	54.48	57.70

*Subsets of groups—number of pairs of which have means that differ by more than the shortest significant range for a subset of that kind.

Table 11

A t-Test Summary: Test for a Relationship between Attitude toward Biology (2 Levels) and Total Points Earned (Achievement)

Total Points Earned (Achievement)	Count	Mean	SD	df	t*
Group 1 (positive att.)	154	253.64	25.43	328	5.11
Group 2 (negative att.)	176	239.98	23.10		

N = 329

*Pooled Variance Estimate

2-Tailed Probability > 0.001

Table 12

A t-Test Summary: Test for a Relationship between Attitude toward Biology (2 Levels) and Perception of the Level of Difficulty of a Biology Course

Level of Difficulty	Count	Mean	SD	df	t*
Group 1 (positive att.)	153	3.22	0.76	327	-3.91
Group 2 (negative att.)	176	3.55	0.78		

N = 330

*2-Tailed Probability > 0.001

Table 13

A t-Test Summary: Test for a Relationship between Theoretical Orientation and Perception of the Level of Difficulty of a Biology Course

Level of Difficulty	Count	Mean	SD	df	t*
Group 1 (high TO)	112	3.28	0.81	327	-2.06
Group 2 (low TO)	217	3.46	0.78		

N = 329

*Pooled Variance Estimate

2-Tailed Probability = 0.040

TO = Theoretical Orientation

II. DISCUSSION OF THE RESULTS

The results of this study indicated that:

1. A student's attitude toward biology influenced the rating of his biology instructor's performance; students with a more positive attitude gave higher ratings.
2. A student's theoretical orientation, i.e., the level of cognitive development, influenced the rating of his biology instructor's performance; students with a high theoretical orientation gave higher ratings.
3. Students who included chemistry in their high school program rated their biology instructor's performance no differently than students who did not include chemistry in their high school program.
4. Students have the ability to differentiate among their instructors in a multisection biology course.
5. Students in freshman biology classes reported their cumulative grade point average with a high degree of accuracy.
6. Students in freshman biology classes reported their sex with a high degree of accuracy.
7. Students in freshman biology classes reported their class level with a high degree of accuracy.
8. The grade a student expected in a freshman biology class corresponded closely with the grade assigned by the instructor.

9. The grade a student expected in a freshman biology class had very little influence on the rating of his instructor's performance.
10. There was a weak positive correspondence between a student's achievement in a biology course and the rating of his instructor's performance.
11. A student's rating of biology instruction on 20 specific performance items and the student's global rating of performance corresponded highly.
12. There was a significant correspondence in a student's attitude toward biology and his achievement in a freshman biology course. Students with a more positive attitude achieved at a higher level.
13. The reaction of a student to class size influenced the rating of his biology instructor's performance. Students who thought the class size was too large gave the highest ratings; those who said class size "didn't matter" gave the lowest ratings.
14. The reaction of a student to class load corresponded with the rating of his biology instructor's performance. Students who perceived the work load as much heavier than in other courses gave much lower ratings than students who perceived the work load as being about right; students who perceived the work load as much lighter gave the highest ratings.
15. The reaction of a student to course difficulty corresponded with the rating of his biology instructor's performance.

Students who thought the difficulty of the course was very elementary or very difficult gave the lowest ratings; the ones who perceived course difficulty as somewhat elementary or somewhat difficult gave higher ratings; students who thought the course difficulty as about right gave the highest ratings.

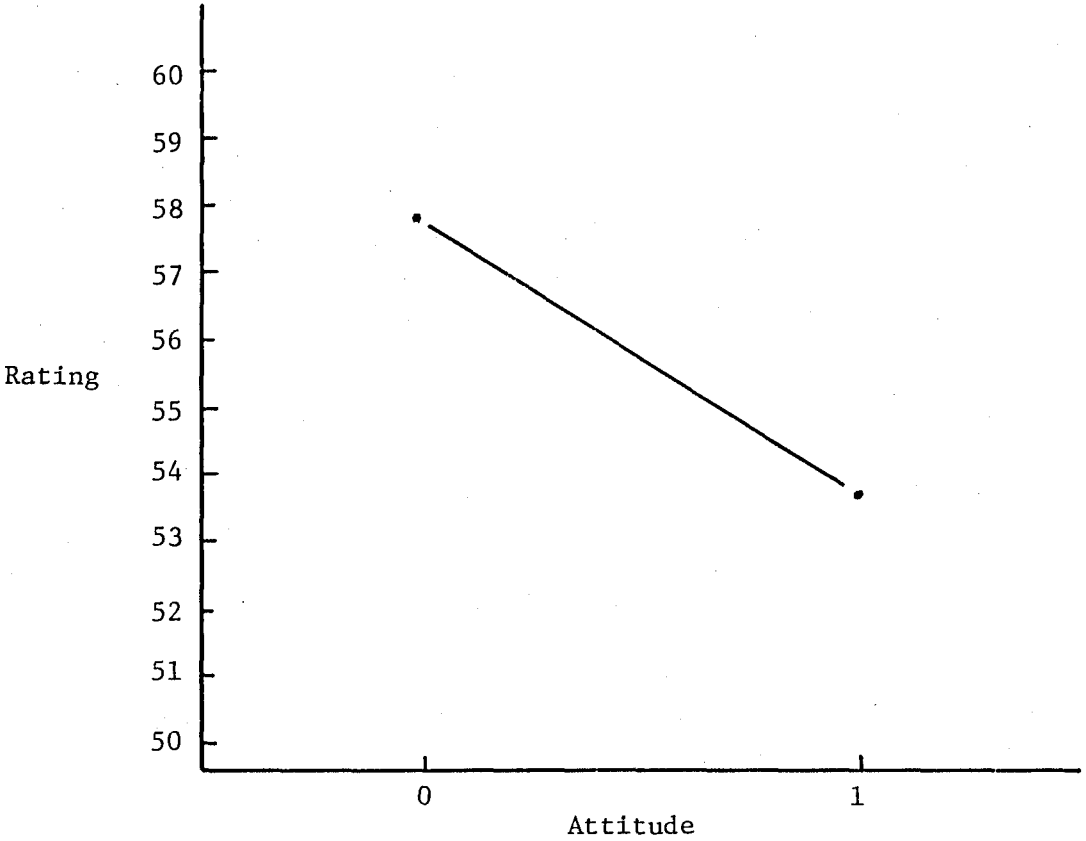
16. Students with a positive attitude toward biology achieved at a higher level than students with a positive attitude toward biology.
17. Students who had a positive attitude toward biology perceived the course as less difficult than students who had a negative attitude.
18. Students with a high theoretical orientation perceived the course as less difficult than students with a low theoretical orientation.
19. The interaction variability in the multianalysis of variance model was attributable to chance, i.e., there was no significant interaction in the eleven interaction terms.

Findings of this study lend themselves to three types of interpretation. First, more knowledge of the extent of relationships between instructor and course evaluation measures and various demographic, attitudinal and personality variables (in the context of a particular course and instructor) clarifies our general understanding of the evaluation process and points to its complexity. Second, knowledge of these relationships can provide departmental staff members with information necessary for a better "fit" of the curriculum to the student

population enrolling for these courses. Third, since future research can benefit from a narrowing of the field of study, results reported here should have implications for further such investigations.

Results presented above both confirm and disconfirm preconceptions held by the author. The relationship between ratings and student attitudes toward biology was expected. Significantly higher ratings were assigned by students who displayed a positive attitude toward biology; this indicated that attitude toward biology partitioned the student population into two distinct subpopulations. The probability of obtaining this difference in means for the two attitude groups is less than one in a thousand. Figure 3 is a graphic representation of this finding. It is interesting to note that the student population was about evenly divided—157 subjects were in the positive attitude group; 176 subjects were in the negative attitude group. This is good news for biology teachers. Figure 4 reveals more detailed analysis of the data. Students with a more positive attitude gave the higher ratings to all 3 instructors; however, the order in which they ranked the instructors as well as the difference between instructors was virtually identical to rankings given by the negative attitude group. In other words, all students, regardless of attitude agreed on the relative position of each instructor in the ranking order. More impressive was the agreement in the rating group with regard to high, low, or average ratings given to the three instructors. The lowest mean rating was given by the negative attitude group to instructor B.

The results clearly demonstrate the relationship between the students' theoretical orientation and ratings of instructional performance.



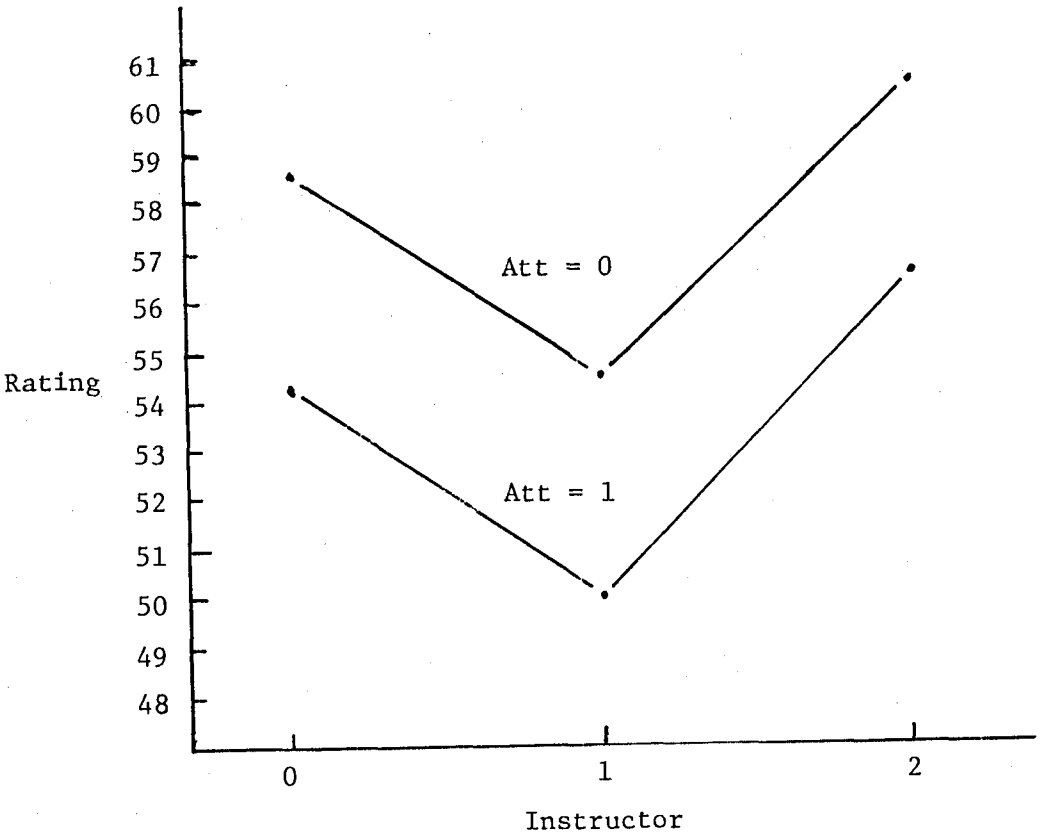
Means

Att	N	Rating
0	157	57.94
1	176	53.74

Code

0 = positive attitude
1 = negative attitude

Figure 3. A graphic representation of the relationship between students' attitudes toward biology and their ratings of instructional performance.



Means

Att	Instr	N	Rating
0	0	29	58.621
0	1	58	54.586
0	2	67	60.537
1	0	33	54.303
1	1	65	49.984
1	2	78	56.628

Code

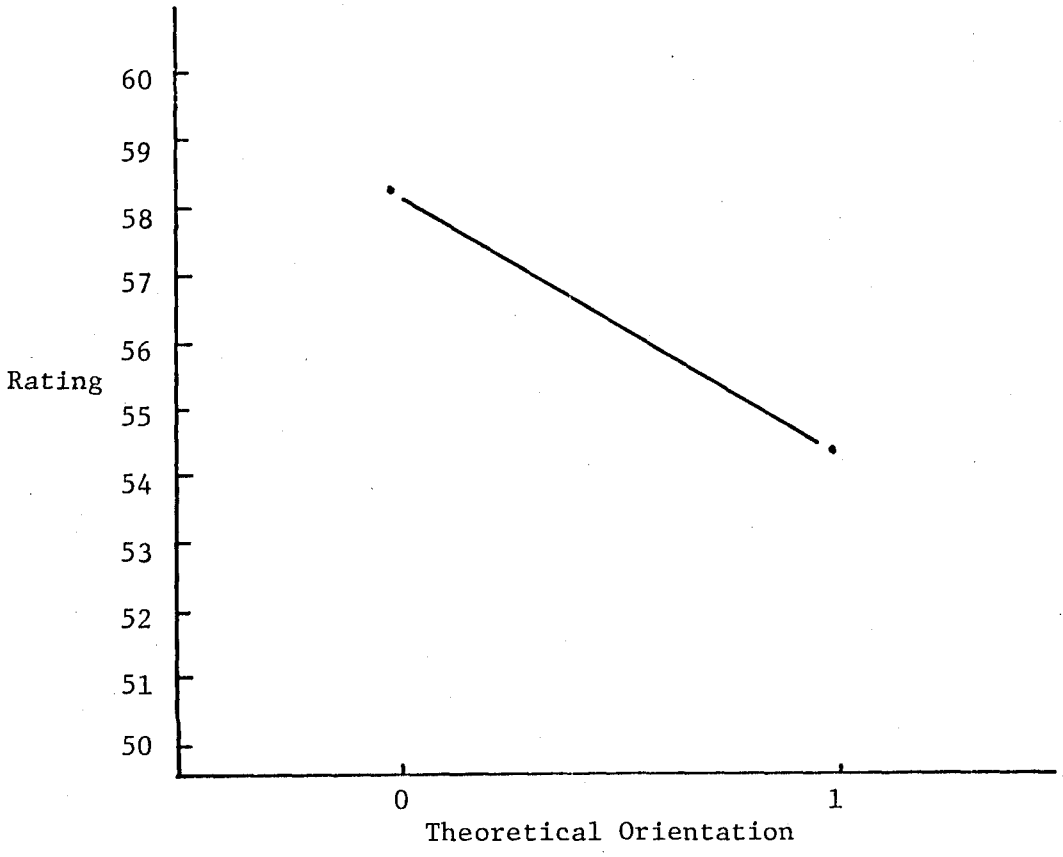
Att = attitude toward biology
0 = a positive attitude toward biology
1 = a negative attitude toward biology

0 = Instructor A
1 = Instructor B
2 = Instructor C

Figure 4. The relationship between the ratings of students with a positive attitude toward biology and students with a negative attitude for three biology instructors.

As expected, students with a high theoretical orientation gave the highest mean ratings to all three instructors. For the low theoretical orientation group instructor C received the highest rating. Again instructor B received the lowest mean ratings from both groups. Figures 5, 6, and 7 display the results. It may be inferred from the data that a student's theoretical orientation influences his rating of instruction in freshman biology courses. These results tend to corroborate the results of a related study by McKeachie (1969b).

The nearly zero relationship found between ratings and the students' chemistry background in high school was not expected. It was assumed that most students who selected chemistry would be in the high theoretical orientation group. However, this was not the case. Only 70 of the 209 students who had high school chemistry were in the high theoretical orientation group; the remaining 139 were in the low theoretical orientation group. It could be that high school counselors, in an effort to "prepare students for college," are pushing courses in science for all students regardless of ability. Also it must be considered that the students in the high theoretical group who did not have high school chemistry (n-43) were students who avoided chemistry in an attempt to have a higher ranking in the graduation class or who aspired to become valedictorian of their class. The outcome of the relationship between the chemistry variable and the theoretical orientation variable may be viewed more clearly in light of the above discussion. This relationship is the only interaction term in the 4-way analysis of variance which approached the .05 significance level set for this study



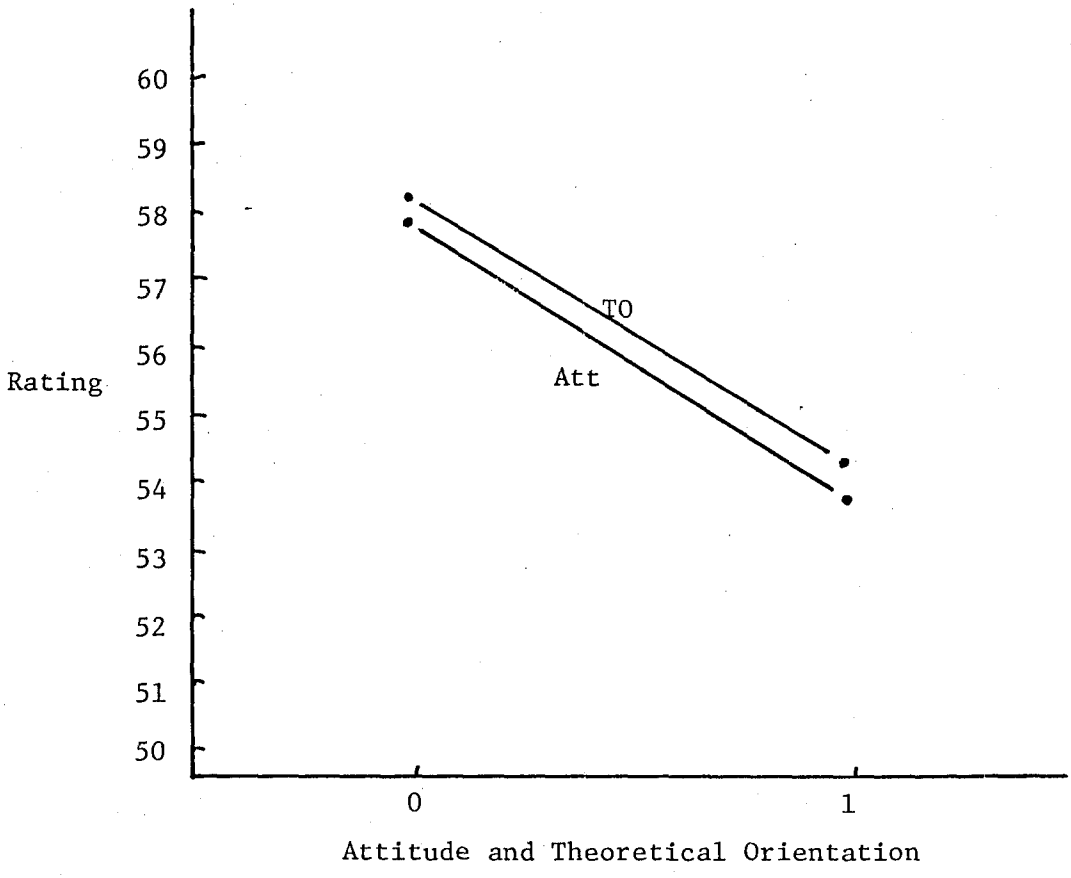
Means

TO	N	Rating
0	113	58.12
1	217	54.43

Code

0 = high theoretical orientation
1 = low theoretical orientation

Figure 5. A graphic representation of the relationship between students' theoretical orientations and ratings of instructional performance.



Code

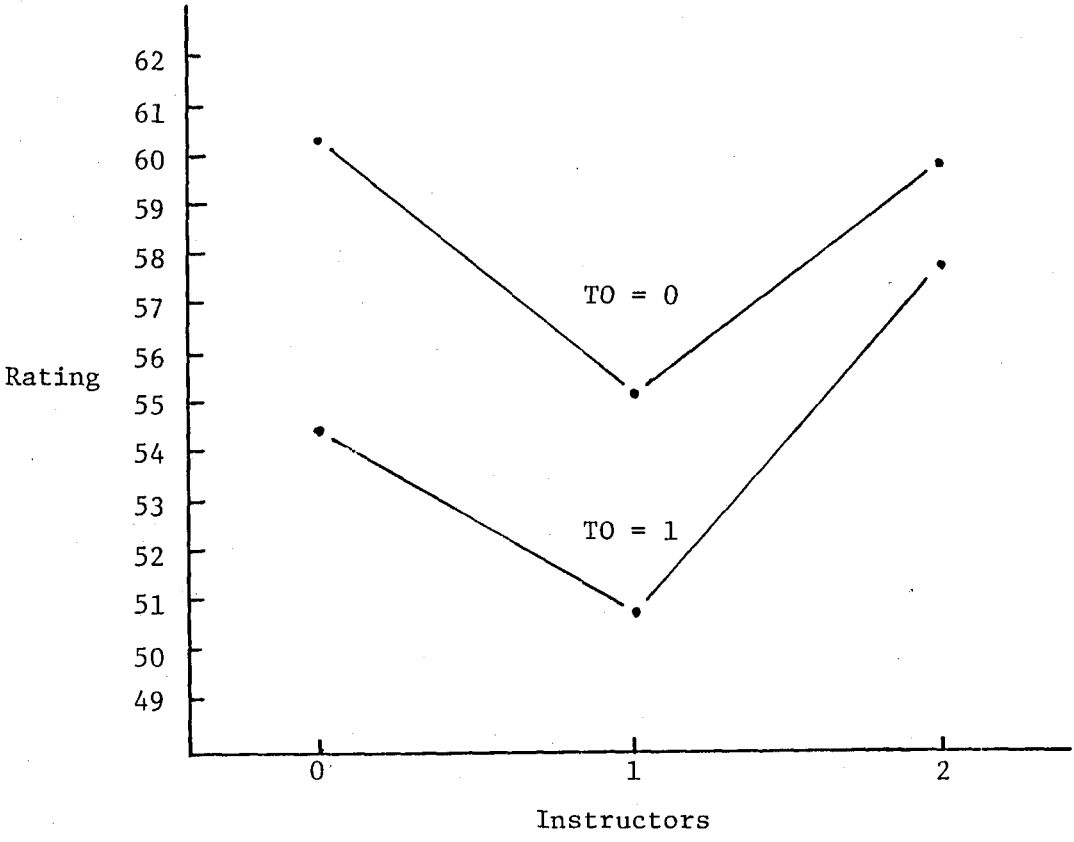
Att

0 = positive attitude
1 = negative attitude

TO

0 = high theoretical orientation
1 = low theoretical orientation

Figure 6. Figure 5 superimposed on Figure 3.



Means

TO	Instr	N	Rating
0	0	20	60.25
0	1	42	55.16
0	2	51	59.72
1	0	42	54.45
1	1	81	50.59
1	2	94	57.73

Code TO

0 = high theoretical orientation
1 = low theoretical orientation

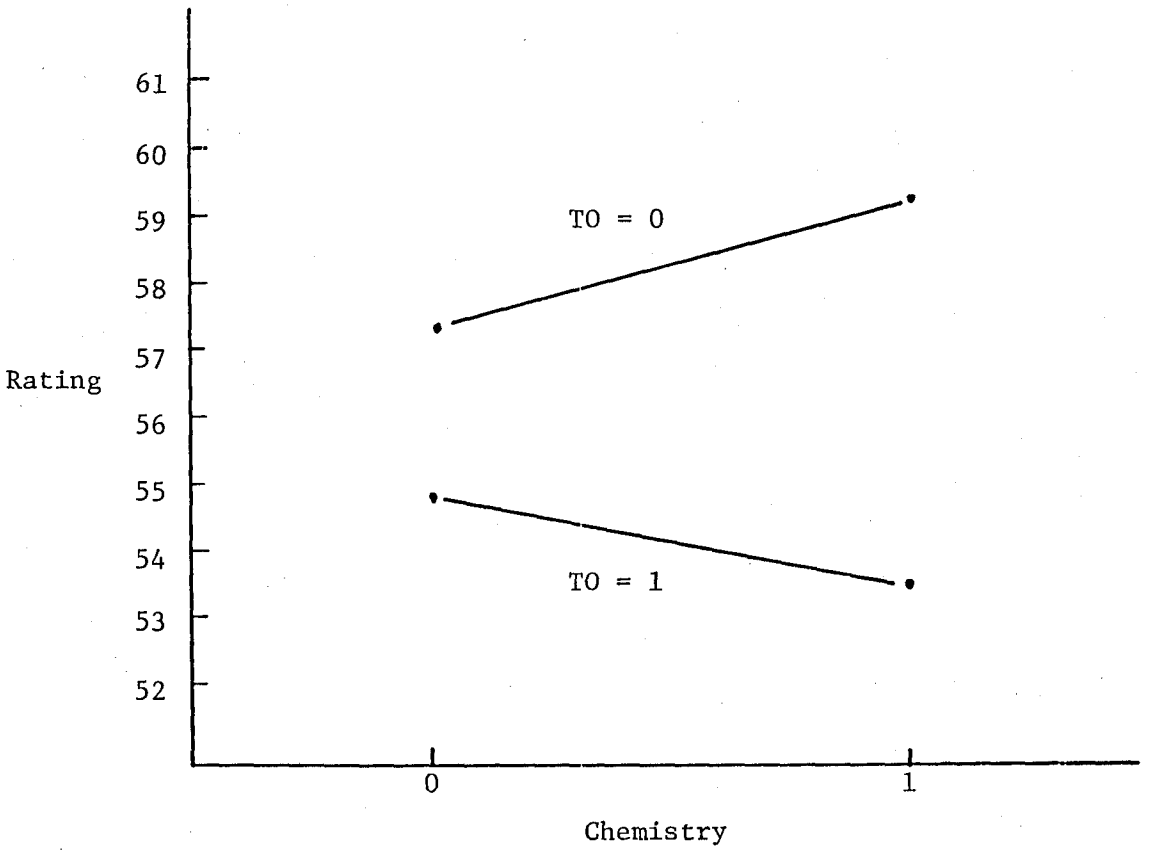
Figure 7. A graphic representation of the relationship between the ratings of students with a high theoretical orientation and students with a low theoretical orientation of three biology instructors.

($p = .07$). The results of this unexpected relationship is graphically displayed in Figure 8; students with a high theoretical orientation who did not have high school chemistry gave the highest ratings while students with a low theoretical orientation who did not have high school chemistry gave the lowest ratings.

The fourth main effect, the instructor factor, showed the strongest relationship to the dependent variable, student ratings. Results represent strong support for the contention that students can clearly discriminate among the teaching performance of different instructors. Graphic representation of these results is displayed in Figure 9. A difference as large as that observed would be expected on a chance basis less than one in ten thousand times.

An analysis of covariance using math and English ACT scores to adjust for differences in academic abilities generated a suprising result—it had virtually no effect on the dependent variable, i.e., mean ratings of students with higher abilities did not differ from mean ratings of students with lower abilities. Table 14 shows the means for each group of the 4 main effects both before and after adjusting for the abilities variable. The before and after adjustment means for the instructor factor are practically identical. These data indicate that ratings did not systematically covary with ACT test scores in math and English.

A detailed analysis of the hypotheses designed to determine how accurately students report information about themselves is shown in Tables 15, 16, and 17. These data reveal that students report information about themselves with a high degree of accuracy.



Means

Chem	TO	N	Rating
0	0	70	57.44
0	1	139	54.89
1	0	43	59.23
1	1	78	53.62

Code

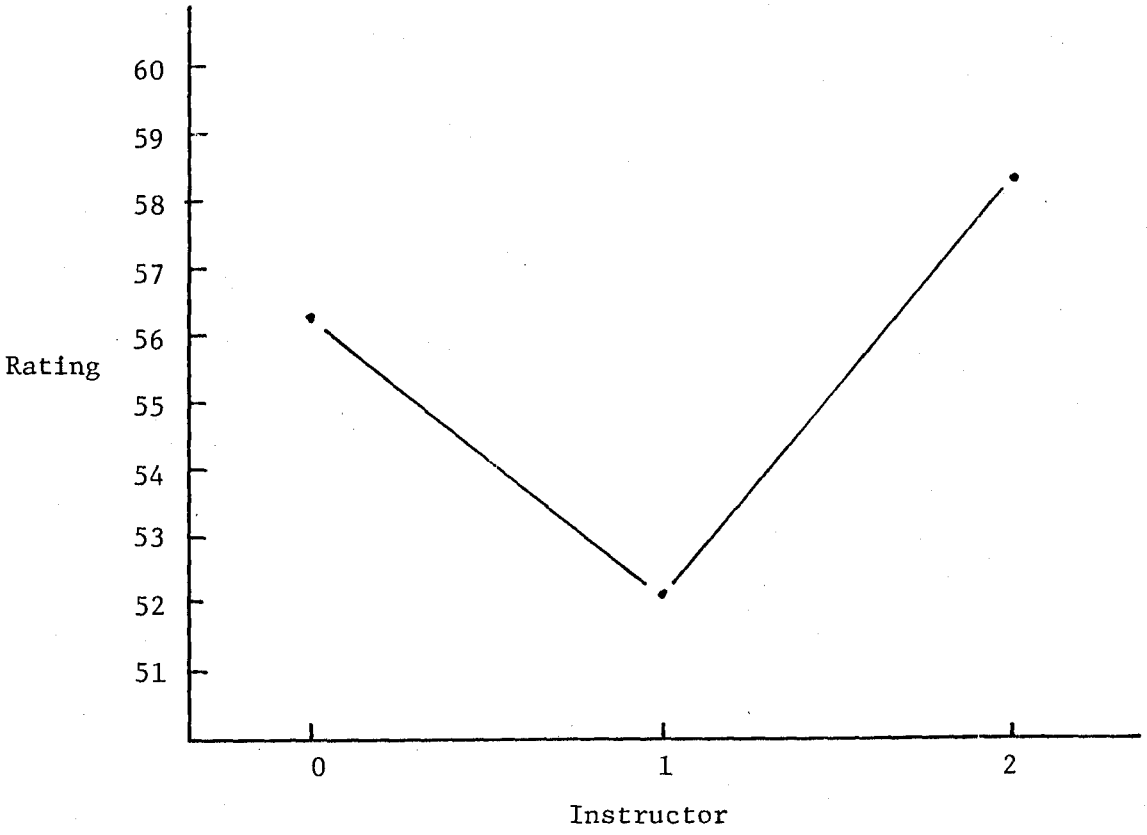
TO

0 = high theoretical orientation
1 = low theoretical orientation

Chem

0 = no high school chemistry
1 = high school chemistry

Figure 8. The relationship between chemistry background in high school, theoretical orientation and the instructor rating.



Means

Instr	N	Rating
0	62	56.32
1	123	52.15
2	145	58.43

Code

0 = Instructor A
1 = Instructor B
2 = Instructor C

Figure 9. The relationship between the instructor variable and the dependent variable, student ratings of instructor performance.

Table 14

Analysis of Covariance for the 4-Way Analysis of Variance

	Means Before Adjusting -Rating-	Means After Adjusting -Rating-
Attitude		
Positive	57.4597	57.5719
Negative	53.7324	53.6344
Theoretical Orientation		
High	57.5341	57.6956
Low	54.4494	54.3696
Instructor		
A	56.2340	56.2573
B	51.9010	51.9037
C	58.2203	58.2087

Table 15

Reported Sex and Actual Sex

296 responded correctly—approx. 89.4%
35 responded incorrectly—approx. 10.5%

17 females said they were males
18 males said they were females

52.3% of responding population were males

47.7% of responding population were females

r = 0.7882

*Forty-three students failed to respond to this question.

Table 16

Reported Class Level and Actual Class Level

272 students (79.8%) reported accurately
Only 17 students (5%) report deviated from actual level by more than one class level

8 freshmen reported they were sophomores
2 freshment reported they were juniors
2 freshmen reported they were seniors
3 freshmen reported they were graduate students

27 reported they were sophomores when records showed freshmen
0 reported they were sophomores when records showed juniors
1 reported they were sophomores when records showed seniors
1 reported they were sophomores when records showed graduate student

7 reported they were juniors when records showed freshmen
11 reported they were juniors when records showed sophomores
0 reported they were juniors when records showed seniors
1 reported they were juniors when records showed graduate student

1 reported they were senior when records showed freshman
0 reported they were seniors when records showed sophomores
5 reported they were seniors when records showed juniors
0 reported they were seniors when records showed graduate students

$r = 0.61$

Table 17

Reported GPA and Actual GPA

62.1% reported GPA accurately
37.9% reported GPA inaccurately

1 reported 1.00-1.49 actually had 2.00-2.49

1 reported 1.50-1.99 actually had 0.50-0.99
2 reported 1.50-1.99 actually had 2.00-2.49
1 reported 1.50-1.99 actually had 3.00-3.49

16 reported 2.00-2.49 actually had 1.50-1.99
5 reported 2.00-2.49 actually had 2.50-2.99
6 reported 2.00-2.49 actually had 3.00-3.49
1 reported 2.00-2.49 actually had 3.50-4.00

1 reported 2.50-2.99 actually had 1.00-1.49
2 reported 2.50-2.99 actually had 1.50-1.99
33 reported 2.50-2.99 actually had 2.00-2.49
2 reported 2.50-2.99 actually had 3.00-3.49
5 reported 2.40-2.99 actually had 3.50-4.00

5 reported 3.00-3.49 actually had 1.50-1.99
7 reported 3.00-3.49 actually had 2.00-2.49
17 reported 3.00-3.49 actually had 2.50-2.99
5 reported 3.00-3.49 actually had 3.50-4.00

1 reported 3.50-3.99 actually had 1.50-1.99
3 reported 3.50-3.99 actually had 2.00-2.49
5 reported 3.50-3.99 actually had 2.50-2.99
11 reported 3.50-3.99 actually had 3.00-3.49

28 reported lower than actual
14 of the 28 were off 1 block

85 reported higher than actual
78 of the 85 were off 1 block

$r = 0.7042$

The most interesting result concerned the correspondence between actual GPA and reported GPA. Approximately 62 percent of the students' reported GPA corresponded exactly to the actual GPA. Another 22 percent (92 students) were off only one block in either direction. Tables 3, 4, and 5 (pages 59, 60, and 62) contain computer-generated contingency tables for these variables. Tables 15, 16 and 17 contain more detailed information from the contingency tables in Tables 3, 4, and 5.

In examining whether students report their sex (male or female) accurately there were some interesting results. Seventeen females said they were males and 18 males said they were females. Forty-three students failed to respond to the question, and 52.3 percent of the responding population were males.

It was found that students in the study were able to predict their final grade with a high degree of accuracy. A contingency table for these factors appears in Table 6, page 63. Table 18 contains a detailed breakdown of grade expectancies. Only 11 of the 344 students (5 percent) were off over one letter grade in predicting their actual grade. A very weak negative correlation (-.18) was obtained when students' expected grades were correlated with the rating they gave their instructor. It appears that these students were well aware of their real progress in the course but that it had very little influence in how they rated their instructor. Students with high grades had a slight tendency to give lower ratings.

Expected results were obtained when the factors of class size, class load and course difficulty were investigated in relationship to student ratings. Analyses are displayed in Figures 10, 11 and 12. It is

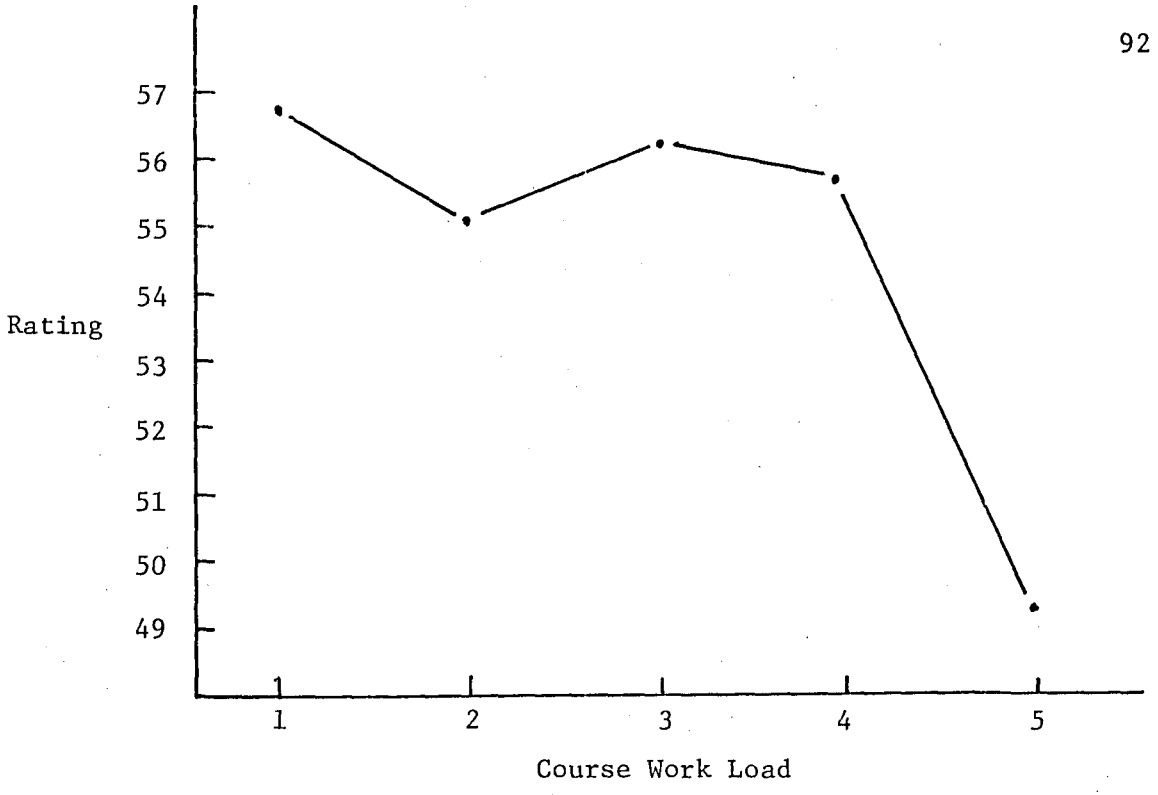
Table 18

Expected Grade and Actual Grade

218	got what they expected	62%
42	got better than they expected	12%
91	got lower than they expected	26%
24	expected an A but received a B	
5	expected an A but received a C	
1	expected an A but received a D	
9	expected a B but received an A	
51	expected a B but received a C	
1	expected a B but received a D	
24	expected a C but received a B	
8	expected a C but received a C	
1	expected a C but received an F	
1	expected a D but received an A	
2	expected a D but received a B	
6	expected a D but received a C	

$r = 0.607$

*Data extracted from the contingency table in Table 6, page 63.



Question to which students responded:
The work load for this course in relation to other courses of equal credit was:

Group	Count	Mean
1 = much lighter	16	56.81
2 = lighter	46	55.17
3 = about the same	171	56.23
4 = heavier	82	55.72
5 = much heavier	14	49.29

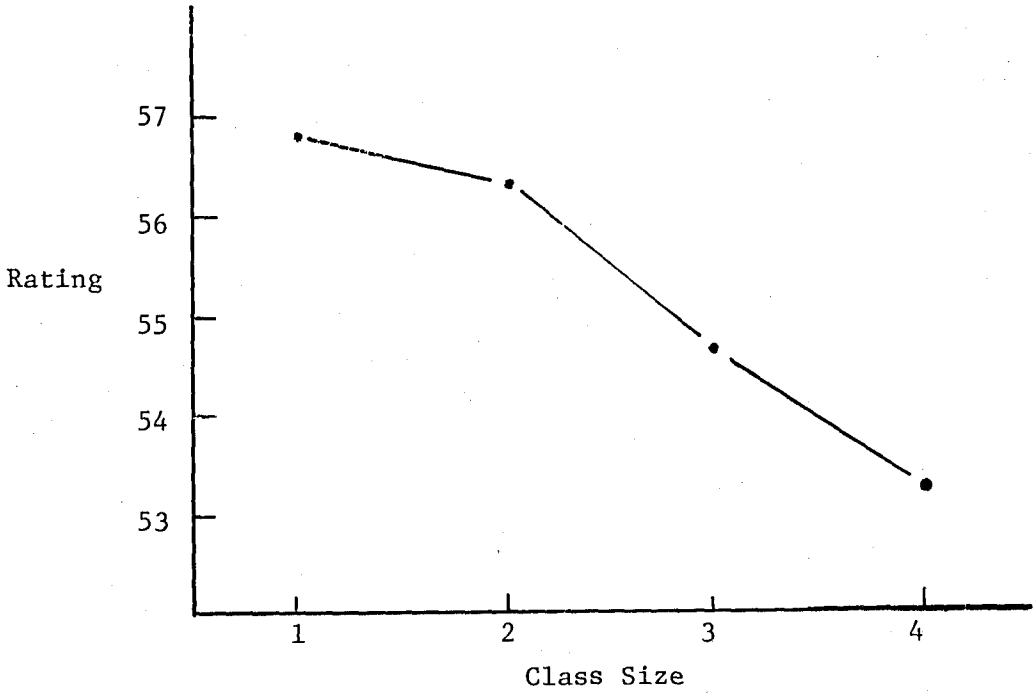
Test of homogeneity of variances
Barlett-Box F P = .794

Multiple range test
Scheffe Procedure (ranges for the .05 level) Ranges—4.37

Homogeneous subset

Group	5	2	4	3	1
Mean	49.29	55.17	55.57	56.23	56.81

Figure 10. Course work load and instructor's rating.



Question to which students responded:
Was class size satisfactory for the method of conducting the class?

Group	Count	Mean
1 = Yes, most of the time	158	56.84
2 = No, class was too large	45	56.40
3 = No, class was too small	3	54.67
4 = It didn't make any difference one way or another	105	53.30

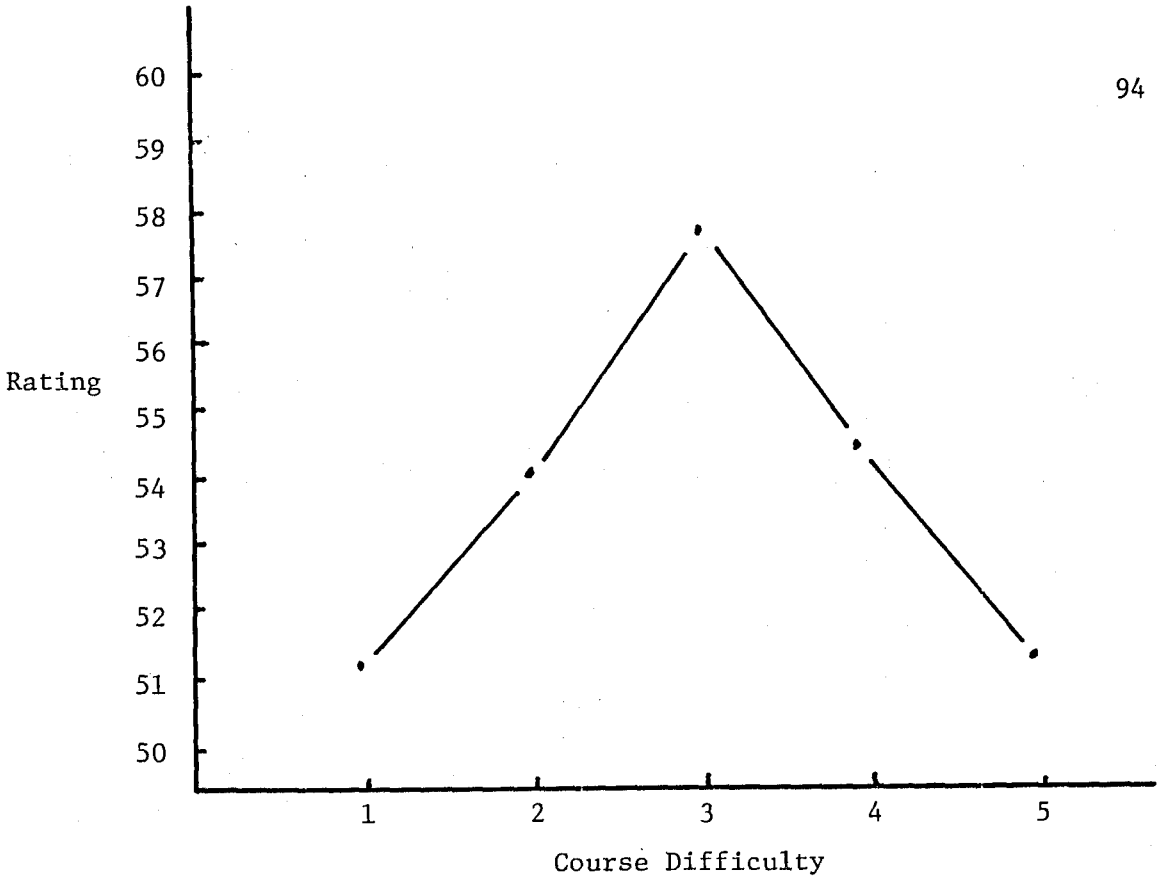
Test of homogeneity of variances
Barlett-Box F P = 0.015

Multiple range test
Scheffe Procedure (ranges for the .05 level) Ranges—3.97

Homogeneous subset

Subset 1:	Group	4	3	2
	Mean	53.30	54.67	56.40
Subset 2:	Group	3	2	1
	Mean	54.67	56.40	56.84

Figure 11. Class size and instructor's rating.



Question to which students responded:
For my preparation and ability, the level of difficulty of this course was:

Group	Count	Mean
1 = very elementary	5	51.20
2 = somewhat elementary	27	54.04
3 = about right	149	57.70
4 = somewhat difficult	127	54.48
5 = very difficult	21	51.38

Test of homogeneity of variances
Barlett-Box F P = 0.065

Multiple range test
Scheffe Procedure (ranges for the .05 level) Ranges—4.37

Homogeneous subset

Group	1	5	2	4	4
Mean	51.20	51.38	54.04	54.48	57.70

Figure 12. Course difficulty and instructor's rating.

interesting to note that students who perceived the course work load as much heavier than other courses gave the lowest ratings. Also students who thought the difficulty level to be very elementary or very difficult gave significantly lower ratings than students who thought the course difficulty to be "about right." A total of 149 students were in the about right category; a total of 127 students were in the "somewhat difficult" category.

The results of the validity study is presented in terms of the correlation between total exam points earned by students and instructional ratings. Math ACT scores were available for almost all the students so an adjusted (linear regression) total exam score for each student was used for correlational analysis. This regression procedure provides an improved measure of how much students learned in the course.

There was a weak positive correlation between total exam points and ratings (.135). The correlation between achievement residualized for ability and ratings was .205. An Eta (test for curvilinear tendency in the regression line) generated a correlation rato of .45 (the regression of achievement on rating, i.e., the regression of x on y). A computer-generated scattergram of these variables is shown in Figure 13. The scatter is clearly shown as being almost entirely restricted to the upper right quadrant. A possible explanation for this narrow range is that (1) students are allowed to drop a course as late as 4 weeks before the end of the quarter. This removes students from the class who are failing or those who wish to avoid a below average grade (2) students are "soft" on teachers—they generally rate them above the numerical average on all items (Centra 1972).

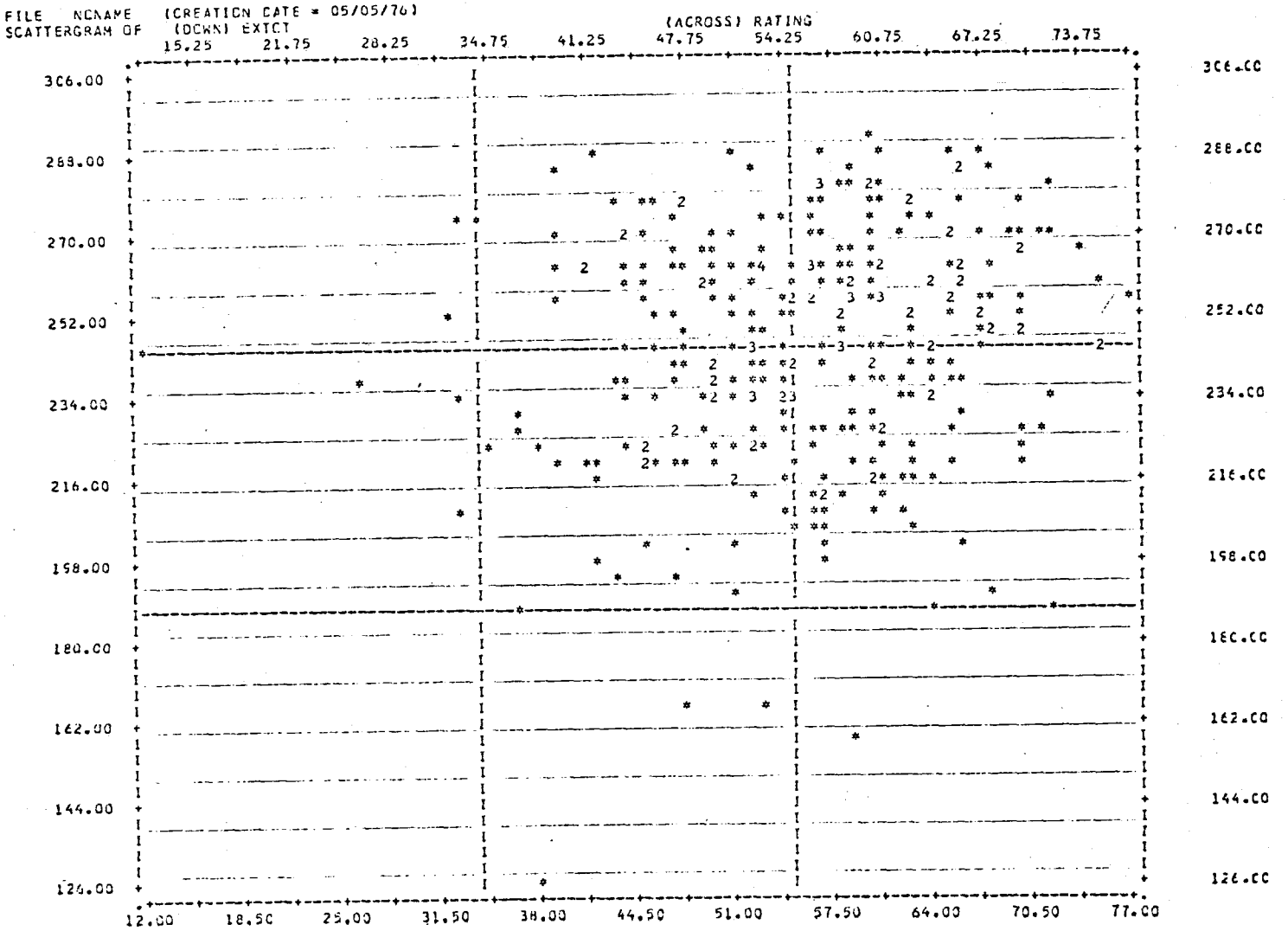


Figure 13. A computer-generated scattergram of the variables, total exam-points (achievement) and student ratings of instruction.

The importance of this finding should not be underestimated, particularly in relation to the widely publicized (and subsequently criticized) negative findings of Rodin and Rodin (1972). It provides additional support, albeit weak, for the contention that at least some information from student ratings is positively related to student achievement; a trend which must be substantiated if widespread use of student ratings for tenure, promotion, and/or instructional improvement is to be continued.

CHAPTER V

SUMMARY AND CONCLUSIONS

This study was designed to investigate the effects of three student characteristics on student ratings of instruction in biology courses at the college level. Included as variables in the study were the students' theoretical orientation, attitude toward biology, and background in chemistry courses in high school, as well as the course instructor. Multianalysis of variance was employed to examine the influence of the three student characteristics on students' ratings of instructorial or professorial performance and to ascertain whether students can differentiate among instructors.

A second purpose of the study was to learn more about the extent students report accurately such information as sex, cumulative grade point average, and class level. This was accomplished by the use of correlation analyses and chi square tests.

Correlation analyses were used to determine the correspondence of each student's expected grade to the grade assigned by the instructor and to learn more about the effects of a student's perceived progress on his rating of instruction.

One-way analyses of variance and correlational analyses were used to determine relationships between students' ratings of instruction and reactions to class size, course work load, and difficulty of the course.

A validity study of the Student Instructional Report (SIR) was attempted by use of correlation analyses to determine the relationship

between achievement and the rating; and between the rating responses computed from 20 specific items and a global rating.

By use of t-tests the relationships between attitude and achievement, course difficulty and theoretical orientation; and course difficulty and attitude were evaluated.

The basic population for the study was comprised of students enrolled in the third quarter of freshman biology classes during the spring quarter, 1975, at East Tennessee State University.

For the multivariate analysis of variance, the subjects were divided into two groups on the basis of their scores on a biology attitude scale. Two groups were established on the basis of chemistry courses taken in high school; two groups were formed on the basis of scores on the theoretical orientation scale of a personality inventory. Finally the group was divided into three groups identified by the instructor assigned to their section; this completed the $2 \times 2 \times 2 \times 3$ factorial design. The procedure of unweighted means analysis was used to analyze the problem of unequal n.

The results of the study indicated that: (1) students with a more positive attitude toward biology gave significantly higher ratings than those with a less positive (i.e., negative) attitude toward biology; (2) students with a high theoretical orientation gave significantly higher ratings than those with a low theoretical orientation; (3) students who included chemistry in their high school program gave similar ratings to those who did not include chemistry in their high school program; (4) students differentiated among their biology instructors' performances; (5) students reported with a high degree of accuracy such information as their GPA, sex, and class level; (6) expected grades of students had

virtually no influence on ratings of instructional performance; (7) the grades students expected corresponded closely to the grades assigned by the instructor; (8) there was a weak relationship between students' achievement in a biology course and the ratings of their biology instructors' performance on 20 specific performance items and the students' general ratings of their biology instructors' performance; (10) the reactions of students to class size, work load in the class, and course difficulty influenced the ratings of their biology instructors' performance.

Student ratings of undergraduate teaching fall far short of a complete assessment of an instructor's teaching contribution. Nevertheless, if teaching performance is to be evaluated, either for purposes of tenure, promotion, or for individual improvement, a systematic measure of student attitudes, opinions and observations can hardly be ignored. However, sufficient information needs to be accumulated in order to correct the ratings to allow for systematic effects that might penalize instructors at certain levels and in certain categories. In regard to student ratings of instructional performance the information base is still too incomplete to permit a fair and just application. Clearly, the influence of student characteristics on student ratings is a matter for further research.

LIST OF REFERENCES

LIST OF REFERENCES

Aleamoni, L. M. Typical faculty concern about student evaluation of instruction. Invited address presented at the Israel Institute of Technology, Haifa, Israel, March 24, 1974.

American Association of University Professors Committee C on Teaching, Research, and Publication. Statement of teaching evaluation. AAUP Bulletin, 1974, 60(2), 166-170.

Balvin, T. O., and Boedeker, R. R. Nonscience student attitudes toward science, physics and two introductory physical science courses. The Journal of College Science Teaching, 1975, 4 , 179-182.

Bendig, A. W. Relation of level of course achievement of students, instructors, and course ratings in introductory psychology. Educational and Psychological Measurement, 1953, 13, 437-488(a).

Bendig, A. W. Student achievement in introductory psychology and student ratings of the competence and empathy of their instructors. Journal of Psychology, 1953, 36, 427-433.

Bendig, A. W. Ability and personality characteristics of introductory psychology instructors rated competent and empathetic by the students. Journal of Educational Research, 1955, 48, 705-709.

Bennett, R. M. Effects of previous high school programs on achievement in college biology. The Journal of College Science Teaching, 1975, 4, 242-246.

Blum, M. L. An investigation of the relation existing between students' grades and their ratings of the instructors' ability to teach, Journal of Educational Psychology, 1936, 27, 217-221.

Borg, W. R. Personality and interest measures as related to criteria of instructor effectiveness. Journal of Educational Research, 1957, 50, 701-709.

Boyle, D. G. Students Guide to Piaget. New York: Pergamon Press, 1969.

Bresler, J. B. Teaching effectiveness and government awards. Science, 1968, 160, 164-167.

Brooks, D. W. and Levenson, H. Questions and answers about questions and answers. The Journal of Chemical Education, 1974, 51, 161-162.

Caffrey, B. Lack of bias in student evaluations of teachers. Proceedings of the 77th Annual Convention of the American Psychological Association, 1969, 4, 641-642.

Cameron, W. B. Mashed potatoes, oysters, and teacher-rating scales. Phi Kappa Phi Journal, 1976, 54, 33-36.

Centra, J. A. Michigan State University student instructional rating scale analysis. Unpublished report, Office of Institutional Research, Michigan State University, 1964.

Centra, J. A. Problems of evaluating teaching. Current Issues in Higher Education, Association for Higher Education, Washington, D.C., 1966.

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

Centra, J. A. The utility of student ratings for instructional improvement. Project Report 72-16. Princeton, N.J.: Educational Testing Service, 1972(b).

Centra, J. A. The student instructional report: Comparison with alumni ratings, item reliabilities, the factor structure, SIR Report Booklet No. 3. Princeton, N.J.: Educational Testing Service, 1973.

Centra, J. A. and Linn, Robert L. Student Points of View in Ratings of College Instruction. Research Bulletin (RB7360). Princeton, N.J.: Educational Testing Service, October, 1973.

Clark, K. E. and Keller, R. J. Student ratings of college teaching. In R. A. Eckert (Ed.) A university looks at its program. Minneapolis: University of Minnesota Press, 1954.

Cohen, J. and Humphreys, L. G. Memorandum to faculty. University of Illinois, Department of Psychology, 1960.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-286.

The College Board News, January, 1976, p. 5.

Cosgrove, D. C. Diagnostic rating of teacher performance. Journal of Educational Psychology, 1959, 50, 200-204.

Comparative Data for the Student Instructional Report. Princeton, N.J.: Educational Testing Service, 1973-1974.

Costin, F. Survey of opinions about lecturers. University of Illinois, Department of Psychology, 1968.

Costin, Frank, Greenough, William T. and Menges, Robert J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 41, No. 5, 1971.

Costin, F., Greenough, W. J. and Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1972, 41, 511-535.

Crannell, C. W. An experiment in the rating of instructors by their students. College and University, 1948, 24, 5-11.

Crawford, P. L. and Bradshaw, H. T. Perception of characteristics of effective university teachers: A scaling analysis. Educational and Psychological Measurement, 1968, 28, 1079-1085.

Crittenden, K. S. and Norr, J. L. Some remarks on "student ratings": the validity problem. American Education Research Journal, 1975, 12, 429-433.

Crittenden, K. S. and Norr, J. L. Student values and teacher evaluation: A problem in person perception. Sociometry, 1973, 36, 143-151.

Crittenden, K. S., Norr, J. L., and LeBailey, R. K. Size of university classes and student evaluations of teaching. Journal of Higher Education, 1975, 46, 461-470.

Drucker, A. J. and Remmers, H. H. Do alumni and students differ in their attitudes toward instructors? Purdue University Studies in Higher Education, 1950, 70, 62-64.

Drucker, A. J. and Remmers, H. H. Do alumni and students differ in their attitudes toward instructors? Journal of Educational Psychology, 1951, 42, 129-143.

Downie, W. W. Student evaluation of faculty. Journal of Higher Education, 1952, 23, 495-496, 503.

Doyle, K. O., and Whitely, Susan E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11, 259-274.

Eckert, R. E. Ways of evaluating college teaching. School and Society, 1950, 71, 65-69.

Elliot, D. H. Characteristics and relationships of various criteria of college and university teaching. Purdue University Studies in Higher Education, 1950, 70, 5-61.

French, G. M. College students' concepts of effective teaching determined by an analysis of teacher ratings. Dissertation Abstracts, 1957, 17, 1380-1381.

Frey, Peter W. The ongoing debate: Student evaluation of teaching. Change, February, 1974, 47-48.

Frey, P. W. Student evaluations. Technical comment. Science, 1975, 187, 558-559.

Frey, P. W. Student instructional ratings and faculty performance. Science, 1973, 182, 83-85.

Frey, P. W. Student ratings of reachings: Validity of several rating factors. Science, 1973, 182, 83-85.

Frey, P. W., Dale, W. L., and Beatty, W. W. Student rating of instruction: Validation research. American Educational Research Journal, 1975, 12, 435-447.

Gage, N. L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.

Garverick, C. S. and Carter, H. D. Instructor ratings and expected grades. California Journal of Educational Research, 1962, 13, 218-221.

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

Goodhartz, A. S. Student attitudes and opinions relating to teaching at Brooklyn College. School and Society, 1948, 68, 345-349.

Grangin, K. L. and Painter, J. J. A new evaluation for students' course evaluation tendencies. American Educational Research Journal, 1973, 10, No. 2, 115-124.

Griffiths, D. Unpublished Ed.D. Dissertation, Rutgers University, New Brunswick, N.J., June 1973.

Grush, J. E. and Costin, F. The student as consumer of the teaching process. American Educational Research Journal, 1975, 12, 55-56.

Guthrie, E. R. The evaluation of teaching. Educational Record, 1949, 30, 109-115.

Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: University of Washington, 1954.

Hadelson, E. The validity of student rating of instructors. School and Society, 1951, 73, 265-266.

Harvey, J. N. and Barker, D. G. Student evaluation of teaching effectiveness. Improving College and University Teaching, 1970, 18, 275-278.

Hayes, J. R. Research, teaching, and faculty fate. Science, 1971, 172, 227-230.

Heilman, J. D. and Armentrout, W. D. The rating of college teachers on ten traits by their students. Journal of Educational Psychology, 1936, 27, 197-216.

Heist, P. and Yonge, G. Omnibus Personality Inventory—Form F Manual. New York: The Psychological Corporation, 1968.

Hildebrand, M. and Wilson, R. C. Effective University Teaching and Its Evaluation. Berkeley, California: University of California, 1970.

Hodgson, T. F. The General and Primary Factors in Student Evaluation of Teaching Ability. Seattle, Washington: University of Washington, 1958.

Hoyt, D. P. and Spangler, R. K. Faculty research involvement and instructional outcomes, Research Report No. 30. Kansas State University, Office of Eduational Research, 1974.

Hoyt, D. P. Instructional effectiveness. I. Measurement of effectiveness. Research Report No. 6. Kansas State University, Office of Educational Research, 1969.

Isaacson, R. L., McKeachie, W. J., and Milholland, J. E. Correlation of teacher personality variables and student ratings. Journal of Educational Psychology, 1963, 54, 110-117.

Kolodiy, G. Piagetian theory and college science teaching. The Journal of College Science Teaching, 1974, 3, 261-262.

Kolodiy, G. The cognitive development of high school and college science students. Journal of College Science Teaching, 1975, 5, 20-22.

Lanthrop, R. G. and Richmond, C. College students' evaluation of courses and instructors. Chico State College, California, 1967.

Lovell, G. D. and Haner, C. F. Forced-choice applied to college faculty rating. Educational and Psychological Measurement, 1955, 15, 291-304.

Mager, R. F. Developing Attitudes toward Learning. Palo Alto, California: Fearon, 1968.

Mann, W. R. Changes in the level of attitude sophistication of college students as a measure of teacher effectiveness. Dissertation Abstracts, 1969, 29(8-A), 2443-2444.

Maslow, A. H. and Zimmerman, W. College teaching ability, scholarly activity and personality. Journal of Educational Psychology, 1956, 47, 185-189.

McDaniel, E. D. and Feldhusen, J. F. Relationships between faculty ratings and indexes of service scholarship. Proceedings of the 78th Annual Convention of the American Psychological Association, 1970, 5, 619-620.

McKeachie, W. J. Student ratings of faculty. AAUP Bulletin, 1969(a), 55, 439-444.

McKeachie, W. J. Teaching Tips. A Guidebook for the Beginning College Teacher. (6th Ed.) Lexington, Mass.: D. C. Heath, 1969(b).

McKeachie, W. J., Lin, Y. and Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.

McKeachie, W. J., and Solomon, D. Student ratings of instructors: A validity study. Journal of Educational Research, 1958, 51, 379-382.

McKinnon, J. W. Earth science, density and college freshmen. Journal of Geological Education, 1971(a), 19, 218-220.

McKinnon, J. W. Are colleges concerned with intellectual development? American Journal of Physics, 1971(b), 39, 1047-52.

McNemar, Q. Psychological Statistics. (4th Ed.) New York: John Wiley and Sons, Inc., 1969.

Menges, R. J. The New Reporters: Students rate instruction. In C. R. Pace (Ed.) New directions in higher education: Evaluating learning and teaching. San Francisco: Jossey-Bass, 1973.

Menne, J. Teacher evaluation: Performance or effectiveness, Iowa State University. (Mimeographed.)

Rayden, N. F. College student ratings of instructors. Journal of Experimental Education, 1968, 37, 76-81.

Remmers, H. H. Appraisal of college teaching through ratings and student opinion. In 27th Yearbook of the National Society of College Teachers of Education. Chicago: University of Chicago Press, 1939.

Remmers, H. H. Reliability and halo-effect of high school and college students' judgement of their teachers. Journal of Applied Psychology, 1934, 18, 619-630.

Remmers, H. H. and Elliot, D. N. The Indian College and University staff-evaluation program. School and Society, 1949, 70, 168-171.

Remmers, H. H., Morton, F. D. and Elliot, D. N. Are students' ratings of instructors related to their grades? Purdue University Studies in Higher Education, 1949, 66, 17-26.

Remmers, H. H. and Stalnaker, J. M. Can students discriminate traits associated with success in teaching? Journal of Applied Psychology, 1928, 12, 602-610.

Rezler, A. G. The influence of needs upon the students' perception of his instructor. The Journal of Educational Research, 1965, 58, 282-286.

Rily, J. W., Ryan, B. F. and Lifschitz, M. The Student Looks at His Teacher. New Brunswick, N.J.: Rutgers University Press, 1950.

Rodin, M. and Rodin, B. Student evaluation of teachers. Science, 1972, 177, 1164-1166.

Rosenshine, B., Cohen, A. and Furst, N. Correlates of student preference ratings. Journal of College Student Personnel, 1973, 14, 269-272.

Royce, J. D. Popularity and the teacher. Education, 1956, 77, 233-237.

Russell, J. and Hollander, S. A biology attitude scale. The American Biology Teacher, 1975, 37, 270-273.

Sayre, S. A. and Ball, D. W. Piagetian development in students, The Journal of College Science Teaching, 1975, 5, 23.

Scahaff, M. E. and Siebring, B. R. What do chemistry professors think about evaluation of instruction. The Journal of Chemical Education, 1974, 51, 152-154.

Schwebel, M. Logical thinking in college freshmen. U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Research Final Report Project No. 0-B-105. Grant No. OEG-2-000329 (509), 1972, 22.

Scott, C. S. Some remarks on "student ratings": Validation. American Educational Research Journal, 1975, 12, 444-447.

Senghal, S. and Stallings, W. Some observations of the relationships between research productivity and student evaluations of course and teaching. ERIC Report ED 027 838.

Siebring, B. R. and Schaff, M. E. A survey of the literature concerning student evaluation of teaching. The Journal of Chemical Education, 1974, 51, 159-151.

Silberman, M. L. and Allender, J. S. A semiprojective technique for assessing students' reaction to college classes. The Journal of Higher Education, 1974, 6, 450.

Smith, G. M. A Simplified Guide to Statistics. New York: Holt, Rinehart and Winston, Inc., 1970.

Spencer, R. E. The Illinois course evaluation questionnaire: Manual of interpretation (Rev. ed.) Research report No. 270. Champaign, Ill.: University of Illinois, Office of Instructional Resources, Measurement and Research Division, 1968.

Spencer, R. E. and Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 7, 209-210.

Stewart, C. T. and Malpass, L. F. Estimates of achievement and ratings of instructors. Journal of Educational Research, 1966, 59, 347-350.

Tetenbaum, Toby J. The role of student needs and teacher orientations in student ratings of teachers. American Educational Research Journal, 1975, 12, 417-433.

Thorndike, E. L. A constant error in psychological ratings. Journal of Applied Psychology, 1920, 4, 25-29.

Treffinger, D. J. and Feldbuser, J. F. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 1968, 59, 297-301.

Trent, J. W. and Cohen, A. M. Research on Teaching in Higher Education. In R. M. W. Trauers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973, 997-1071.

Walker, B. D. An investigation of selected variables relative to the manner in which a population of junior college students evaluate their teachers. Dissertation Abstracts, 1969, 29(9-b), 3474.

Walsh, W. B. Validity of self-report. Journal of Counseling Psychology, 1967, 14, 18-23.

Weaver, C. H. Instructor rating by college students. Journal of Educational Psychology, 1960, 51, 21-25.

Wherry, R. J. Control of bias in ratings. PRS Reports, 914, 915, 919, 920, and 921. Department of the Army, Adjutant General's Office, Personal Research and Procedures Division, Personnel Research Branch, 1952.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Wood, K., Linsky, A. S. and Straus, M. A. Class size and student evaluation of faculty. Journal of Higher Education, 1974, 45, 524-534.

Vocks, V. W. Publications and teaching effectiveness. Journal of Higher Education, 1962, 33, 212.

Vocks, V. W. and French, S. M. Are student-ratings of teachers affected by grades? Journal of Higher Education, 1960, 31, 330-334.

Yonge, G. D. and Sassenath, J. M. Student personality correlated of teacher ratings. Journal of Educational Psychology, 1968, 59, 44-52.

Zelby, T. W. Student faculty evaluation. Science, 1974, 183, 1267-1270.

APPENDIXES

APPENDIX A

EAST TENNESSEE STATE UNIVERSITY

M E M O R A N D U M

TO: College Deans

FROM: Mack P. Davis, Vice President for Academic Affairs

SUBJECT: Meeting with Mrs. Clara Counts

DATE: April 15, 1974

On Monday afternoon, April 29, at 3:15 o'clock, you are invited to this office to meet with Mrs. Clara Counts, a member of our faculty on leave for doctoral study.

Mrs. Counts is considering, for her dissertation, making a study which would involve personnel at East Tennessee State University. I have told her that your knowledge of her proposal and your attitude toward it will give her a firm idea of how she should proceed.

Thank you.

MPD:maw

cc: Mrs. Clara Counts

MEMORANDUM

TO: O. E. Price

VIA: Dr. T. Copeland

FROM: Clara Counts

DATE: May 1, 1975

Please provide me with a printout of all freshman biology classes for spring quarter, 1975. This information is necessary in order to complete data collecting for my dissertation.

EAST TENNESSEE STATE UNIVERSITY

Johnson City, Tennessee

College of Education
Department of General
Science and Science
Education

May 1, 1975

Dr. T. P. Copeland
Chairman, Biology Department
East Tennessee State University
Johnson City, TN 37601

Dear Dr. Copeland:

Thank you for sacrificing time in your staff meeting of Monday, April 28, for me to present a proposal for collecting data necessary in completing my dissertation.

The reception of the proposal by you and the Biology Department Staff plus the willingness on everyone's part to cooperate in the data gathering process is greatly appreciated. This certainly reflects the fact that the members of the Biology Department are a research-oriented group.

Thank you again for your help which will enable me to complete the above research endeavor.

Sincerely,

Clara Counts

vdn

cc: Vice-President Mack Davis
Dean George Dove
Professor Dick Widows, Coordinator, Freshman Biology Program

MEMORANDUM

TO: All Biology Staff Teaching Freshman Biology

FROM: Clare Counts

SUBJECT: Collecting Data for Research

DATE: May 8, 1975

As you know, because of the time factor, Dr. Copeland has requested that classes resume following the collecting of data. However, it is critical (in order to make this study comparable to other studies) that the instructor whose class is participating not be present during the data-collecting process. Please advise on how instructors should be notified upon completion of the procedure.

Thank you for being truly a research-oriented group.

MEMORANDUM

TO: All Biology Staff Teaching Freshman Biology

FROM: Clare Counts

SUBJECT: Student-Ratings Research

DATE: June 5, 1975

Thank you again for your participation in the research activities concerned with student ratings. Enclosed is:

1) A copy of the Student Instructional Report completed by the students.

2) A copy of the Likert-Type Attitude Scale completed by the students.

3) A copy of the Personality Inventory completed by the students.

4) A description of the fourteen (14) scales contained in the Personality Inventory. The student personality traits which will be used in the study are marked on this enclosure.

5) A self-report form for instructors. The questions are identical to the ones on the student-rating sheet but phrased for the instructor.

   Using this self-report you will have the opportunity to compare your own perceptions with those of your students when you receive your SIR reports. In a recent study with SIR, instructors at several colleges found this report to be useful. (A report discussing this study has been published and is available, write to: Student Instructional Report, Educational Testing Service, Princeton, NJ 08540), or see me—I have a copy of all of Centra's studies with SIR.

6) Comparative data for a four-year college. Note: The two shaded vertical columns on the right contain the mean and standard deviation for each item where it is appropriate to compute them. The number, not the percentage, of all students giving a particular response is used to calculate all statistics. With the exceptions of items 6 and 16, a high mean is "better" or more positive.

The standard deviation is best described as the spread among the responses to any item. The smaller the standard deviation, the greater the consensus among students in the class. Another indication of divergence is the percentage of students who respond to each alternative for each item.

CC:jk

Enclosures (6)

APPENDIX B

Instructor's General Statement*

Some of the biology classes have been selected to participate in a different activity today.

You will have an opportunity to:

(1) rate this course and your instructor.

(2) provide us with information about your attitudes, opinions, and feelings on a variety of subjects.

This information may help us to improve our course offerings and instruction. We want to make available biology courses which best fill the needs of our students.

Please complete as many answers as time permits. Ratings will be sent to Educational Testing Service, Princeton, New Jersey, for scoring.

This is an official biology class meeting. A sheet of paper for taking the roll will be passed around after you have started working on the ratings.

*Statement read to the class by the instructor. After reading the statement, the instructor left the classroom.

Procedure*

Tell students to listen carefully to all instructions before beginning.

Hand out booklets (with answer sheets enclosed) and soft lead pencils.

Instructions for students:

1. Open booklet. Remove answer sheets. Explain answer sheets.
2. Write the SIR No. on first answer sheet.
3. Write OPI on last answer sheet.
4. Do not sign your name on any materials.
5. There are no correct answers.
6. Time is at a premium. Work rapidly.
7. When you have finished put answer sheets inside of the OPI booklet—Student Rating Sheet on top, attitude scale in the middle and the OPI answer sheet last. Clip together with paper clip. Drop pencil in the box provided, give materials to person in charge.
8. If you have any questions raise your hand and we will come to your desk and answer the questions.
9. Take roll (very carefully).

*Instruction provided for each proctor. In an orientation session just prior to collection of data each proctor was encouraged to follow the instructions outlined above.

APPENDIX C

Data are summarized and categorized for the Combined Reports and for the Percentile Distributions in the following ways:

Type of Colleges

Two-Year Colleges
Four-Year Colleges

Course Level

Freshman/Sophomore
Junior/Senior

Type of Class

Lecture Course
Lecture/Discussion
Discussion
Lecture/Lab
Lab

Subject Field

Agriculture and Forestry
Biological Sciences
Business, Commerce, Management
Education, Industrial Arts, Health, Physical Education
Engineering
Fine Arts
Foreign Languages & Literature
Home Economics
Humanities
Physical Sciences, Math, Geography
Social Sciences and Pre-law

Table C-1

Percentile Distribution of SIR Item Means for: Freshman/Sophomore Level Classes

90%	80%	70%	60%	Median 50%	40%	30%	20%	10%	SIR ITEMS	Item Nos.	Mean	Standard Deviation	Average Standard Deviation
3.72	3.61	3.54	3.44	3.36	3.29	3.20	3.08	2.88	1. The instructor's objectives for the course have been made clear	1.	3.33	.32	.58
3.67	3.55	3.46	3.38	3.30	3.22	3.15	3.00	2.82	2. There was considerable agreement between the announced objectives of the course and what was actually taught	2.	3.28	.33	.64
3.78	3.67	3.57	3.49	3.40	3.29	3.17	3.00	2.69	3. The instructor used class time well	3.	3.32	.41	.62
3.80	3.67	3.60	3.53	3.45	3.38	3.29	3.18	3.03	4. The instructor was readily available for consultation with students.	4.	3.42	.30	.58
3.55	3.42	3.28	3.19	3.11	3.00	2.91	2.74	2.54	5. The instructor seemed to know when students didn't understand the material	5.	3.07	.40	.70
2.45	2.25	2.15	2.05	1.97	1.88	1.78	1.69	1.57	6. Lectures were too repetitive of what was in the textbook(s)	6.	1.99	.36	.75
3.72	3.60	3.50	3.38	3.29	3.21	3.10	3.00	2.79	7. The instructor encouraged students to think for themselves	7.	3.27	.37	.61
3.75	3.63	3.52	3.43	3.32	3.20	3.09	2.93	2.72	8. The instructor seemed genuinely concerned with students' progress and was actively helpful	8.	3.27	.40	.64
3.58	3.41	3.27	3.15	3.00	2.89	2.75	2.60	2.35	9. The instructor made helpful comments on papers or exams	9.	2.99	.47	.70
3.63	3.48	3.36	3.26	3.15	3.04	2.93	2.78	2.57	10. The instructor raised challenging questions or problems for discussion	10.	3.12	.40	.65
3.75	3.64	3.55	3.47	3.39	3.31	3.21	3.08	2.89	11. In this class I felt free to ask questions or express my opinions	11.	3.35	.35	.66
3.85	3.77	3.71	3.63	3.55	3.47	3.36	3.23	3.00	12. The instructor was well-prepared for each class	12.	3.49	.33	.54
3.68	3.56	3.46	3.38	3.29	3.23	3.11	3.00	2.79	13. The instructor told students how they would be evaluated in the course	13.	3.26	.36	.65
3.70	3.59	3.50	3.45	3.37	3.30	3.20	3.08	2.91	14. The instructor summarized or emphasized major points in lectures or discussions	14.	3.33	.31	.61
3.57	3.40	3.27	3.14	3.02	2.90	2.75	2.59	2.29	15. My interest in the subject area has been stimulated by this course	15.	2.97	.48	.81
2.20	2.00	1.93	1.85	1.79	1.71	1.64	1.57	1.45	16. The scope of the course has been too limited; not enough material has been covered	16.	1.81	.30	.68
3.50	3.36	3.25	3.17	3.07	3.00	2.86	2.71	2.50	17. Examinations reflected the important aspects of the course	17.	3.01	.42	.71
3.56	3.41	3.32	3.22	3.12	3.03	2.94	2.82	2.69	18. I have been putting a good deal of effort into this course	18.	3.12	.34	.71
3.67	3.55	3.46	3.38	3.29	3.20	3.10	3.00	2.78	19. The instructor was open to other viewpoints	19.	3.25	.35	.61
3.75	3.63	3.53	3.45	3.38	3.29	3.20	3.04	2.82	20. In my opinion, the instructor has accomplished (is accomplishing) his or her objectives for the course	20.	3.33	.36	.60
3.81	3.62	3.50	3.39	3.27	3.17	3.08	3.00	2.86	21. For my preparation and ability, the level of difficulty of this course was	21.	3.30	.38	.69

Table C-1 (continued)

90%	80%	70%	60%	Median 50%	40%	30%	20%	10%	SIR ITEMS	Item Nos.	Mean	Standard Deviation	Average Standard Deviation
3.80	3.58	3.40	3.25	3.09	2.98	2.84	2.69	2.48	22. The work load for this course in relation to other courses of equal credit was	22.	3.12	.54	.70
3.67	3.48	3.33	3.23	3.14	3.07	3.00	2.93	2.80	23. *For me, the pace at which the instructor covered the material* during the term was	23.	3.19	.35	.58
3.88	3.80	3.73	3.67	3.59	3.50	3.39	3.27	3.10	24. *To what extent did the instructor use examples or illustrations to help clarify the material?*	24.	3.53	.31	.58
4.29	4.06	3.93	3.80	3.67	3.54	3.33	3.11	2.77	32. Overall, I would rate the textbook(s)	32.	3.59	.58	.92
4.25	4.00	3.89	3.71	3.55	3.42	3.25	3.00	2.67	33. Overall, I would rate the supplementary readings	33.	3.51	.62	.90
4.17	4.00	3.83	3.70	3.58	3.44	3.28	3.07	2.79	34. Overall, I would rate the quality of the exams	34.	3.52	.54	.89
4.55	4.38	4.23	4.09	3.94	3.76	3.58	3.33	3.00	35. I would rate the general quality of the lectures	35.	3.84	.61	.85
4.35	4.14	4.00	3.80	3.62	3.43	3.24	3.00	2.64	36. I would rate the overall value of class discussions	36.	3.56	.67	.95
4.50	4.33	4.00	3.93	3.71	3.50	3.23	3.00	2.50	37. Overall, I would rate the laboratories	37.	3.60	.78	.96
4.48	4.30	4.12	4.00	3.83	3.67	3.48	3.24	2.97	38. I would rate the overall value of this course to me as	38.	3.76	.59	.95
4.52	4.31	4.16	4.00	3.83	3.67	3.50	3.27	2.90	39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course?	39.	3.77	.62	.86

Source: Comparative Data for the Student Instructional Report, Educational Testing Service, 1973.

Table C-2

Reliability of the Student Instructional Report Items*

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

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

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION
EDUCATIONAL TESTING SERVICE, PRINCETON, N.J. 08540

ETS

BIOLOGICAL SCIENCES, HEALTH FIELDS

ADMIN. DATE	REPORT DATE	COLLEGE CODE	SIR REPORT NUMBER	CLASS ENROLL-MENT	
1972 /73				3323	1

SECTION I — All response indicators are PERCENTAGE of the total number responding, which is: 169	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Stongly Disagree	Mean*	Standard Deviation*
1. The instructor's objectives for the course have been made clear.	0	2	46	44	6	1	3.33	.32
2. There was considerable agreement between the announced objectives of the course and what was actually taught.	0	4	41	44	8	2	3.30	.30
3. The instructor used class time well.	0	3	44	42	9	2	3.32	.39
4. The instructor was readily available for consultation with students.	1	5	50	36	6	2	3.42	.36
5. The instructor seemed to know when students didn't understand the material.	1	7	31	43	15	3	3.09	.41
6. Lectures were too repetitive of what was in the textbook(s).	2	32	6	8	34	19	1.99	.46
7. The instructor encouraged students to think for themselves.	3	3	48	38	6	2	3.39	.41
8. The instructor seemed genuinely concerned with students' progress and was actively helpful.	1	4	48	37	9	2	3.38	.33
9. The instructor made helpful comments on papers or exams.	1	28	23	31	14	3	3.05	.48
10. The instructor raised challenging questions or problems for discussion.	1	14	31	39	14	2	3.15	.41
11. In this class I felt free to ask questions or express my opinions.	1	3	54	33	7	2	3.44	.36
12. The instructor was well-prepared for each class.	1	6	51	36	5	1	3.46	.35
13. The instructor told students how they would be evaluated in the course.	1	5	42	40	10	2	3.28	.38
14. The instructor summarized or emphasized major points in lectures or discussions.	1	12	40	39	7	1	3.34	.35
15. My interest in the subject area has been stimulated by this course.	0	2	46	33	13	4	3.27	.46
16. The scope of the course has been too limited; not enough material has been covered.	0	5	4	9	45	36	1.81	.32
17. Examinations reflected the important aspects of the course.	1	30	20	35	10	4	2.96	.50
18. I have been putting a good deal of effort into this course.	1	1	45	42	9	2	3.32	.32
19. The instructor was open to other viewpoints.	2	7	33	43	7	3	3.27	.35
20. In my opinion, the instructor has accomplished (is accomplishing) his objectives for the course.	1	5	46	43	6	1	3.38	.34

*OMITS and NOT APPLICABLE responses are excluded in computing the mean and standard deviation.

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION
EDUCATIONAL TESTING SERVICE, PRINCETON, N.J. 08540

BIOLOGICAL SCIENCES, HEALTH FIELDS

ADMIN. DATE	REPORT DATE	COLLEGE CODE	SIR REPORT NUMBER	CLASS ENROLLMENT
1972/73				5323

2

SECTION II — All percentages in this section are a PERCENTAGE of the total number responding, which is: 163

	Omits	1 Very Elementary	2 Somewhat Elementary	3 About Right	4 Somewhat Difficult	5 Very Difficult	Mean*	Standard Deviation*
21. For my preparation and ability, the level of difficulty of this course was:	1	2	7	56	29	5	3.30	.83

	Omits	Much Lighter	Lighter	About the same	Heavier	Much Heavier	Mean*	Standard Deviation*
22. The work load for this course in relation to other courses of equal credit was:	1	2	13	51	25	7	3.23	.82

	Omits	Very Slow	Somewhat slow	Just about right	Somewhat fast	Very fast	Mean*	Standard Deviation*
23. For me, the pace at which the instructor covered the material during the term was:	3	1	8	69	21	3	3.16	.54

	Omits	4 Frequently	3 Occasionally	2 Seldom	1 Never		Mean*	Standard Deviation*
24. To what extent did the instructor use examples or illustrations to help clarify the material?	2	63	27	6	1	0	3.55	.65

	Omits	Yes, most of the time	No, class was too large	No, class was too small	It didn't make any difference
25. Was class size satisfactory for the method of conducting the class?	1	59	13	1	15

	Omits	Major requirement or elective within major field	Minor requirement or required elective outside major field	College requirement but not part of my major or minor field	Elective not required in any way	Other
26. Which one of the following best describes this course for you?	1	51	7	14	21	3

	Omits	Friend(s) recommended it	Faculty advisors recommend	Teacher's excellent reputation	Thought I could make a good grade	Could use pass/no credit option	It was required	Subject was of interest	Other
27. Which one of the following was your most important reason for selecting this course?	2	2	4	5	1	0	39	44	3

	Omits	A	B	C	D	Fail	Pass	No Credit	Other
28. What grade do you expect to receive in this course?	3	28	45	13	1	0	6	0	1

	Omit	3.50-4.00	3.00-3.49	2.50-2.99	2.00-2.49	1.50-1.99	1.00-1.49	Less than 1.00	None yet-freshman or transfer
29. What is your approximate cumulative grade-point average?	4	11	30	28	13	3	1	0	9

	Omits	Freshman	Sophomore	Junior	Senior	Graduate	Other
30. What is your class level?	3	22	33	29	21	1	2

	Omits	Female	Male
31. Sex:	6	60	34

*OMITS responses are excluded in computing the mean and standard deviation

SECTION III		Mean	Standard Deviation
32. Overall, I would rate the textbook(s):	1 20 23 29 16 6 4	3.76	.72
33. Overall, I would rate the supplementary readings:	2 43 14 26 10 4 2	3.80	.57
34. Overall, I would rate the quality of the exams:	1 21 12 36 19 7 4	3.63	.55
35. I would rate the general quality of the lectures:	3 9 32 35 13 6 3	3.99	.62
36. I would rate the overall value of class discussions:	3 13 23 32 16 8 4	3.72	.66
37. Overall, I would rate the laboratories:	4 28 24 25 11 6 3	3.92	.62
38. I would rate the overall value of this course to me as:	2 1 40 35 13 7 3	4.05	.58
	Omits	Among the lowest 10%	Among the next 20%	About Average	In the next 20%	In the lowest 10%	Mean	Standard Deviation
39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course?	5	33	31	23	6	2	3.90	.63

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION
EDUCATIONAL TESTING SERVICE, PRINCETON, N.J. 08540

ADMIN. DATE	REPORT DATE	COLLEGE CODE	SIR REPORT NUMBER	CLASS ENROLLMENT	
0975	0975	1198	30275	0	1

SECTION I — All response summaries are a PERCENTAGE of the total number responding, which is: 24

	Omits	Not Applicable	1 Strongly Agree	2 Agree	3 Disagree	4 Strongly Disagree	Mean*	Standard Deviation*
1. The instructor's objectives for the course have been made clear.	0	0	75	21	4	0	3.71	.55
2. There was considerable agreement between the announced objectives of the course and what was actually taught.	0	0	63	25	4	8	3.42	.93
3. The instructor used class time well.	0	0	71	25	0	4	3.63	.71
4. The instructor was readily available for consultation with students.	0	8	88	4	0	0	3.95	.21
5. The instructor seemed to know when students didn't understand the material.	0	0	50	38	13	0	3.38	.71
6. Lectures were too repetitive of what was in the textbook(s).	4	13	4	0	46	33	1.70	.73
7. The instructor encouraged students to think for themselves.	0	4	46	42	8	0	3.39	.66
8. The instructor seemed genuinely concerned with students' progress and was actively helpful.	0	0	71	25	4	0	3.67	.56
9. The instructor made helpful comments on papers or exams.	4	8	42	25	13	8	3.14	1.01
10. The instructor raised challenging questions or problems for discussion.	4	4	29	50	13	0	3.18	.66
11. In this class I felt free to ask questions or express my opinions.	4	0	54	38	4	0	3.52	.59
12. The instructor was well-prepared for each class.	0	4	88	8	0	0	3.91	.29
13. The instructor told students how they would be evaluated in the course.	4	0	50	46	0	0	3.52	.51
14. The instructor summarized or emphasized major points in lectures or discussions.	0	0	71	25	4	0	3.67	.56
15. My interest in the subject area has been stimulated by this course.	0	0	50	29	13	8	3.21	.98
16. The scope of the course has been too limited; not enough material has been covered.	0	0	0	8	46	46	1.63	.65
17. Examinations reflected the important aspects of the course.	4	0	58	33	4	0	3.57	.59
18. I have been putting a good deal of effort into this course.	0	0	33	54	8	4	3.17	.76
19. The instructor was open to other viewpoints.	0	13	46	38	4	0	3.48	.60
20. In my opinion the instructor has accomplished (is accomplishing) his or her objectives in the course.	4	0	67	29	0	0	70	.47

*OMITS and NOT APPLICABLE responses are excluded in computing the mean and standard deviation.

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION
EDUCATIONAL TESTING SERVICE, PRINCETON, N.J. 08540

ADMIN. DATE	REPORT DATE	COLLEGE CODE	SIR REPORT NUMBER	CLASS ENROLLMENT	
0975	0975	1198	30275	0	2

SECTION II — All response summaries are a PERCENTAGE of the total number responding, which is: **24**

	Omits	1 Very Elementary	2 Somewhat Elementary	3 About Right	4 Somewhat Difficult	5 Very Difficult	Mean*	Standard Deviation*
21. For my preparation and ability, the level of difficulty of this course was:	0	0	4	67	29	0	3.25	.53
	Omits	Much Lighter	Lighter	About the same	Heavier	Much Heavier	Mean*	Standard Deviation*
22. The work load for this course in relation to other courses of equal credit was:	0	13	25	54	8	0	2.58	.83
	Omits	Very Slow	Somewhat slow	Just about right	Somewhat fast	Very fast	Mean*	Standard Deviation*
23. For me, the pace at which the instructor covered the material during the term was:	0	0	4	83	13	0	3.08	.41
	Omits	4 Frequently	3 Occasionally	2 Seldom	1 Never		Mean*	Standard Deviation*
24. To what extent did the instructor use examples or illustrations to help clarify the material?	0	92	8	0	0	0	3.92	.28

	Omits	Yes, most of the time	No, class was too large	No, class was too small	It didn't make any difference
25. Was class size satisfactory for the method of conducting the class?	0	54	8	0	38
	Omits	Major requirement or elective within major field	Minor requirement or required elective outside major field	College requirement but not part of my major or minor field	Elective not required in any way	Other
26. Which one of the following best describes this course for you?	0	33	8	58	0	0
	Omits	Friend(s) recommended it	Faculty advisers recommend	Teacher's excellent reputation	Thought I could make a good grade	Could use pass/no credit option	It was required	Subject was of interest	Other
27. Which one of the following was your most important reason for selecting this course?	0	8	8	13	0	0	67	4	0
	Omits	A	B	C	D	Fail	Pass	No Credit	Other
28. What grade do you expect to receive in this course?	0	42	38	21	0	0	0	0	0
	Omits	3.50-4.00	3.00-3.49	2.50-2.99	2.00-2.49	1.50-1.99	1.00-1.49	less than 1.00	None yet—freshman or transfer
29. What is your approximate cumulative grade-point average?	0	13	29	38	0	13	0	0	8
	Omits	Freshman	Sophomore	Junior	Senior	Graduate	Other	
30. What is your class level?	8	50	33	8	0	0	0	
	Omits	Female	Male					
31. Sex:	13	71	17					

*OMITS responses are excluded in computing the mean and standard deviation

STUDENT INSTRUCTIONAL REPORT

INSTITUTIONAL RESEARCH PROGRAM FOR HIGHER EDUCATION
EDUCATIONAL TESTING SERVICE, PRINCETON, N.J. 08540

ADMIN. DATE	REPORT DATE	COLLEGE CODE	SIR REPORT NUMBER	CLASS ENROLL-MENT	
0975	0975	1198	30275	0	3

SECTION III — All response summaries are a PERCENTAGE of the total number responding, which is: 24	Omits	Not applicable (Use if items or parts of items were omitted)	5 Excellent	4 Good	3 Satisfactory	2 Fair	1 Poor	Mean	Standard Deviation*
32. Overall, I would rate the textbook(s):	0	13	0	38	4	25	21	2.67	1.28
33. Overall, I would rate the supplementary readings:	0	67	0	8	17	8	0	3.00	.76
34. Overall, I would rate the quality of the exams:	0	0	29	50	13	8	0	4.00	.88
35. I would rate the general quality of the lectures:	0	0	67	25	8	0	0	4.58	.65
36. I would rate the overall value of class discussions:	4	4	21	29	29	8	4	3.59	1.10
37. Overall, I would rate the laboratories:	4	79	4	0	8	4	0	3.25	1.26
38. I would rate the overall value of this course to me as:	0	0	38	50	8	0	4	4.17	.92
	Omits		Among the top 10%	Among the top 30%	About Average	In the lowest 30%	In the lowest 10%	Mean	Standard Deviation*
39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course?	0		67	21	4	4	4	4.42	1.06

SECTION IV — All response summaries are a PERCENTAGE of the total number responding, which is: 24		Omits	Not Applicable 0	1	2	3	4	5	6	7	8	9
INSTRUCTOR'S SUPPLEMENTARY QUESTIONS AND RESPONSE OPTIONS	40.											
	41.											
	42.											
	43.											
	44.											
	45.											
	46.											
	47.											
	48.											
	49.											

*OMITS and NOT APPLICABLE responses are excluded in computing the mean and standard deviation.

I.N. 397838

APPENDIX D

Table D-1

Correlations between the OPI and the Allport-Vernon-Lindzey Study of Values (Revised Edition)
(N = 171 College Freshmen)

	STUDY OF VALUES							
OPI	Theoretical	Economic	Aesthetic	Social	Political	Religious	Mean	SD
TI	.05	-.63	.47	.24	-.17	.04	31.6	7.1
TO	.62	-.32	.10	-.10	.00	-.21	23.1	5.6
Es	-.16	-.50	.61	.26	-.20	-.05	15.2	5.0
Co	.11	-.36	.40	.04	-.06	-.13	17.9	5.6
Au	.16	-.29	.44	.11	-.20	-.23	30.3	6.2
RO	.32	-.03	.32	.02	.16	-.66	16.6	5.5
SE	-.14	-.09	.03	.21	-.04	.05	23.3	7.1
IE[a]	-.03	.02	.25	-.06	.18	-.31	28.1	9.7
PI	.12	-.03	-.11	.09	-.05	.01	30.6	9.0
AL	.16	.00	.02	-.04	-.10	-.04	12.5	4.6
Am	-.18	-.48	.17	.46	-.30	.28	23.3	5.9
PO	-.02	.62	-.42	-.18	.19	-.15	9.6	5.2
MF	.55	.45	-.46	-.45	.32	-.29	26.4	7.0
RB	.23	-.24	-.02	.01	-.10	.13	14.6	4.2
Mean	46.8	33.0	45.0	38.6	39.3	37.3		
SD	8.4	9.9	10.2	7.9	7.0	10.7		

Note.—r of .20 required, to be significant at .01 level.

[a]The correlations between IE and the other variables in the table are based on a version of IE which contained five more items than the present scale. On the present IE scale, the mean would be about 2.5 points lower and the standard deviation would be about 1.1 points smaller.

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory," Manual, Form F, The Psychological Corporation, 1968.

Table D-2

Correlations of the OPI with Certain Experimental Measures and the College Entrance Examination Board Scholastic Aptitude Test (SAT)
(N = 94 College Freshmen)

	EXPERIMENTAL MEASURE						
		Complex		SAT[d]		OPI[e]	
OPI	Construct Complexity[a]	Figure Preference[b]	Problem Solving[c]	Verbal	Mathematical	Mean	SD
TI	.42	.28	.31	.51	.12	26.0	8.1
TO	.25	.04	.53	.52	.46	21.4	5.4
Es	.37	.32	.19	.28	-.10	11.7	5.1
Co	.44	.32	.32	.32	.23	15.1	5.2
Au	.38	.11	.38	.55	.39	24.5	7.8
RO	.31	.08	.22	.26	.27	13.5	5.6
SE[f]	-.05	.06	-.15	-.14	-.14	22.5	6.6
IE[f]	.20	.08	.22	.32	.14	27.1	9.7
PI	-.15	-.24	-.14	-.04	.10	31.0	10.7
AL	-.10	-.11	-.08	-.15	-.08	12.5	4.9
Am	-.01	.23	-.09	-.10	-.12	19.8	5.8
PO	-.48	-.23	-.41	-.43	-.30	14.2	5.7
MF	-.23	-.38	.19	.04	.51	29.4	6.8
RB	-.03	-.12	.14	.03	.15	14.1	3.9
Mean	8.2	11.1	6.8	541.6	585.2		
SD	3.3	5.8	3.1	100.7	106.7		

Note.—For N = 94, r of .27 required, to be significant at .01 level; for N = 43, r of .40 is necessary.

[a]Student was required to generate his own categories for grouping various objects. Score range: 1-20.

[b]Student was required to select the 12 most preferred figure drawings from a pool of 24. Score range: 0-24.

[c]Student was asked to solve problems which required changes in his mental set. Score range: 0-14.

[d]SAT scores were available for only 43 students.

[e]Based on N of 94.

[f]See footnote a, Table D-1, page 135.

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory," Manual, Form F, The Psychological Corporation, 1968.

Correlations between the OPI and the Guilford-Zimmerman Temperament Survey
(N = 98 Male College Freshmen)

	G-Z TEMPERAMENT SURVEY										OPI	
OPI	General Activity	Restraint	Ascend-ance	Socia-bility	Emotional Stability	Objec-tivity	Friend-liness	Thought-fulness	Personal Relations	Mascu-linity	Mean	SD
TI	.14	.46	.33	.11	-.01	.13	.09	.68	-.21	.08	27.8	7.0
TO	.08	.33	.04	-.04	.12	.21	.03	.51	.06	.16	23.0	5.1
Es	.16	.25	.32	.07	-.12	-.12	-.01	.44	-.28	.02	12.3	4.7
Co	.14	.13	.15	-.08	-.18	-.05	-.01	.38	-.15	.15	16.3	4.9
Au	-.07	-.01	-.11	-.13	-.01	.28	.16	.06	-.03	.02	27.7	6.4
RO	-.13	-.16	-.28	-.13	-.03	.11	.09	-.08	-.11	-.07	15.1	5.3
SE	.42	-.24	.68	.84	.27	.19	-.18	.06	.08	.36	24.6	6.9
IE[a]	.18	-.47	.23	.25	-.32	-.38	-.46	-.09	-.29	.09	30.6	9.5
PI	.15	.00	.27	.40	.77	.67	.29	-.08	.43	.33	32.8	10.3
AL	.19	-.18	.17	.43	.71	.70	.24	-.11	.41	.40	13.5	4.4
Am	.20	.27	.32	.36	.27	.32	.30	.25	.26	.12	20.2	5.6
PO	.04	-.35	.05	.16	-.07	-.30	-.31	-.37	-.05	-.05	11.6	5.6
MF	-.04	-.04	-.18	-.06	.33	.24	.04	-.17	.16	.12	30.8	5.9
RB	.26	.32	.33	.27	.52	.42	.17	.33	.22	.22	15.0	4.6
Mean	15.9	17.7	16.9	17.9	16.4	17.0	13.9	21.0	16.7	19.2		
SD	5.5	4.7	6.4	6.6	6.3	5.1	4.8	4.9	5.1	3.7		

Note.—r of .27 required, to be significant at .01 level.

[a]See footnote a, Table 1, page

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory," Manual, Form F, The Psychological Corporation, 1968.

Correlations between the OPI and the Activities Index (AI)
(N = 165 College Freshmen)

AI	OPI TI	TO	Es	Co	Au	RO	SE	IE[a]	PI	AL	Am	PO	MF	RB	AI Mean	AI SD
Abasement	.07	−.05	.01	−.11	−.08	−.17	.01	−.21	.01	−.04	.20	.02	−.19	.08	4.1	1.6
Achievement	.20	.28	.10	.16	−.12	−.06	.23	.15	.08	.17	.18	.04	.04	.27	6.1	2.3
Adaptiveness	.18	.12	.01	.05	.14	.14	.09	−.22	.21	.18	.25	−.16	−.08	.16	5.8	2.3
Affiliation-Rejection	−.15	−.21	−.14	−.32	−.39	−.22	.57	−.07	.11	.10	.16	.33	−.09	.07	4.8	2.7
Aggression-Blameavoidance	.00	.00	.10	.24	.06	.00	−.12	.47	−.30	−.17	−.28	−.03	.06	−.30	4.8	2.1
Change-Sameness	.04	.04	.22	.46	.23	.12	−.01	.31	−.15	−.08	−.01	−.22	−.23	−.20	7.0	2.0
Conjunctivity-Disjunctivity	.14	.12	−.01	−.41	−.28	−.07	.19	−.23	.16	.14	.14	.27	.04	.36	4.9	2.4
Counteraction-Infavoidance	.30	.30	.03	.23	.04	.01	.22	−.04	.26	.33	.35	−.12	.02	.31	6.8	2.2
Deference	.08	.05	.07	−.31	−.32	−.14	.31	−.27	.23	.19	.29	.25	−.09	.28	6.0	2.2
Dominance	.07	.09	.02	.06	−.11	−.06	.31	.25	.10	.15	.05	.10	.15	.16	6.3	2.4
Ego Achievement	.29	.17	.08	.08	−.03	−.09	.31	.07	.17	.25	.28	−.05	.10	.29	6.8	2.5
Emotionality-Placidity	.05	−.11	.22	.27	.21	.07	−.09	.21	−.23	−.22	.06	−.23	−.44	−.29	5.4	2.2
Energy-Passivity	.07	.12	.03	.18	.07	.01	.20	.04	.15	.19	.18	−.10	−.04	.13	7.3	1.5
Exhibitionism-Infavoidance	.05	.00	.05	.07	−.06	−.05	.45	.28	.13	.15	.08	.05	.07	.05	3.9	2.3
Fantasied Achievement	−.05	−.02	.07	−.09	−.26	−.17	.11	.30	−.16	−.09	−.11	.26	.14	.02	3.7	2.3
Harmavoidance	.00	−.11	.03	−.26	−.14	−.05	.05	−.38	.03	−.09	.09	.15	−.21	.04	4.2	2.5
Humanism	.35	.14	.34	.10	.02	−.07	.08	−.04	.07	.09	.22	−.08	−.25	.22	8.4	1.8
Impulsion-Deliberation	−.02	−.15	.13	.31	.19	.00	−.10	.42	−.29	−.19	−.15	−.15	−.19	−.41	6.5	2.1
Narcissism	−.09	−.11	.11	−.16	−.25	−.15	.11	.16	−.22	−.25	−.06	.21	−.23	−.13	4.6	2.4
Nurturance-Rejection	.05	−.06	.02	−.09	−.14	−.12	.29	−.03	.13	.07	.36	.13	−.18	.13	7.1	2.0
Objectivity	.17	.23	.08	.15	.17	.24	.10	.10	.07	.10	.00	−.10	.13	.14	9.0	1.2
Order	−.10	−.05	−.02	−.49	−.47	−.20	.19	−.17	.00	.05	.00	.42	−.06	.17	4.1	3.0
Play	−.18	−.16	−.03	−.02	−.03	.02	.17	.18	−.12	−.12	−.11	.09	−.06	−.26	4.3	2.2
Pragmatism	.01	.24	−.04	−.16	−.32	−.19	.25	−.03	.12	.20	.13	.29	.17	.30	5.3	2.5
Reflectiveness	.27	.24	.31	.18	.16	.02	−.09	.13	−.19	−.12	.07	−.19	−.16	.02	7.9	1.5
Scientism	.16	.47	.11	.00	−.05	−.08	.05	−.13	.13	.17	.10	.02	.18	.40	5.3	3.2
Sentience	.06	−.08	.24	.24	.04	.18	−.11	.24	−.17	−.15	−.09	−.08	−.25	−.24	5.5	1.6
Sex-Prudery	−.09	−.16	.03	−.04	−.13	−.10	.20	.11	−.04	−.11	.04	.12	−.24	−.09	4.5	2.3
Succorance-Autonomy	−.05	−.05	−.02	−.28	−.34	−.24	.29	−.19	.15	.10	.21	.25	−.08	.16	5.6	2.2
Understanding	.33	.44	.19	.11	.01	−.05	.14	−.01	.11	.15	.22	−.09	.04	.36	7.7	1.7
Mean	32.7	20.8	17.3	22.1	32.2	17.7	22.0	36.2	27.9	10.6	22.6	9.2	24.7	12.0		
SD	5.8	4.8	3.9	5.0	6.9	4.2	7.1	9.7	11.3	4.8	5.4	5.5	5.5	4.4		

Correlations of the OPI with the Miller Analogies Test and Faculty Ratings
(N = 37 Graduate Students)

OPI	Miller Analogies Test	FACULTY RATING[a] A	B	C	D	E	F	G	H	OPI Mean	OPI SD
TI	-.18	.41	.42	.06	.35	.35	.33	.42	.41	35.6	4.3
TO	.17	.28	.35	.04	.30	.28	.26	.29	.33	23.6	3.7
Es	.03	-.13	-.12	-.27	-.17	-.15	-.28	-.24	-.19	14.9	4.6
Co	-.14	.07	.22	-.03	.09	.08	-.06	.02	.06	18.8	4.8
Au	-.08	.06	.14	-.01	-.11	.12	-.16	-.08	-.01	34.2	5.9
RO	-.03	.00	.08	-.19	-.16	-.03	-.32	-.22	-.12	15.1	6.1
SE	-.35	.03	.03	.30	.25	.07	.41	.33	.18	24.0	6.5
IE[b]	.19	-.07	.11	-.19	-.08	-.05	-.14	-.18	-.06	28.3	11.7
PI	-.15	.17	.11	.27	.21	.18	.40	.40	.28	40.6	10.2
AL	.02	.03	.01	.23	.10	.10	.24	.28	.17	14.1	4.9
Am	.40	.20	.20	.18	.25	.23	.35	.44	.27	24.1	5.0
PO	.15	-.33	-.39	-.11	-.22	-.31	-.06	-.22	-.27	5.9	3.2
MF	.29	.11	.15	.17	.09	.18	.19	.20	.22	28.5	5.7
RB	.11	.27	.18	.23	.28	.17	.43	.42	.38	16.7	5.0
Mean	73.0	55.7	54.1	54.8	53-0	54.0	55.3	56.6	54.3		
SD	15.6	9.1	9.4	9.3	8.6	9.3	8.0	10.3	10.0		

Note.—r of .44 required, to be significant at .05 level.

[a]A: Power of assimilation and logic
B: Self-reliance, originality, and freedom in thinking
C: Oral assignment presentation
D: Written performance
E: Master of field (in course or field rated)

F: Individual vigor
G: Attitudes and reactions toward work
H: Overall evaluation (by professor)

[b]See footnote a, Table D-1, page 135.

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory, Manual, Form F, The Psychological Corporation, 1968.

Scale	INTERNAL CONSISTENCY: Freshman at 37 Colleges[b] (N = 7283) = r_{tt}[c]	Mean	SD	Freshmen at One College (N = 400) r_{11}[d]	Mean	SD	TEST-RETEST[a]: Women at Three Colleges (N = 67) r_{12}	$Mean_1$	SD_1	$Mean_2$	SD_2	Upperclassmen at One College (N = 71) r_{12}	$Mean_1$	SD_1	$Mean_2$	SD_2
TI	.85	25.3	7.9	.86	25.2	8.1	.94	30.0	8.1	30.1	8.0	.89	23.9	7.6	23.4	8.0
TO	.78	19.6	5.7	.80	21.3	5.4	.84	20.9	4.9	21.6	5.8	.87	17.3	5.0	17.1	5.7
Es	.82	12.2	5.2	.84	11.3	5.0	.89	14.7	4.9	15.7	5.0	.89	11.3	4.9	11.3	5.6
Co	.76	15.3	5.5	.73	14.8	5.0	.91	17.7	6.2	18.1	6.3	.93	14.5	5.2	15.0	5.8
Au	.86	23.4	8.4	.82	24.9	7.1	.88	28.1	7.5	28.9	8.3	.87	28.8	6.0	28.9	6.6
RO	.86	11.8	6.2	.91	13.7	6.1	.92	14.3	5.7	14.4	6.3	.91	12.4	5.4	12.8	5.4
SE	.83	23.4	7.1	.88	23.0	7.3	.87	22.6	6.8	21.7	7.4	.92	23.2	7.3	22.7	7.7
IE	.83	25.6	8.9	.82	25.5	8.9	.87	23.2	9.1	23.7	9.6	.93	26.2	10.2	26.4	11.0
PI	.89	29.9	10.5	.91	32.3	10.2	.87	33.7	9.4	34.7	10.2	.91	36.7	11.1	35.9	12.1
AL	.82	12.3	4.6	.84	13.0	4.4	.79	13.8	4.0	13.8	4.6	.84	13.6	4.5	13.3	4.7
Am	.74	20.8	5.6	.83	19.6	5.7	.81	24.4	4.7	23.8	4.6	.90	23.4	5.9	23.0	6.0
PO	.84	14.8	6.4	.79	14.2	5.5	.89	9.8	5.6	10.1	6.0	.89	13.8	4.8	13.5	5.4
MF	.73	28.4	7.1	.76	31.0	7.3	.87	24.6	5.6	24.8	5.6	.88	28.0	6.5	28.2	6.7
RB	.67	13.4	4.4	.65	14.2	4.2	.84	14.9	4.2	14.9	4.2	.86	12.7	4.5	12.6	4.6

[a]The time interval between the two test administrations was between three and four weeks for all students.

[b]Normative sample.

[c]Based on Kuder-Richardson Formula 21.

[d]Split-half correlation corrected by Spearman-Brown formula.

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory," Manual, Form F, The

Intercorrelation of the OPI Scales
(Based on Normative Sample of 7283 College Freshmen)[a]

OPI	TI	TO	Es	Co	Au	RO	SE	IE[b]	PI	AI	Am	PO	MF	Rb	MEN Mean	MEN SD
TI		.62	.65	.57	.50	.27	.10	.04	.10	.09	.47	-.65	-.35	.39	24.5	8.0
TO	.63		.32	.49	.43	.35	.02	.00	.14	.22	.29	-.50	.10	.53	21.1	5.5
Es	.64	.42		.50	.29	.20	.03	.27	-.19	-.13	.26	-.33	-.59	.13	10.6	5.2
Co	.56	.50	.50		.58	.40	-.02	.38	-.13	.01	.20	-.62	-.22	.06	15.6	5.3
Au	.49	.38	.30	.61		.62	-.11	.21	.10	.11	.18	-.74	-.11	.06	24.0	8.3
RO	.32	.35	.26	.46	.63		-.19	.28	-.08	.01	-.12	-.40	.02	-.07	12.6	6.2
SE	.12	.05	.01	.01	-.07	-.14		.08	.42	.31	.45	-.01	-.14	.26	22.6	7.3
IE[b]	.20	.17	.35	.54	.34	.37	.04		-.49	-.31	-.30	.02	-.13	-.44	30.7	9.8
PI	.08	.06	-.18	-.18	.06	-.09	.43	-.47		.69	.50	-.19	.26	.60	30.3	10.4
AL	.14	.21	-.05	.00	.12	.05	.32	-.27	.65		.36	-.08	.41	.57	12.5	4.6
Am	.42	.25	.17	.17	.23	-.06	.41	-.23	.52	.41		-.33	-.19	.46	19.2	5.6
PO	-.61	-.41	-.26	-.61	-.76	-.44	-.03	-.15	-.17	-.08	-.33		.25	-.07	15.1	6.4
MF	-.24	.17	-.37	-.15	-.12	.03	-.14	-.14	.20	.38	-.23	.27		.19	33.1	5.7
RB	.37	.53	.17	.00	.00	-.05	.21	-.34	.53	.57	.53	-.18	.28		13.7	4.5
WOMEN Mean	26.0	18.2	13.7	15.0	22.9	11.1	24.1	25.7	29.4	12.0	22.4	14.6	23.9	13.0		
WOMEN SD	7.8	5.5	4.8	5.6	8.4	6.0	6.8	9.6	10.3	4.6	5.2	6.4	5.4	4.3		

[a]Correlations above diagonal are for 3540 men; correlations below diagonal are for 3743 women.

[b]See footnote a, Table D-1, page 135.

Source: Paul Heist and George Yonge, "Omnibus Personality Inventory," Manual, Form F, The Psychological Corporation, 1968.

APPENDIX E

An example of a negative item and a positive item from the Biology Attitude Likert-Type Scale; and an example of a positive item and a negative item from the Semantic Differential Scale*

Item 2—I have always enjoyed studying biology in school. (Positive Item)

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
(5)	(4)	(3)	(2)	(1)

Item 12—It makes me nervous to even think about doing a Biology Experiment. (Negative Item)

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
(5)	(4)	(3)	(2)	(1)

Biology is:

Item 15	Good	A (5)	B (4)	C (3)	D (2)	E (1)	Bad
Item 18	Cruel	A (1)	B (2)	C (3)	D (4)	E (5)	Kind

*The Semantic Differential Scale was originally developed by Osgood, Suci, and Tannebaum and published in The Measurement of Meaning, University of Illinois Press, Urbana, Ill., 1957.

Source: The American Biology Teacher, May, 1975.

APPENDIX F

STUDENT INSTRUCTIONAL REPORT

SIR Report Number

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

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

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

		NA	SA	A	D	SD
1.	The instructor's objectives for the course have been made clear	0	4	3	2	1
2.	There was considerable agreement between the announced objectives of the course and what was actually taught	0	4	3	2	1
3.	The instructor used class time well	0	4	3	2	1
4.	The instructor was readily available for consultation with students	0	4	3	2	1
5.	The instructor seemed to know when students didn't understand the material	0	4	3	2	1
6.	Lectures were too repetitive of what was in the textbook(s)	0	4	3	2	1
7.	The instructor encouraged students to think for themselves	0	4	3	2	1
8.	The instructor seemed genuinely concerned with students' progress and was actively helpful	0	4	3	2	1
9.	The instructor made helpful comments on papers or exams	0	4	3	2	1
10.	The instructor raised challenging questions or problems for discussion	0	4	3	2	1
11.	In this class I felt free to ask questions or express my opinions	0	4	3	2	1
12.	The instructor was well-prepared for each class	0	4	3	2	1
13.	The instructor told students how they would be evaluated in the course	0	4	3	2	1
14.	The instructor summarized or emphasized major points in lectures or discussions	0	4	3	2	1
15.	My interest in the subject area has been stimulated by this course	0	4	3	2	1
16.	The scope of the course has been too limited; not enough material has been covered	0	4	3	2	1
17.	Examinations reflected the important aspects of the course	0	4	3	2	1
18.	I have been putting a good deal of effort into this course	0	4	3	2	1
19.	The instructor was open to other viewpoints	0	4	3	2	1
20.	In my opinion, the instructor has accomplished (is accomplishing) his or her objectives for the course	0	4	3	2	1

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

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

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

23. For me, the pace at which the instructor covered the material during the term was:
1 Very slow
2 Somewhat slow
3 Just about right
4 Somewhat fast
5 Very fast

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

Questionnaire continued on the other side.

572MRC114P200X
283562

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

(1) Yes, most of the time
(2) No, class was too large
(3) No, class was too small
(4) It didn't make any difference one way or the other

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

(1) Major requirement or elective within major field
(2) Minor requirement or required elective outside major field
(3) College requirement but not part of my major or minor field
(4) Elective not required in any way
(5) Other

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

(1) Friend(s) recommended it
(2) Faculty advisor's recommendation
(3) Teacher's excellent reputation
(4) Thought I could make a good grade
(5) Could use pass/no credit option
(6) It was required
(7) Subject was of interest
(8) Other

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

(1) A
(2) B
(3) C
(4) D
(5) Fail
(6) Pass
(7) No credit
(8) Other

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

(1) 3.50-4.00
(2) 3.00-3.49
(3) 2.50-2.99
(4) 2.00-2.49
(5) 1.50-1.99
(6) 1.00-1.49
(7) Less than 1.00
(8) None yet--freshman or transfer

30. What is your class level?

(1) Freshman
(2) Sophomore
(3) Junior
(4) Senior
(5) Graduate
(8) Other

31. Sex:

(1) Female
(2) Male

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

		Not applicable, don't know, or there were none.	Excellent	Good	Satisfactory	Fair	Poor
32.	Overall, I would rate the textbook(s)	(0)	(5)	(4)	(3)	(2)	(1)
33.	Overall, I would rate the supplementary readings	(0)	(5)	(4)	(3)	(2)	(1)
34.	Overall, I would rate the quality of the exams	(0)	(5)	(4)	(3)	(2)	(1)
35.	I would rate the general quality of the lectures	(0)	(5)	(4)	(3)	(2)	(1)
36.	I would rate the overall value of class discussions	(0)	(5)	(4)	(3)	(2)	(1)
37.	Overall, I would rate the laboratories	(0)	(5)	(4)	(3)	(2)	(1)
38.	I would rate the overall value of this course to me as	(0)	(5)	(4)	(3)	(2)	(1)

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

One of the most effective (among the top 10%)	More effective than most (among the top 30%)	About average	Not as effective as most (in the lowest 30%)	One of the least effective (in the lowest 10%)
(5)	(4)	(3)	(2)	(1)

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

	NA										NA									
40.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	45.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
41.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	46.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
42.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	47.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
43.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	48.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
44.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	49.	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

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

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

BIOLOGY ATTITUDE SCALE

Each of the statements below expresses a feeling toward biology. Please rate each statement on the extent to which you agree. For each, you may (A) strongly agree, (B) agree, (C) be undecided, (D) disagree, or (E) strongly disagree.

After you have made your choice, place letter which best represents your feelings in the space at the left of each statement.

A	B	C	D	E
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

____ 1. Biology is very interesting to me.

____ 2. I don't like biology, and it scares me to have to take it.

____ 3. I am always under a terrible strain in a biology class.

____ 4. Biology is fascinating and fun.

____ 5. Biology makes me feel secure, and at the same time it is stimulating.

____ 6. Biology makes me feel uncomfortable, restless, irritable, and impatient.

____ 7. In general, I have a good feeling toward biology.

____ 8. When I hear the word biology, I have a feeling of dislike.

____ 9. I approach biology with a feeling of hesitation.

____10. I really like biology.

____11. I have always enjoyed studying biology in school.

____12. It makes me nervous to even think about doing a biology experiment.

____13. I feel at ease in biology and like it very much.

____14. I feel a definite positive reaction to biology; it's enjoyable.

Below are some scales on which we would like you to rate your feelings toward biology. On each scale, you can rate your feelings toward biology as an A, B, C, D, or E. There are no correct answers. Also, some of the scales seem to make more sense than others. Don't worry about it. Just rate your feelings toward biology on these scales as best you can. Please don't leave any scales blank.

For your response to each scale, circle letter selected.

BIOLOGY IS:

15.	Good	A	B	C	D	E	Bad
16.	Clean	A	B	C	D	E	Dirty
17.	Worthless	A	B	C	D	E	Valuable
18.	Cruel	A	B	C	D	E	Kind
19.	Pleasant	A	B	C	D	E	Unpleasant
20.	Sad	A	B	C	D	E	Happy
21.	Nice	A	B	C	D	E	Awful
22.	Fair	A	B	C	D	E	Unfair

SAMPLE QUESTIONS FROM THE OMNIBUS PERSONALITY INVENTORY*

1. I dislike mathematics.

2. I prefer having a principle or theory explained to one rather than attempting to understand it on my own.

3. I like to solve puzzles.

4. Novelty has a great appeal to me.

5. I like to fool around with new ideas, even if they turn out later to have been a total waste of time.

6. I prefer to engage in activities from which I can see definite results rather than those from which no tangible or objective results are apparent.

7. I question the accuracy of statements made in my textbooks or reference books.

8. I much enjoy thinking about some problem which is a challenge to the experts.

*There are 395 questions in the inventory. Students were asked to mark true if the statement is true or mostly true for him and false or not usually true as applied to him.

SOURCES OF RATING FORMS AND RELATED ISSUES

Investigators at Purdue University and at the University of Washington, Illinois, Iowa, Miami, Michigan, Minnesota, and Texas, among others, have studied student rating forms for some years and have developed a variety of useful instruments (Bendig, 1925b, 1952c, 1954; Brandenburg & Remmers, 1927; Costin (in press); Crannell, 1953; Eckert, 1950; Gibb, 1955; Guthrie, 1954; Holmes, 1971; Isaacson, McKeachie, Millholland, Lin, Hofeller, Baerwaldt, and Zinn, 1964; Leftwich & Remmers, 1962; McKeachie, 1969a, 1969b; Remmers, 1960; Remmers & Brandenburg, 1927; Remmers & Elliot, 1949; Spencer, 1968; Spencer & Aleamoni, 1970; Stuit & Ebel, 1952). An interesting variation on the usual type of rating form is a unidimensional scale developed by Hayes (1963a, 1963b) which yields a single score of teaching effectiveness based on students' responses to items cast in the familiar multiple-choice form of achievement test.

Werdell (1967) and Eble (1970) provided excellent brief histories of the student evaluation movement in American colleges, described a number of instruments currently being used in various colleges and universities, and discussed the practice and importance of using rating forms (including the issues involved in publishing the results).

Simpson and Seidman (1962) prepared a manual to help teachers develop their own evaluation of teaching; these behavioral items were ranked in the order in which students, alumni, and faculty regarded their value. Gadzella (1968) and Gray (1969) also derived empirically extensive lists of criteria which can be useful in developing rating scales, and Deshpande, Webb, and Marks (1970) derived through factor analysis 14 dimensions of teaching behavior represented by 147 behavioral items.

To make the evaluation of teaching behavior as meaningful as possible, Farrar (1969) suggested the "critical incidents" approach (Flanagan, 1954). He advocated collecting "critical incidents" of faculty behavior from both faculty members and students, classifying them, and then having students judge the frequency with which they occurred in the classroom.

VITA

Clare Potter Counts was born in Johnson City, Tennessee, on September 26, 1931. She attended the elementary school on the campus of East Tennessee State University and was graduated from the University High School in 1949. The following September she entered East Tennessee State University, and in June 1956, she received a Bachelor of Science degree in English. In August of 1956 she accepted a teaching position at Boone High School, Orlando, Florida. During the period of 1957-1961 she taught in the Norfolk, Virginia, school system, the Washington County, Tennessee, school system and the Bristol, Tennessee, school system. In the fall of 1960 she began study toward a Master's degree in Biology. She received this degree in August 1961. In the fall of 1962, she accepted an instructor's position in the Biology Department of East Tennessee State University. In the Spring of 1963 she accepted an instructor's position at Purdue University. From 1965-1973 she taught in the Department of General Science and Science Education at East Tennessee State University.

She entered Graduate School at the University of Tennessee in September 1973 and received the Doctor of Education degree with a major in Curriculum and Instruction and minors in Science Education and Zoology in June 1976. Presently, she is teaching in the Department of General Science and Science Education at East Tennessee State University.

She is married to Perry Dalmond Counts, formerly of Dickinson County, Virginia. They have three children, Scott, Marc, and Kim.